MACQUARIE University

People do not empathise and help others based on facial resemblance

Master of Research Thesis

By: **Andrejs Enriquez**

Supervisor: Dr. Ian Stephen

A thesis submitted in partial fulfilment of the requirements of a Master of Research Degree

Macquarie University

Faculty of Psychology

December 2017

Word count: 20086

Statement of Candidate

I hereby confirm that all material contained in this project are my original authorship and ideas, except where the work of others has been acknowledged or referenced. I also confirm that the work has not been submitted for a higher degree to any other university or institution. The research project was approved by the Macquarie University Human Research Ethics Committee (Approval No. 5201700157) Signed:

Acknowledgements

I would like to acknowledge everyone who played a role in my academic achievements.

I would like to thank my supervisor Dr Ian Stephen, who has kept me on track, gets back to my many questions and drafts with unbelievable speed even during busy times, and has taught me that if something seems impossible, there's often a simpler way to do it.

I would like to thank my Master of Research coordinator Dr Simon Boag, who has set up many insightful and motivating workshops and events this year.

Finally I will like to thank my mother and father who have proofread my drafts, supported me finically for recruiting and drives to Sydney, and supported me emotionally which has kept me going strong.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Abstract

Background: Altruism may have evolved through kin selection by maximising inclusive fitness, so people may have evolved to preferentially help kin over non-kin. Previous research suggests that empathy may be one the main contributors to helping behaviour, and that people trust others who look similar to themselves. However, whether this leads to helping, and whether empathy towards others changes according to their physical similarity remains unknown.

Aim: To examine the effect of within-race self-similarity on cognitive empathy, emotional empathy, and helping.

Method: The faces of 56 Caucasian participants (17 male, 27 female; aged 18-29) were secretly recorded responding to happy and sad video clips. Happy and sad composite faces (validated by a rating task) were morphed to look similar and dissimilar to each participant. A total of 15 male and 18 female participants returned to make judgements of each face, including how well they recognise the faces' emotional expression, how much they feel that emotion themselves, and whether they chose to help people with a similar or a dissimilar face in a low risk scenario.

Results: There was no significant difference for either cognitive empathy, emotional empathy, or helping similar compared to dissimilar faces. There was also no significant difference in cognitive compared to emotional empathy for the preference of similar over dissimilar faces. Nor did a self-preference in cognitive empathy nor a self-preference in emotional empathy significantly predict a self-preference in helping

Conclusion: This study has demonstrated that empathy and helping are not affected by self-resemblance, and offers an alternative way that kin selection functions in humans.

Table of Contents

Kin Selection

An action which benefits another person even at the short-term expense to oneself is considered altruistic. Charles Darwin struggled with the paradox that people can be altruistic. With the information available at the time, he was unable to understand why people often help others even when it is at a cost to themselves. In The Descent of Man (Dawin, 1871, p.110) he stated "He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature." It wasn't until almost a century later that Hamilton (1964) proposed his theory of inclusive fitness stating that genes can propagate along two routes: a direct route, and an indirect route. An individual can be evolutionarily successful by directly passing their own genes down through their own reproduction, or by indirectly passing their genes down through assisting the reproductive success of family members with whom they share a portion of their genes (Gardner, West, & Wild, 2011). Maynard Smith called this indirect route kin selection (Smith, 1964). Therefore, it has been suggested that people may have evolved to preferentially help kin over non-kin.

Altruistic actions towards relatives is known as kin altruism. Hamilton's rule ($rB - C > 0$) determines the likelihood of altruistic behaviour based on the probability that a gene may be shared with a relative. If probability that two individuals share an allele due to recent common descent (r) multiplied by the benefit to the altruist in terms of inclusive fitness (B) minus the cost of helping (c) is greater than zero, then altruism is possible (Hamilton, 1964). Therefore, although an individual acting altruistically may incur a personal cost which hinders their own fitness, their genes still benefit.

The sting of a bee is a powerful example of the theory of inclusive fitness. A colony of bees share a large portion of their genes with each other, so if the bee perceives you as a potential threat to its colony, it will sting you at the cost of its life because the inclusive

fitness benefit to saving the colony is greater than the inclusive fitness cost of dying itself. Similarly, kin selection can be viewed in terms of multiple people, because the more people that can be helped who are likely to share copies of the same genes with each other, the greater likelihood of helping them, and the greater chance a person would risk their own self to help them. A famous example of this was illustrated J. R. S. Haldane, who light-heartedly wrote that he would not jump in a river to save a brother from drowning, but he would gladly do it for two brothers or eight cousins, because mathematically the inclusive fitness advantage of helping them is equal to him surviving (1955). In reality, kin selection is more complex than this simple mathematical calculation, but this example's underlying notion, that people are more likely to help others based on the probability that they may share genes is the foundation of kin selection.

Kin Recognition

Identifying the degree if genetic relatedness between others and oneself is known as kin recognition. In many cases such as first encounters, it is unknown whether people are related to each other, thus physical cues such as facial similarity can be used to recognise whether other people are related. A person's face is a more accurate cue of relatedness than other physical features of people such as their body. Bodies can undergo drastic changes based on non-genetic factors that come from different lifestyles including weight and body composition changes, and although faces do change appearance with changes in weight (Kau, Zhurov, Bibb, Hunter, & Richmond, 2005), these changes are minor compared to those reflected by the body (Marre & Bestard, 2009). On the other hand, facial features are relatively stable throughout a person's life, unless a person changes their eye colour with contacts, undergoes cosmetic surgery, or sustains a major facial injury.

People can also identify kin through learning such as knowing who you grew up with and being told directly who relatives are, and can identify kin through phenotypic matching

sensory cues such as facial similarity and scent similarity (Mateo, 2015). Most support for kin selection comes from studies which assume that humans can recognise their kin through facial cues. However, this assumption has been criticised by many theorists such as Bellard, McAuliffe, Burch (2006), who argued that studies which use self-resembling faces may be tapping into different concepts such as self-bias and self-recognition. They suggested that the environment of ancestral humans, which lacked reflective surfaces besides still lakes, would not have given humans a sufficient understanding of what their own face looked like to the level that people do today due to mirrors and cameras. Thus if people recognise self-similar faces as related, then self-similar faces may be a valid image for what people believe to be themselves, and may thus also want to help them more than self-dissimilar faces.

Self-resemblance

To show that self-resembling faces are in fact a cue for kin recognition, DeBruine (2005) manipulated composite faces to resemble each participant. The participants rated the trustworthiness, short-term attractiveness, and long-term attractiveness of the manipulated composite faces and the original composite faces. It is expected that people will be more trusting of relatives, while at the same time finding them less sexually attractive, compared to non-relatives. If the participants were only showing a familiarity or a self-bias, then they would have rated the self-resembling faces as more trustworthy and attractive for both short and long relationships. Instead, the participants rated the self-resembling faces as more trustworthy than the original faces, less attractive for a short-term relationship than the original faces, and no different to the original faces for a long-term relationship. This effect was the same for both men and women. This means that the self-resembling facial cues were able to increase trust, while decreasing sexual attraction towards people who look related. Therefore, this study was able to successfully support the assumption that self-resembling faces mainly reflect kin recognition as opposed to a familiarity or self-bias.

Unconvinced, some researchers such as Alvergne, Faurie, and Raymond (2009), have proposed that instead of comparing oneself to others as a way of judging relatedness, people compare their close family to others as a way of judging relatedness. In spite of Alvergne, Faurie, and Raymond's (2009) argument, support for the self as a cue for kin selection was found by Bressan and Zucchi (2009) in their study titled, "Human kin recognition is self- rather than family-referential". The findings revealed that dizygotic and monozygotic twins were more inclined to help people in danger who looked similar to themselves, as opposed to people who looked more like their twin, and they were more likely to endorse people to be married into their family who looked like similar to themselves compared to people that looked more like their twin. These findings were similar for both dizygotic and monozygotic twins, because although monozygotic twins are commonly referred to as identical twins, there are actually minor differences in their appearance that occur from differences in weight and height which also affect their facial appearance (Kau, Zhurov, Bibb, Hunter, & Richmond, 2005), and from slight DNA changes throughout their lives (Bruder et al., 2008). In another study by Platek and Kemp (2009), fMRI was used to examine the neural substrates of perceiving the different faces. The findings revealed that the same neural substrates were active when looking at faces of family and faces of self-resembling composites, but different neural substrates were active when looking at faces of friends and strangers. These studies further support the validity of using self-resembling faces to elicit relatedness in people.

Self-resembling faces have also been shown to influence the decisions people make in public goods games. In Krupp, Debruine, & Barclay's (2008) variant of the game, each participant was given $10 at the start of each of the four rounds, and asked to anonymously donate a portion of it to a common goods pool. Each round the money in the common goods pool was doubled and shared equally between the four participants of a game. The goal was to have the most money by the end of the four rounds. At the end of each round, the amount

donated and each participant's total earnings was made public. Participants were then given the option to punish other players for free-riding. For every $1 spent in punishing, the punished would lose $3. Krupp, Debruine, & Barclay (2008) set up the game to have participants play against three AI players of whom up to two had their face morphed with the participant. The results showed that participants were more likely to donate money to the public good if one of the other players resembled themselves. Furthermore, the greater the number of players that resembled themselves in a game, the greater the amount of money donated to the public good. This study shows how facial resemblance is able to increase cooperation and reduce selfishness within the setting of a virtual public goods game.

There have been many studies which deal with positive relatedness in faces, which refers to faces that have been morphed to look similar to a participant. Consequently, Krupp, DeBruine, Jones, & Lalumière (2012) set out to test whether people can detect negative relatedness in faces, which refers to faces that have been morphed in the opposite direction to self-resembling faces, so that they look dissimilar to a participant. They examined trustworthiness and attraction, as positive relatedness studies have consistently shown that similar looking faces are trustworthy (DeBruine, 2005) but not sexually attractive (DeBruine, 2004; DeBruine, 2005; Penton-Voak, Perrett, & Peirce, 1999). Their results showed that participants had positive preferences for self-resembling faces and negative preferences for anti-self-resembling faces across contexts. They suggest that negative relatedness may be a cue facilitating spite in people who do not share the same genes. It follows that if people are more trusting of similar faces, then they may also be more helpful towards similar faces.

Paternity Uncertainty

Kin who are equally likely to share copies of the same genes with each other do not all receive equal amounts of help and investment. For example, people share twenty-five percent of their genes with their grandparents on their mother's side of the family, and

twenty-five percent of their genes with their grandparents on their father's side of the family, yet Euler & Weitzel (1996) found that the grandparents on the mother's side of the family provided significantly more solicitude than the grandparents on the father's side of the family. Additionally, grandmothers provided significantly more solicitude than grandfathers despite also having an equal probability that a gene may be shared with either relative. These findings may be explicable in terms of the theory of paternity uncertainty, in which men are uncertain of whether their partner's child is their own biological child or whether it is actually another man's child (Buss, 1996).

In a hypothetical situation, if mothers are one hundred percent sure whereas fathers are fifty percent sure that their child is their true biological offspring, then a father's mother would also be fifty percent sure that she is the child's biological grandmother, as she would be one hundred percent sure that her son would be her own offspring since she gave birth to him, but she would be just fifty percent sure that her son's child would be his offspring because the son did not give birth to that child. On the other hand, the father's father would only be twenty-five percent sure that he would be the child's biological grandfather, as he would only be fifty percent sure that his own son would be his own offspring as he did not give birth to him, and a further fifty percent less sure that his son's child would be the son's offspring because the son did not give birth to the child. Therefore, the maternal side of the family are more likely to provide help than relatives from the paternal side of the family. This effect has been found for many forms of investment, such as care (Danielsbacka, Tanskanen, Jokela, & Rotkirch, 2011; Euler & Weitzel, 1996), contact (Danielsbacka et al., 2011; Pollet, Nettle, & Nelissen, 2006), and closeness (Bishop, Meyer, Schmidt, & Gray, 2009; Michalski & Shackelford, 2005). This further supports the notion that people treat others differently based on how related they are perceived to be, not necessarily on how related they actually are.

Helping

Dawkins (1979) has suggested that kin selection should not refer to discriminating between kin and non-kin, because everyone would technically fall within the kin category to some degree. Instead kin selection should refer to discriminating people on a spectrum of genetic relatedness. Monozygotic twins would fall at 100% genetic relatedness, followed by dizygotic twins, sibling, and parents at 50% genetic relatedness, and so on until the level of genetic relatedness comes close to 0%. In line with this suggestion, Stewart-Williams (2007) asked undergraduate students how likely they would be to help either a sibling, cousin, acquaintance, or friend, in low, medium, and high cost requests. He found that across all types of helping, siblings were more likely to be helped than cousins, and cousins were more likely to be helped than acquaintances. He also found that as the cost of helping increased, the more related the person was to the participant, the more likely they would receive help. For low cost help friends were valued greater than siblings, for medium cost help friends and siblings were equally valued, and for high cost help siblings were valued more than friends. This suggests that people are not just preferentially helping others based on their degree of relatedness for kinds of helping, but are also taking the cost of helping into consideration.

Sex Differences in Helping

Sex may be another factor which influences helping based on cost. Andreoni and Vesterlund (2001) found that in dictator games, women were more likely than men to act altruistically when the cost was high, while men were more likely than women to act altruistically when the cost was low. There is conflicting evidence concluding that women are more altruistic than men (Fabes & Eisenberg, 1998; Piper & Schnepf, 2008), that men are more altruistic than women (Benenson et al., 2009; Johnson, 1996), or that there is no sex difference in altruism (Nowell & Tinkler, 1994). However, in accordance with the study by

Andreoni and Vesterlund (2001), the content and context of the altruistic action may determine which sex is more likely to help others.

In investment games, Chaudhuri and Gangadharn (2003) discovered that men were more trusting than women by giving more money during a first encounter, yet women were more reciprocal by giving more money back. A possible reason for this is that women are more risk-aversive (Eckel & Grossman, 2008). Chaudhuri and Gangadharn's (2003) result has been replicated cross-culturally by Croson and Buchan (1999), who also found that women were more reciprocal by giving more money back. However, they also found no difference between men and women's trust, as both sexes gave equally during the first encounter. This effect was the seen across cultures. From recruiting a large sample of over ten-thousand, Baez et al., (2017) found that men were more generous than women when the player benefited by being generous, and in all other times women were more generous. Unlike the men, the women appeared to be displaying altruism by being generous to others even when it did not benefit them. In each of these studies, it may have been the case that the women were more focused building a good reputation, which could have meant more to them than winning the game. Regardless, due to the strict parameters of economic games, caution should be taken in extrapolating these results to naturalistic scenarios.

In naturalistic scenarios, women are shown to be more helpful then men in many contexts. Piper and Schnepf (2008) looked at charity records and found that regardless of what the charity was for, women gave more than men. They also found that married men gave more than single men, but still not as much as married women. Additionally, a meta-analysis of prosocial behaviours in naturalistic contexts has also shown that girls demonstrate more prosocial behaviour than boys, such as, comforting, sharing, and instrumental help (Fabes & Eisenberg, 1998). Furthermore, research by Schwartz and Rubel (2005) shows that cross-culturally, women prioritise benevolence more than men, which they define as caring

about the welfare of others they frequently encounter. These studies demonstrate that women are more helpful in long-term contexts that favour the gender role of females as nurturing and caring, compared to the gender role of males as heroic and chivalrous (Eagly & Crowley, 1986).

Eagly and Crowley (1986) conducted a meta-analysis on studies that focused on helping in the short-term context of heroic and chivalrous actions based on the male gender role, such as putting oneself at risk to help a stranger. The results showed that men were more likely to help than women in these scenarios, and women were more likely to be helped than men, however the difference between men and women offering help was only small. Conversely, Iredale, Van Vugt, & Dunbar (2008) do not attribute gender roles as the main influence behind men helping strangers more than women. Instead, they argued that in these settings, men are using helping as a mating strategy by showing off to others around them. Their experiment involved asking heterosexual men and women to donate money to charity either alone, with a same sex observer, or with an opposite sex observer. Females contributed the same percentage of their wage to charity regardless of whether they were alone, or with a same or opposite sex observer. Males on the other hand, contributed a significantly higher percentage of their wage to charity if they were with an opposite sex observer, compared to a same sex observer or if they were alone. This suggests that sex may influence kin selection studies of helping behaviour, so when comparing helping of kin to strangers, men may show a stronger inclination to help strangers than women, especially if surrounded by women.

Alternative Explanations of Altruism and Helping

People perform altruistic actions with a selfless regard for others. There are multiple perspectives explaining why people perform altruistic actions, which can be divided into proximate and ultimate explanations. Proximate explanations refer to how a behaviour occurs by determining the immediate variables and processes which contribute to its production,

while ultimate explanations refer to why a behaviour occurs by determining the way in which the behaviour has successfully been passed down from ancestor to ancestor (Scott-Phillips, Dickins, & West, 2011). Social-psychologists have focused on proximate explanations of altruism by attempting to understand what situational factors contribute to people going out of their way to help others, while evolutionary theorists have focused on ultimate explanations of altruism by attempting to understand why a behaviour which is considered to be immediately costly to an individual has been successful in being passed down. Both explanations are complementary to each other and are required to completely understand the behaviour.

Evolutionary Explanations of Altruism

Besides kin selection, evolutionary theorists have posed several other ultimate explanations of altruism, such as reciprocal altruism, and reputation preservation (Fehr & Fischbacher, 2003). However, these other explanations are not counted as true altruism by some theorists, because unlike kin selection, the individual themselves still benefits directly from these behaviours (Becker, 1976).

Reciprocal altruism refers to benefiting others at a cost to oneself with the expectation that others will return the benefit. In game theory, the most common use of reciprocal altruism is the strategy known as tit for tat, which is intended to reduce exploitation (Hamilton & Axelrod, 1981). Reciprocal altruism explains altruism in situations where people are likely to meet multiple times as opposed to one off encounters. Therefore, Trivers (1971) asserts that it would not be a viable behaviour around strangers, unless you know you will meet them again and consequently they will be able to repay you for your help. Contrary to this assertion, (Henrich & Fehr, 2003) argues that one off encounters with other humans were rare for our ancestors, which means people help others who they may never encounter again. This is supported by research in which people playing economic games help others at a

cost to themselves even when they believe they will never encounter those people again (Sally, 1995).

People are not only more likely to help others who have previously helped them, but are more likely to help other people who also help others. Indirect reciprocity refers to people rewarding those who help others with a good reputation and those who do not with a bad reputation. As seen in public good games (Brandt, Hauert, & Sigmund, 2003) and dictator games (Servátka, 2009), people with good reputations are rewarded, while those with bad reputations are punished. Haley & Fessler (2005) suggests that the reason people help others who they will never encounter again is not because reciprocal altruism is maladaptive in high population modern society, but because helping others preserves their good reputation. In their (2005) study, they found that even in economic games with only one-off encounters, people were more generous when given cues that signalled the presence of others. This highlights the importance that reputation has in motivating people to help others.

Social Psychological Explanations of Altruism and Helping

Social psychologists usually examine helping from the point of the individual as a voluntary attempt to help others (McAndrew, 2002). Their research has discovered many immediate factors and situations in which people are more likely to help others, such as the social and cultural norms, individual trait differences in helping and empathy, and the individual's emotional state. For example, people are more likely to participate in volunteering when they are in a positive mood (Schnall, Roper, & Fessler, 2010), and they feel happier after volunteering which increases their motivation to continue volunteering in future events (Meier & Stutzer, 2008). One of the most well-known proximate explanations of helping behaviour is the bystander effect, which refers to the inverse relationship between the amount of people present around someone in need of help and the likelihood of helping that person (Fatkin & Lansdown, 2017). The bystander effect has also been demonstrated in

the online environment. Brody and Vangelisti (2016) examined incidents of cyberbullying, finding that the number of bystanders and the closeness of the bystander and the victim were the most significant factors that contributed to whether the bystander intervened. This reveals that the mechanisms which contribute to whether a person helps someone may not require in person contact.

Culture and learning also contribute considerably to whether a person is likely to help others. During a person's upbringing, learning through the actions of others serves as a powerful tool that is used to know who, when, and the degree of assistance that is socially normal to be given to others (Robinson III & Curry, 2005). Bandura's (1977) Social Learning Theory, states that people learn vicariously through the actions of others. To examine this, an experiment by Simmons and Sands-Dudelczyk (1983) looked at the way children helped each other in three different types of preschools; a Transactional Analysis preschool in which discussing feelings is encouraged, a Montessori preschool which focuses on child directed learning, and a traditional preschool which aimed at preparing the children for adulthood. The likelihood of helping another child who lost their necklace was the same across all types of preschools. However, each preschool showed differences in the way help was given. The Transactional Analysis preschool children were more likely to comfort the distressed child, the Montessori preschool children were more likely to search for the missing necklace, and the traditional preschool children were more likely to recruit an adult to help. This shows how learning may not just serve as a moderator for helping but also a mediator in the way it is carried out. Furthermore, learning may also be a non-genetic reason as to why different ethnicities with different cultures and role models show different helping behaviours (Kumar, Calvo, Avendano, Sivaramakrishnan, & Berkman, 2012; Salamon & Sokolowski, 2001).

The current debate within social psychological explanations of helping is whether people help others with the true intention of helping other people or whether people help

others with the intension that it will actually benefit themselves. On one hand, the negative-state relief model people are driven to reduce negative moods by engaging in behaviours such as helping, which make them feel better about themselves (Cialdini & Kenrick, 1976). On the other hand, Batson's empathy-altruism hypothesis predicts that if people feel empathetic towards someone then they will help that person (Batson et al., 1989). Empathy according to Batson et al., (1989) referred to cognitive empathy, which is the ability to understand what others are feeling. In support of the empathy-altruism hypothesis, a meta-analysis by Eisenberg and Miller (1987) has shown a positive relationship between emotional empathy which refers to feeling another person's emotions and altruism, and a more recent meta-analysis by Jolliffe and Farrington (2004) has shown a negative relationship between both emotional and cognitive empathy and offending. This suggests that variation in altruism may be due to factors which influence empathy between people, known as state empathy, and an individual's own level of empathy, known as trait empathy.

The Concept of Empathy

Empathy has been postulated as one of the main contributors to helping behaviour, however what is meant by the word empathy has been highly contested since its first use. In 1909 the British psychologist Edward Bradford Titchener who worked under Wilhelm Wundt, created the word empathy to be an English translation of Einfühlung (Titchener, 1909), which was the German word at the time for having an understanding of another person or object by projecting oneself into them (Lipps, 1903). It wasn't until the 1950s that empathy became associated with emotions. The rise of client-centred therapy which emphasised an emotional understanding of clients led to a definition of empathy which was used to describe the ability to understand another's emotions (Wispé, 1986). The scales which were developed to measure empathy followed suit, by measuring the accuracy to which a person can understand what others are feeling (Rogers, 1957). Beginning in the

1970s, a division was made between understanding the emotions of others; which is known as cognitive empathy, and responding emotionally to others emotions; which is known as emotional empathy (Mehrabian & Epstein, 1972).

Support for separating cognitive and emotional empathy has come from recent evidence showing a double dissociation between the two. This means that the areas of the brain which are central to cognitive empathy are different to the areas of the brain which are central to emotional empathy, thus they can function independently. Shamay-Tsoory, Aharon-Peretz and Perry (2009) were the first to show a double dissociation between cognitive and emotional empathy using the lesion method in people with brain damage. They found that people with damage to their ventromedial prefrontal cortex scored low on cognitive empathy scales, but scored normally on emotional empathy scales. Conversely, people with damage to their inferior frontal gyrus scored low on emotional empathy scales, but scored normally on cognitive empathy scales. To assess cognitive empathy, the perspective taking and fantasy subscales of the Interpersonal Reactivity Index (Davis, 1983) were combined and measured. These scores were correlated more with the participants' performance on a facial emotion recognition task which measured cognitive empathy, than with their performance on a false-belief task which measured the similar concept; theory of mind. To assess emotional empathy, the empathetic concern and emotional distress subscales of the Interpersonal Reactivity Index (Davis, 1983) were combined and measured. However, unlike cognitive empathy, this measure was not validated against non-self-report measures.

In response to the limitation of Shamay-Tsoory, Aharon-Peretz and Perry's (2009) study, Nummenmaa, Hirvonen, Parkkola, & Hietanen (2008) examined the brain activation of healthy participants using fMRI while administering a task. The task produced emotional empathy by asking the participants to empathise with people expressing negative emotions in photographs. On the other hand, the task produced cognitive empathy by asking the

participants to empathise with people not expressing any emotions in photographs, which meant the participants would only be able to recognise the face in absence of feeling a certain emotion. The results revealed that compared to cognitive empathy, emotional empathy produced an increased activation of the limbic system which is involved in processing emotions, and the mirror neuron system in the premotor cortex which is involved in copying the movements of others. This supports the notion that there are two separate components of empathy; cognitive empathy which describes recognising emotions, and emotional empathy which describes feeling and reproducing the emotions of others as a response to them.

Therefore in this thesis, empathy will refer to the dual dimensional construct consisting of cognitive empathy and emotional empathy. Consequently, empathy will be defined as a set of abilities which are involved in understanding and responding to the emotions of others (Reniers, Corcoran, Drake, Shryane, & Völlm, 2011).

Empathy's Effect on Helping

A simple experiment was set up by Klimecki, Jusyte, Scheeff, & Schönenberg (2016) to test the effect of state empathy on the amount of money given in a dictator game. The participants watched either an empathic state inducing video which featured homeless children who were suffering, or a control video which featured people doing everyday activities such as walking. Then they were each given $10 and asked to give any amount they wish to an opposing player. The results showed that the participants who had just watched the empathy evoking video gave on average 70% more than the participants who had just watched the control video. The presentation of suffering people was chosen to evoke empathetic concern based on Batson, O'Quin, Fultz, Vanderplas, and Isen's (1983) research into emotional responses, which found that watching a nearby person suffering evoked feeling empathy towards that person. This would be considered affective empathy due to reacting to the emotion of person suffering. Therefore, Klimecki, Jusyte, Scheeff, &

Schönenberg (2016) have demonstrated that evoking emotional empathy is an effective strategy for increasing generosity. Naturalistic support for this finding comes from successful charity campaigns that utilise the presentation of suffering people, which leads to more donations (Bendapudi, Singh, & Bendapudi, 1996).

In another study, Edele, Dziobek, and Keller (2013) were interested in examining the effect of trait empathy on the amount of money given during a dictator game. They found that people high in emotional empathy were more likely to give an equal or greater amount to other players than to themselves, while people high in cognitive empathy were more likely to give less than an equal amount to other players than to themselves. The methods used to assess emotional empathy included the empathy subscales of the Interpersonal Reactivity Index and a test which asked participants how much they felt the emotions of others displayed in photos, while the methods used to assess cognitive empathy included the cognitive subscales of the Interpersonal Reactivity Index and a test which asked participants what emotion was being felt by others displayed in photos and videos. Edele, Dziobek, and Keller (2013) conclude that the mere ability of recognising unfortunate situations of others is not associated with altruistic decision making, but the ability to feel and act from others unfortunate situations can influence altruistic decision making positively. These results support earlier research by Davis (1983), which also found that trait emotional empathy but not cognitive empathy was associated with helping through volunteering to save a woman calling out in distress on a radio.

A study on cyberbullying by Ang and Goh (2010) examined the gender differences in trait empathy which contributes to this form of bullying. Their results reveal that boys cyberbully if they are low in cognitive and emotional empathy, while girls cyberbully only if they are low in emotional empathy as trait cognitive empathy did not affect girls' likelihood of cyberbullying. A girl's cognitive awareness of an action being morally right or wrong, did

not affect whether they would carry through with that action, unlike the boys in the study whose cyberbullying behaviour reflected whether they knew the action was morally right or wrong. Ang and Goh (2010) suggest that a girl's inclination to cyberbully may be associated more with reacting emotionally than cognitive decision making when compared to boys. In addition to this, they also found that both cognitive and emotional empathy was lower in boys than girls, and consequently boys were more likely to cyberbully compared to girls. These findings correspond to the gender difference in empathy found by Rueckert and Naybar (2008). In their (2008) experiment, women were rated significantly higher than men on the Mehrabian and Epstein Empathy Questionnaire. This suggests that empathy may be one of the main contributors to the sex difference in helping.

The Current Study

This study aimed to examine the effect of self-similarity on empathy and helping. There were three stages in this study:

Phase one

The aim of phase one was to photograph a neutral expression of people and record them expressing real sadness and happiness.

Stimuli manipulation and validation

The aim of the stimuli manipulation and validation stage was to create happy, sad, and neutral composite faces from the recordings in phase one, validate that those composites' facial expressions are each perceived as happy, sad, and neutral, and then manipulate them into similar and dissimilar versions of each participant.

Phase two

The aim of phase two was for phase one participants to complete an experiment in which they made judgements on self-similar and self-dissimilar faces. These judgements included how well they recognise the faces' emotional expression, how much they feel that emotion themselves, and whether they help people with a similar face or a dissimilar face in a low risk scenario. The following predictions were made:

Hypothesis 1A: Similar faces will be cognitively and emotionally empathised with more than dissimilar faces.

Hypothesis 1b: The difference in empathy for similar over dissimilar faces will be greater in emotional empathy than cognitive empathy.

Hypothesis 2: Similar faces will receive more help than dissimilar faces.

Hypothesis 3: A self-preference in cognitively and emotionally empathising will predict a self-preference in helping.

Phase One: Method

This project met all requirements of the Macquarie University Human Research Ethics Committee, and ethics approval was granted on the 20th April 2017 (see Appendix A).

Participants

The total number of people who participated in the first phase of the study was 52, but the data from 6 participants were excluded due to being of non-Caucasian ethnicity. The remaining 46 participants included 19 men who all identified as male (aged 18 to 27, *M* = 21.16, *SD* = 2.87), and 27 women who all identified as female (aged 18 to 29, *M* = 21.48, *SD* = 3.77). Of these, 80% were first year psychology students who were recruited through Macquarie University's SONA system, and rewarded with either half an hour of course credit or $10 AUD. The other 20% of participants included postgraduate psychology friends and non-academic friends of the researcher, who were approached directly by the researcher and rewarded with $10 AUD each.

Stimuli and Measures

Emotion Eliciting Videos

Six short videos (1 to 5 minutes each) were chosen to elicit an emotional response from participants: three happy, and three sad. Each video for happiness and sadness had a different theme to increase the chance of a participant relating to the content and expressing the desired emotion. The three happy videos featured a cat, a dog, and Mr Bean, while the three sad videos featured a breakup, a refugee, and a dog. All videos were downloaded from YouTube, except for the sad dog video which was a series of edited clips from Futurama's Jurassic Bark. Detailed descriptions and links to each video are provided in Appendix B.

Empathy Quotient Questionnaire (EQ-40)

The EQ-40 (Baron-Cohen, & Wheelwright, 2004) (see Appendix C) is a short and quick measure of empathy in adults with normal intelligence, which assesses cognitive and

emotional empathy together in one empathy scale, and can be used to evaluate the severity of autism spectrum disorder It was developed in response to measures such as the Interpersonal Reactivity Index (Davis, 1983) assessing empathy too broadly. The EQ-40 provided a quick way of knowing whether the participants are able to understand and respond appropriately to the stimuli used this study, and thus whether their responses can be considered valid.

A 40 item version of the original 60 item version was used which contains all the same 40 empathy measuring items, minus the 20 filler items. The purpose of the filler items is to distract the participant from knowing that the questionnaire intends to measure empathy, but since the participants already knew coming in to the experiment that the study would involve empathy, the filler questions were removed to save the participants' time and effort.

The scale within phase one participants had a Cronbach's alpha coefficient of .89 for men, and .87 for women, which are good levels of internal consistency according to DeVellis (2016).

Procedure

Over four months, each participant arrived separately at a university laboratory with the knowledge that the experiment would involve having their photos taken before and after watching some happy and sad videos. Before arriving each participant was reminded to wear no makeup, and to wear contacts if they require glasses to see. Once the participant arrived and signed a consent form, they were instructed to change their top to a plain grey singlet, and removal all facial jewellery. Additionally, to control for the effect that hair colour and style may have on empathy and helping, the participants were also instructed to keep their hair away from their face using a white stocking wave cap (Magic Collections: Hong Kong).

Each participant was seated at a table in a booth to control for background and lighting. The booth had walls painted Munsell N5 neutral grey, and was illuminated with 15 high accuracy d65 fluorescent Philips tubes in high frequency fixtures that reduced flicker

(see Figure 1.). Their photo was taken by a Canon 70D camera 3 meters away, with exposure (1/50 sec), ISO (200), F-stop (f/5.6) and custom white balance held constant. They were instructed to look straight on at the camera for a neutral before-shot. The researcher opened up a laptop's screen in front of the participant on the table so that it was aimed at the participants face. The laptop was secretly recording using its front camera. The camera light was not visible as an undetectable 1x1mm piece of black tape was placed on it blocking the light. The participant was instructed to complete an online experiment after the researcher leaves the room. However, before leaving the room, the researcher announced that the camera which was used to take their neutral photo was being turned off, and placed a cover over the camera lens, so that they didn't behave as if they were being recorded.

The first section of the online experiment asked for demographic information, including their student number or if not available their name, age, sex, gender, and country they have spent most of their life in. The next section of the online experiment played the six emotion eliciting videos, asking the participant after each video "how did this video make you feel?" on a slider scale which displayed very sad on the left at -10, neutral in the middle at 0, and very happy on the right at 10. To reduce order effects, half the participants saw the happy videos first, while the other half saw the sad videos first. The order of happy and sad videos within a block was also randomised. Instead of moving from the three videos of a particular emotion to the three videos of the opposite emotion, the participants were presented with the EQ-40 in the middle to reduce the intensity of the previously elicited emotion. The EQ-40 was also used to identify whether a participant may have significant difficulties in understanding and producing the appropriate emotions.

Once a participant had completed the online experiment, the researcher re-entered the room and asked the participant to look happy and look sad while taking their photos. To make it easier for the participant to fake their emotional expression, the emotion first called

upon was the emotion elicited from the last video they watched. However, only the real expressions of happiness and sadness which were secretly recorded became stimuli for phase two, as previous research (Hall, Gunnery, & Andrzejewski, 2011) has demonstrate that forced emotional expressions are recognised differently to real emotional expressions.

Figure 1. A photo of the booth with the internal lights on. This photo was taken with the camera in the same position it was used when taking photos of the participants' neutral and posed emotions.

Phase One: Results

An alpha level of .05 was set for all analyses.

Empathy Quotient

Assumptions

Results of a box and whisper plot showed one man's total EQ of 14 to be a possible outlier, but further testing using an interquartile range multiplier of 2.2 as recommend by Hoaglin and Iglewicz (1987), revealed this score which was at the 5th percentile of men's total EQ scores to not be an outlier. In accordance with Baron-Cohen and Wheelwright (2004), this score of 14 was within 2.5 SD below the population's mean for males (41.80), and may indicate difficulties in understanding emotions. However, their average cognitive empathy score of 4.83 (recorded in phase two) was higher than the sample's male mean of 3.23, which indicates that the participant can understand emotions. Additionally, the exclusion of his data did not significantly affect the results of this study, therefore it remained in all subsequent analyses.

Linearity for correlation between total EQ and happy and sad videos was confirmed from visual inspection of scatter plots.Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges. The assumption of homogeneity of variance was met for total EQ (Levene's $F(44) = 0.34$, $p = .57$), happy video (Levene's $F(44) = 0.16$, $p = .69$), sad videos (Levene's $F(44) = 0.92$, $p = .34$). Therefore, equal variance was confirmed.

Sex Difference in EQ

Between a possible minimum score of 0 and possible maximum score of 80, the total EQ score for men ranged between 14 and 62, while the total EQ score for women ranged between 31 and 70.

To test if there was a significant difference between men's total EQ scores and women's total EQ scores, an independent samples T-Test was carried out on the dependant variable 'total EQ scores' using a between-subject factor of sex (men, women). Women's total EQ scores (M = 51.70, SD = 10.64) were greater than men's total EQ scores (M = 43.16, SD = 11.37), and the results of the independent samples T-Test revealed that this difference of 8.55, 95% CI [1.94 to 15.15] was significant, $t(44) = 2.60$, $p = .012$, $d = 0.78$.

Relationship between EQ and the Ratings for the Happy and Sad Videos

To test if there was a significant correlation between men and women's total EQ scores and their ratings for the happy and sad videos, a Pearson product-moment correlation was carried out between total EQ scores and happy video ratings, and between total EQ scores and sad video ratings. For men, the relationship between total EQ (M = 43.16, SD = 11.37) did not significantly correlate ($r(17) = -.02$, $p = .94$) with their ratings for happy videos (M = 5.33, SD = 2.38), but did moderately correlate negatively with their ratings for sad videos (M = -5.25, SD = 1.56),which was statistically significant ($r(17) = -.59$, $p = .008$).

Women showed the same pattern, as the relationship between their total EQ (M = 51.70, SD = 10.64) did not significantly correlate ($r(25) = .19$, $p = .35$) with their ratings for happy videos (M = 5.63, SD = 2.11), but did moderately correlate negatively with their ratings for sad videos (M = -5.31, SD = 2.15), which was statistically significant ($r(25) = -.52$, $p = .005$).

Video Ratings

Between a minimum score of -10 and maximum score of 10, the mean of each happy video rating from men ranged between 4.79 and 5.79 (M = 5.33, SD = 2.38), while the mean of each happy video rating from women ranged between 5.59 and 5.70 (M = 5.62, SD = 2.11). Conversely, the mean of each sad video rating from men ranged between -3.42 and -6.53 (M = -5.25, SD = 0.36), while the mean of each sad video ratings from women ranged

between -3.59 and -6.89 (*M* = -5.31, *SD* = 0.41). For the ratings of each individual video compared together, refer to Figure 2.

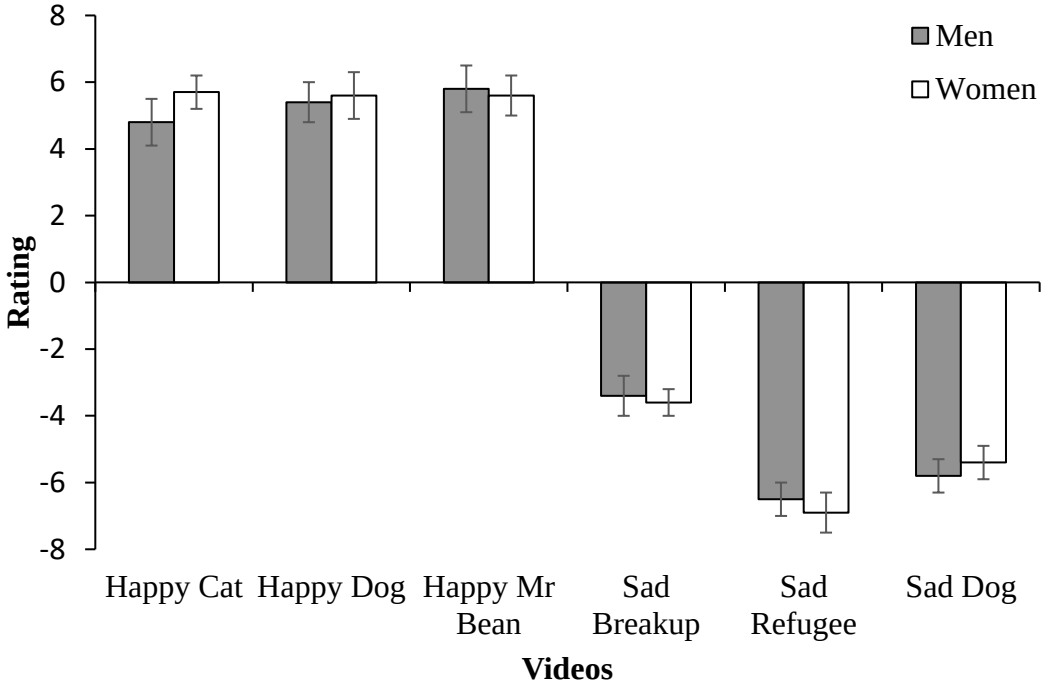


Figure 2. Mean video ratings for all happy and sad videos by men and women. From very sad (-10) to very happy (10). Error bars dictate one standard error around the mean.

Happy and Sad Video Ratings compared to a Neutral Rating of 0

To test if the happy and sad videos were respectively rated as happy and sad, one sample T-Tests were carried out between the happy and sad videos and a test value of 0. The participants rated the happy videos (*M* = 5.51, *SD* = 2.20) higher than a test value of 0, and the results of a one sample T-Test revealed that this difference of 5.51, 95% CI [4.85 to 6.15] was significant, $t(45) = 16.95$, $p < .001$, $d = 2.50$. The participants also rated the sad videos (*M* = -5.28, *SD* = 1.91) lower than a test value of 0, and the results of a one sample T-Test revealed that this difference of 5.28, 95% CI [4.72 to 5.85] was also significant, $t(45) = 18.76$, $p < .001$, $d = 2.76$.

Men compared to Women in the Ratings for the Happy and Sad Videos

To test if there was a significant difference between men and women's ratings of happy and videos, an independent samples T-Test was carried out on the dependant variables 'mean happy video rating' and 'mean sad video rating' using a between-subject factor of sex (men, women). Although women's mean happy video rating (M = 5.62, SD = 2.11) was higher than men's mean happy video rating (M = 5.33, SD = 2.38), the results of the independent samples T-Test revealed that this difference of 0.30, 95% CI [1.04 to 1.64] was not significant, $t(44) = 0.44$, $p = .66$. Similarly, although women's mean sad video rating (M = -5.31, SD = 0.41) was lower than men's mean sad video rating (M = -5.25, SD = 0.36), the results of the independent samples T-Test revealed that this difference of 0.06, 95% CI [-1.10 to 1.23] was not significant, $t(44) = 0.11$, $p = .91$.

Phase One: Discussion

Video Ratings

All six videos which were presented to participants to elicit a natural emotional response were successful. All happy videos were rated as making the participants feel moderately happy to very happy, and despite the sad breakup video being less intense than the other videos, all sad videos were also rated as making the participants feel moderately sad to very sad. The results of one sample T-Tests showed that happy and sad videos were each and on average significantly different from a rating of zero. This very large difference was greater than two and a half standard deviations. Men and women rated each video similarly as the results of a paired sample T-Test showed that there was no significant difference between men and women's ratings for happy and sad videos. Although women have been shown to express more emotion than men in similar situations (Kring & Gordon, 1998), especially in the case of sadness (Chaplin, 2015), the men in this study expressed similar intensities of emotion as that of women. This may have been due to the privacy the participant is given while watching the videos and believing that they are not being observed or recorded.

Empathy Quotient

The results of an independent samples T-Test revealed a significant medium sized sex difference, as women had a significantly higher EQ than men. This sex difference was also found by Baron-Cohen and Wheelwright (2004), and supports their assertion men and women's data should be analysed separately. The results of a Pearson product-moment correlation between men and women's total EQ scores and their ratings for the happy and sad videos, showed a significant large negative correlation for both men and women's EQ scores and their ratings for the sad videos, but not the happy videos. This means that people with low and high EQ are feeling happiness from the happy videos equally, but people with high EQ are feeling sadness from the sad videos considerably more intensely than the people with

low EQ. There are a couple of possible explanations for this, it may be that the EQ scale mainly measures traits associated with feeling sadness as opposed to feeling happiness, or it may be that since happiness is easier to feel than sadness (Gong, 2007), even people with a low ability to empathise can feel the happiness of others at an appropriate level similar to that of people with a high ability to empathise.

Methodological Considerations

To create realistic stimuli in the next stage of the study, the participants from phase one were required to have as much of their face showing as possible. This included keeping their hair out of their face with the white stocking wave cap which also controlled the effect that hair colour and style had on empathy and helping. However, the stimuli looked less representative of the general population, as most people wear their hair out instead of under a cap. Conversely, since face structure is a more stable indicator of relatedness than hair colour and type, this consideration would not significantly affect the experiment's external validity.

The participants were not allowed to wear glasses during the experiment because manipulating their face with the glasses may cause artefacts in the final image and look unrealistic, therefore participants were told to keep their glasses off for the duration of the experiment. This lead to some participants becoming became suspicious as to why they needed to remove their glasses if they were told no photos were being taken while they watched the videos. Fortunately few participants required glasses to see, and those that did, either had contacts with them or were not severely vision impaired so they could see the videos on the laptop without their glasses. It was important to make the participant feel they were not being recorded because observation from others, especially strangers has been shown to inhibit the intensity of facial expressions (Hess, Banse, & Kappas, 1995). Nevertheless the majority of participants were completely surprised to find out that their emotional reactions were recorded, and even those participants who suspected being recorded

still produced at least one instance of happiness and sadness which was required for the stimuli.

Conclusion

The results from phase one add confidence in confirming that the participants' emotional expressions were real in response to the emotion eliciting videos. Furthermore, in young adult Caucasian students, the videos have been validated for use in future studies as stimuli which produce moderate to very high intensities of happiness and sadness.

Stimuli Manipulation and Validation: Method

Participants

The total number of people who completed the stimuli validation experiment was 30. These included 14 men (aged 19 to 28, *M* = 24.14, *SD* = 2.63), and 16 women (aged 18 to 27, *M* = 24.34, *SD* = 3.23). All participants were recruited through ProlificAcademic.com, and each received $3.33 AUD for completing the 10 minute experiment. None of these participants were involved in the main study. All participants were Caucasian, and approximately half of the participants had spent most of their life in Australia, while the other half had spent most of their life in the UK. These demographics in terms of age (18-28) and ethnicity (Caucasian) matched the demographics of the participants in the main study to control for any differences that age and ethnicity may have on perceiving the emotions expressed in young adult Caucasian faces.

Stimuli and Measures

Facial Stimuli

The participants' emotional responses to happy and sad videos were recorded on a laptop's front facing camera in VGA (640x480 pixels), while they were seated in a booth (detailed previously in phase one's method). The laptop was positioned close-up which captured slightly below shoulder height to slightly above their head. Two research assistants who were blind to the study's aims scanned through each participant's recording in the video editing program Vegas Pro 12 and saved still frames of the participant's expressing happiness and sadness as image files. They were instructed to choose the best happy frame and the best sad frame for each participant according to how upright, straight, and close up the participant's head was. Half the participants were randomly allocated to each research assistant, but shared their input with each other when choosing.

While a participant was seated in the booth, their neutral and forced emotional expressions were also captured three meters away on a Canon 70D camera at a resolution of 3648x5472. These photos were cropped to a resolution of 1028 x768 which displayed slightly below shoulder height to slightly above their head. Any loose strands of hair were removed digitally using Photoshop CC.

All images were delineated individually in the computer program Psychomorph to create a two-dimensional vector mask of each face. Psychomorph has been established as a reliable tool for facial transformations (Tiddeman, Stirrat, & Perrett, 2005). The program provides an approximate vector mask which overlays the face. Delineation involved manually adjusting each of the 179 fiducial data points which make up the vector mask to approximately define the shape of the face.

The images of 15 men and 15 women who participated in phase one were selected to be averaged into happy, sad, and neutral composites, so that the participants would not recognise themselves or people they know when completing the final experiment in phase two. The preference for selection included an upright and straight head, minimal loose hair, high image quality, close up providing as much detail as possible, and minimal facial hair (men) as this can produce composites with faded and unrealistic facial hair. Three people of the same sex were averaged into a composite using Psychomorph (see Figure 3), which created 10 happy (5 male, 5 female), 10 sad (5 male, 5 female), and 10 neutral composites (5 male, 5 female). To create a range of different and realistic looking phenotypes, the light and dark haired people were averaged separately from each other, except for one mixed composite because there were not equal amounts of each phenotype. Per sex, this produced three light haired composites, nine dark haired composites, and three mixed hair colour composites. All the composite faces were aligned at the eyes to a straight face template to ensure each face was positioned straight in the centre of each image. A dark circular boarder

was applied to each completed image. For all composites sorted by facial expression and sex, see Appendix D.

Figure 3. The faces of three participants from phase one displaying a neutral expression (a) are averaged into a neutral expression composite face (b).

Empathy Quotient Questionnaire (EQ-40)

The purpose of administering the EQ-40 (see Appendix C) in the stimuli validation questionnaire are for the same reasons outlined in phase one's method.

The scale within phase one participants had a Cronbach's alpha coefficient of .87 for men, and .88 for women, which are good levels of internal consistency according to DeVellis (2016).

Procedure

An online experiment was used to validate the emotional expressions of the composites. The participants accessed the online experiment from a pool of paid experiments on ProlificAcademic.com. The first page required the participants to provide consent before continuing. The following page instructed the participants to enter demographic information, including their age, sex, gender, and country they have spent the most time in. The experiment included two sections.

The first section randomly presented the 10 happy, 10 sad, and 10 neutral composites which were transformed into the similar and dissimilar facial stimuli presented to the participants in phase two of the main study. For each face, the participants were asked "what expression is shown" on a slider scale which displayed very sad on the left at -10, neutral in the middle at 0, and very happy on the right at 10 (see Figure 4).

The second section asked the participants the complete the EQ-40, which was used to identify whether a participant may have significant difficulties in understanding the emotional expressions of the faces in the experiment. At the end of the experiment, the participants were debriefed on the purpose of the experiment, and why the EQ-40 was administered.

Figure 4. A screenshot of rating the emotional expression of a neutral female composite on a slider with very sad on the left, very happy on the right, and neutral in the middle.

Stimuli Manipulation and Validation: Results

An alpha level of .05 was set for all analyses.

Empathy Quotient

Assumptions

Results of a box and whisper plot showed one women's total EQ of 14 to be a possible outlier, but further testing using an interquartile range multiplier of 2.2 as recommend by Hoaglin and Iglewicz (1987), revealed this score which was at the 5th percentile of women's total EQ scores to not be an outlier. In accordance with Baron-Cohen and Wheelwright (2004), this score of 14 was greater than 3 SD below the population's mean for females (47.20), and may indicate difficulties in understanding emotions. However, this woman did not provide extreme scores for any other variable, and the exclusion of her data did not significantly affect the results of this study, therefore it remained in all subsequent analyses.

Linearity for correlation between total EQ and happiness and sadness ratings was confirmed from visual inspection of scatter plots.Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.The assumption of homogeneity of variance was met (Levene's $F(1,28) = 0.15$, $p = .70$). Therefore, equal variance was confirmed.

Sex Difference in EQ

Between a minimum score of 0 and maximum score of 80, the total EQ score for men ranged between 22 and 62, while the total EQ score for women ranged between 14 and 63.

To test if there was a significant difference between men's total EQ scores and women's total EQ scores, an independent samples T-Test was carried out on the dependant variable 'total EQ scores' using a between-subject factor of sex (men, women). Although women's total EQ scores ($M = 35.00$, $SD = 11.05$) were greater than men's total EQ scores

(*M* = 43.31, *SD* = 12.33), the results of the independent samples T-Test revealed that this difference of 8.30, 95% CI [-17.12 to 0.50] was close but not significant, $t(28) = 1.93$, $p = .06$.

Relationship between EQ and the Ratings for the Happy and Sad Faces

To test if there was a significant correlation between men and women's total EQ scores and their ratings for the happy and sad faces, a Pearson product-moment correlation was carried out between total EQ scores and happy face ratings, and between total EQ scores and sad face ratings.

For women, the relationship between their total EQ (*M* = 43.31, *SD* = 12.33) did not significantly correlate with either their ratings for happy faces (*M* = 5.16, *SD* = 1.60, $r(14) = .26$, $p = .28$) or sad faces (*M* = -3.49, *SD* = 1.20, $r(14) = -.13$, $p = .13$).

Similarly for men, the relationship between total EQ (*M* = 35.00, *SD* = 11.04) did not significantly correlate ($r(12) = -.29$, $p = .28$) with their ratings for happy faces (*M* = 4.51, *SD* = 0.99), but unlike women, men's total EQ moderately correlated negatively with their ratings for sad faces (*M* = -3.34, *SD* = 1.13), and was statistically significant ($r(12) = -.59$, $p = .03$.). Further inspection by splitting sad faces in to male sad faces and female sad faces revealed a significant large negative correlation between men's total EQ and male sad faces (*M* = -3.56, *SD* = 1.32, $r(12) = -.60$, $p = .02$), but not female sad faces (*M* = -3.11, *SD* = 1.36).

Emotional Expression Ratings

Between a minimum score of -10 and maximum score of 10, the mean happy face rating for men ranged between 2.40 and 6.30, while the happy face rating for women ranged between 3.10 and 8.20. Conversely, between a minimum score of -10 and maximum rating of 10, the sad face rating for men ranged between -5.90 and -1.90, while the sad face rating for women ranged between -5.50 and -1.50. Additionally, between a minimum score of -10 and maximum rating of 10, the neutral face rating for men ranged between -0.70 and 1.20, while

the neutral face rating for women ranged between -1.70 and 2.20. For the men and women's emotional expression ratings of the happy, sad, and neutral faces compared together, refer to Figure 5.

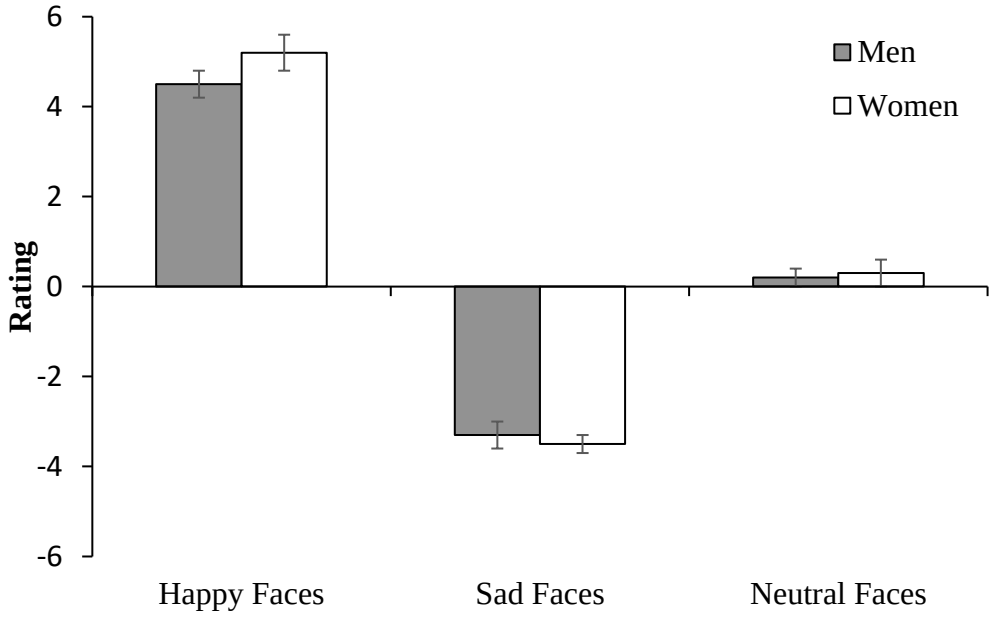


Figure 5. Mean emotional expression ratings for happy, sad, and neutral faces by men and women. From very sad (-10) to very happy (10). Error bars dictate one standard error around the mean.

Happy, Sad, and Neutral Face Ratings for all Composites compared to a Neutral Rating of 0

To test if the happy, sad, and neutral faces were respectively rated as happy, sad, and neutral, one sample T-Tests were carried out between the happy, sad, and neutral faces and a test value of 0. The participants rated the happy faces (M = 4.86, SD = 1.37) higher than a test value of 0, and the results of a one sample T-Test revealed that this difference of 4.86, 95% CI [4.35 to 5.37] was significant, $t(29) = 19.45$, $p < .001$, $d = 3.55$. The participants also rated the sad faces (M = -3.42, SD = 1.15) lower than a test value of 0, and the results of a one

sample T-Test revealed that this difference of 3.42, 95% CI [-3.85 to -2.99] was also significant, $t(29) = -16.27$, $p = < .001$, $d = 2.97$. Additionally, the participants rated the neutral faces ($M = 0.27$, $SD = 0.71$) slightly higher than a test value of 0, but the results of a one sample T-Test revealed that this difference of 0.27, 95% CI [0.00 to 0.53] was not significant, $t(29) = 2.05$, $p = .05$.

Similarly, the data analysed split by the sex of the participants also found a significant difference between the ratings for happy faces (Men $M = 4.51$, $SD = 0.99$; Women; $M = 5.16$, $SD = 1.60$) and a test value of 0 (Men $t(13) = 17.11$, $p < .001$, $d = 4.56$; Women $t(15) = 12.90$, $p < .001$, $d = 3.22$), and between the ratings for sad faces (Men $M = -3.34$, $SD = 1.13$; Women; $M = -3.49$, $SD = 1.20$) and a test value of 0 (Men $t(13) = 11.07$, $p < .001$, $d = 2.96$; Women $t(15) = 11.61$, $p < .001$, $d = 2.91$). Additionally, there was also no significant difference between the ratings of neutral faces (Men $M = 0.24$, $SD = 0.57$; Women; $M = 0.29$, $SD = 0.84$) and a test value of 0 (Men $t(13) = 1.55$, $p = .15$; Women $t(15) = 1.40$, $p = .18$). The same pattern was seen from analysing men and women's scores for the male and female composites separately, as shown in Table 1.

Table 1

One sample T-Tests comparing happy, sad, and neutral face ratings to a neutral rating of 0. Split by composite's sex, and participant's sex.

Results for:	Face rating		Effect size
	Mean	SD	*d*
Happy Male composites			
Men	4.66*	1.18	3.95
women	4.90*	1.61	3.05
Sad Female composites			
Men	-3.56*	1.32	2.70
Women	-3.83*	1.48	2.59
Neutral Male composites			
Men	0.24	0.66	-
women	0.31	0.98	-
Happy Female composites			
Men	4.36*	1.10	3.96
Women	5.43*	1.73	3.13
Sad Female composites			
Men	-3.11*	1.36	2.29
Women	-3.16*	1.18	2.68
Neutral Female composites			
Men	0.23	0.57	-
Women	0.28	0.94	-

Note. * $p < .001$ (2-tailed).

Effect size computed only for significant differences.

Stimuli Manipulation and Validation: Discussion

The aim of the stimuli manipulation and validation stage was to create happy, sad, and neutral composite faces from the recordings in phase one, validate that those composites' facial expressions are each perceived as happy, sad, and neutral, and then manipulate them into similar and dissimilar versions of each participant.

Emotional Expression Ratings

One Sample T-Test results on the stimuli validation experiment confirms that the happy, sad, and neutral composite faces were correctly rated as happy, sad, and neutral. The happy faces were rated as moderately happy which was significantly higher than a neutral rating of zero, and the sad faces were rated as moderately sad which was significantly lower than a neutral rating of zero, while the neutral faces were not rated significantly different to a neutral rating of zero. The very large difference between ratings for happy faces and a neutral rating of zero was greater than three and half standard deviations, while the very large difference between sad faces and a neutral rating of zero was greater than two and half standard deviations. These results were found for all composites combined, for male and female composites separately, by both sexes combined, and by males and females separately. Therefore based on their ratings, no faces were excluded as stimuli for phase two.

Empathy Quotient

The results of an independent samples T-Test revealed that for the participants involved in validating the stimuli, women's EQ was higher than men's EQ but the difference was not significant. This non-significant sex difference indicates that women from ProlificAcedemic.com empathised similar to men, unlike the students from phase one who demonstrated a significant sex difference in EQ. Additionally, both sexes from ProlificAcedmic.com scored lower in EQ compared to the student sample used in phase one. The results of a Pearson product-moment correlation between EQ and the ratings of happy

faces and sad faces was also not significant, except for men's ratings of sad faces, which revealed a significant negative moderate correlation between their EQ and their ratings of sadness. Further inspection showed that there was a significant negative moderate correlation between men's EQ and their ratings for sad male composites, but not for their ratings for sad female composites, and that men's ratings for sad male composites were more intense than their ratings for happy female composites. This means that as men's EQ increased, their ratings for the sad male faces become sadder, while their ratings for the sad female faces remained the same. A possible reason for why men were empathising considerably more with other men than with women may be because the men had a greater sense of self-resemblance when seeing the male composites than they did with the female composites. Additionally, previous research has demonstrated that the more familiar a person is believed to be, the easier it is to empathise with them (Krebs, 1975)

Methodological Considerations

Although the camera for the neutral shots could take photos at ultra-high definition (resolution of 3648x5472 pixels), the front facing laptop camera could only capture in VGA standard definition (resolution of 640x480 pixels). This discrepancy in resolution did not affect the experiment as the high resolution neutral shots taken with the camera were scaled down to 640x480 to fit into the display area of the online experiment.

The happy and sad images of the 15 men and 15 women selected to be averaged into composites had strict selection guidelines to pick the highest quality frame with the most straight and upright face, but for some participants who were constantly moving or had their head in an odd positon while expressing happiness or sadness, a less than ideal facial expression frames had to be picked. For faces which were tilted left or right, this was improved by aligning the face in Psychomoprh with a straight average face, but faces tilted up and down could be improved through an alignment. This only affected some of the happy

and sad composites, but had no effect on their ability to recognise happiness and sadness or their expression's intensity, according to the result of the validation experiment.

Conclusion

The results from the stimuli validation experiment have successfully validated the stimuli as correctly displaying 5 moderately happy, 5 moderately sad, 5 neutral expression male composites, and 5 moderately happy, 5 moderately sad, and 5 neutral expression female composites.

Phase Two: Method

Participants

The participant retention rate between the two phases was 71.74%, with 33 of the 46 phase one participants completing phase two. These included 15 men (aged 18 to 27, M = 21.27, SD = 3.03), and 18 women (aged 18 to 29, M = 22.11, SD = 3.80). 75.76% of these participants were the first year psychology students who were followed up about participating in the second phase through the email address and phone number they gave in the first phase. The other 24.24% of participants were the postgraduate psychology friends and non-academic friends of the researcher who were followed up about participating in the second phase through Facebook messenger. Each participant had the option of additional $10 AUD for their participation in phase two. All participants had spent most of their life in Australia, except for one who spent most of their life in the UK, one in Egypt, and one in Brazil.

Post Hoc Statistical Power Analysis

A post hoc power analysis was conducted using the software package, GPower (Faul and Erdfelder, 1992). The sample size of 33 was used with an alpha level of $p < .05$.

For paired sample T-Tests, the post hoc power analyses revealed the statistical power for this study was .20 for detecting a small effect (d = 0.2), whereas the power exceeded .80 for the detection of a moderate (d = 0.5) to large effect (d = 0.8). Thus, there was adequate power (power = .80) to detect moderate to large effects, but less than adequate statistical power to detect small effects.

For multiple regression with two predictors, the post hoc power analyses revealed the statistical power for this study was .20 for detecting a small effect (f^2 = 0.2), .58 for the detecting moderate effects (f^2 = 0.5), and .91 for detecting large effects (f^2 = 0.8). Thus, there was adequate power (power = .80) to detect large effects, but less than adequate statistical power to detect small and moderate effects.

Stimuli and Measures

Facial Stimuli

Similar and dissimilar versions of each of the 30 validated composites were created for each participant. Each composite was transformed by uploading into Psychomorph a participant's neutral face, an average face of the same sex, and the composite face that is going to be transformed. When transforming the average face, Psychomorph detects the differences between the participant's face and the average face, and applies these differences to the composite face in the desired direction. For example to transform the composite face to look more similar to a participant, a slider in Psychmorph is moved away from the average face towards the participant's face. The opposite is done to transform the composite face to look dissimilar to the participant, by moving the slider away from the participant's face towards the average face. For a visual representation of this process, refer to Figure 6. This was repeated for all participants for each of the 10 happy, 10 sad, and 10 neutral composites. Thus, for each participant, this produced 10 happy, 10 sad, and 10 neutral faces which looked similar them, and 10 happy, 10 sad, and 10 neutral faces which looked dissimilar to them.

All faces were aligned at the eyes to a straight face template to ensure each face was positioned straight in the centre of each image. Before the transformation, the dark circular boarder around each composite image was removed, but reapplied to each completed similar and dissimilar image after the transformation and alignment.

Only shape transformations were made, as opposed to also transforming across differences in colour and texture, because although some information about similarity is inferred from colour and texture, shape only transformations result in more realistic looking stimuli, especially if a participant has facial hair or scars. The level of change was set at 50% as this provided enough of a subtle difference in similarity and dissimilarity for each composite without looking too similar which can produce demand characteristics if

participants recognise themselves in the similar faces. Furthermore, previous research by Krupp et al., (2012) on perceiving positive and negative relatedness, supports the use of shape transformations of +50% to infer relatedness, and -50% to infer negative relatedness.

Figure 6. To transform one of the composite faces (d) to look similar to a participant's face (e), the composite face (d) is transformed on a continuum away from an average face of the same sex (c) and towards the participants face (e), to form one of the participant's similar composites (b).

To transform one of the composite faces (d) to look dissimilar to a participant's face (e), the composite face (d) is transformed on a continuum towards an average face of the same sex (c) and away from the participant's face (e), to form one the participant's dissimilar composites (a).

Empathy Quotient Questionnaire (EQ-40)

The purpose of administering the EQ-40 (see Appendix C) in the stimuli validation questionnaire are for the same reasons outlined in phase one's method.

Procedure

The participants accessed an online questionnaire that was sent to them one to four months after completing phase one. The first page informed the participants about the questionnaire and required them to provide consent before continuing, and asked the participants to complete the questionnaire in a quiet environment where they were free from distractions. The following page asked for their name or student number to ensure that they received the correct questionnaire. The questionnaire included two blocks.

The first block randomly presented 40 faces, one per page. Half the faces were the male and female composites that were transformed to look similar to the participant taking the questionnaire, while the other half were the composites that were transformed to look dissimilar to the participant taking the questionnaire. Thus, each participant sees faces that have been transformed to look similar to themselves, and faces that have been transformed to look dissimilar to themselves. For each face, the participants were asked "what expression is shown?" and "how does this make you feel?" on two separate slider scales, which displayed very sad on the left at -10, neutral in the middle at 0, and very happy on the right at 10 (see Figure 7.). Asking what expression is shown, measures cognitive empathy, as it is the ability to understand what others are feeling. Conversely, asking how the facial expression makes them feel, measures emotional empathy, as it is the ability to feel and reproduce the emotions of others.

The second block randomly presented 10 pairs of faces, and a scenario which forced the participants to choose to help one of two faces. Each pair of faces was presented side by side on a separate page. Each pair included one of the neutral composites that were

transformed to look similar to the participant, and one of the neutral composites that were transformed to look dissimilar to the participant. Similar and dissimilar faces were randomly presented as either on the left or the right. The scenario stated “In a future experiment the following people will be locked in a classroom for a whole day. Their objective is to leave the classroom as soon as possible. The more points they earn, the earlier they can leave. The answers you give now will help contribute to who will get to leave early. You will see 10 pairs of people, and will be required to choose who out of each pair you will give a point to and help leave early. You can only choose one person from each pair to help.” (see Figure 8.). The scenario was low risk but high value, and chosen due to most of the participants being busy students who understand time constraints to be extremely costly. At the end of the questionnaire, the participants were debriefed on the purpose of the study, and how their facial images from phase one were manipulated and incorporated into the questionnaire.

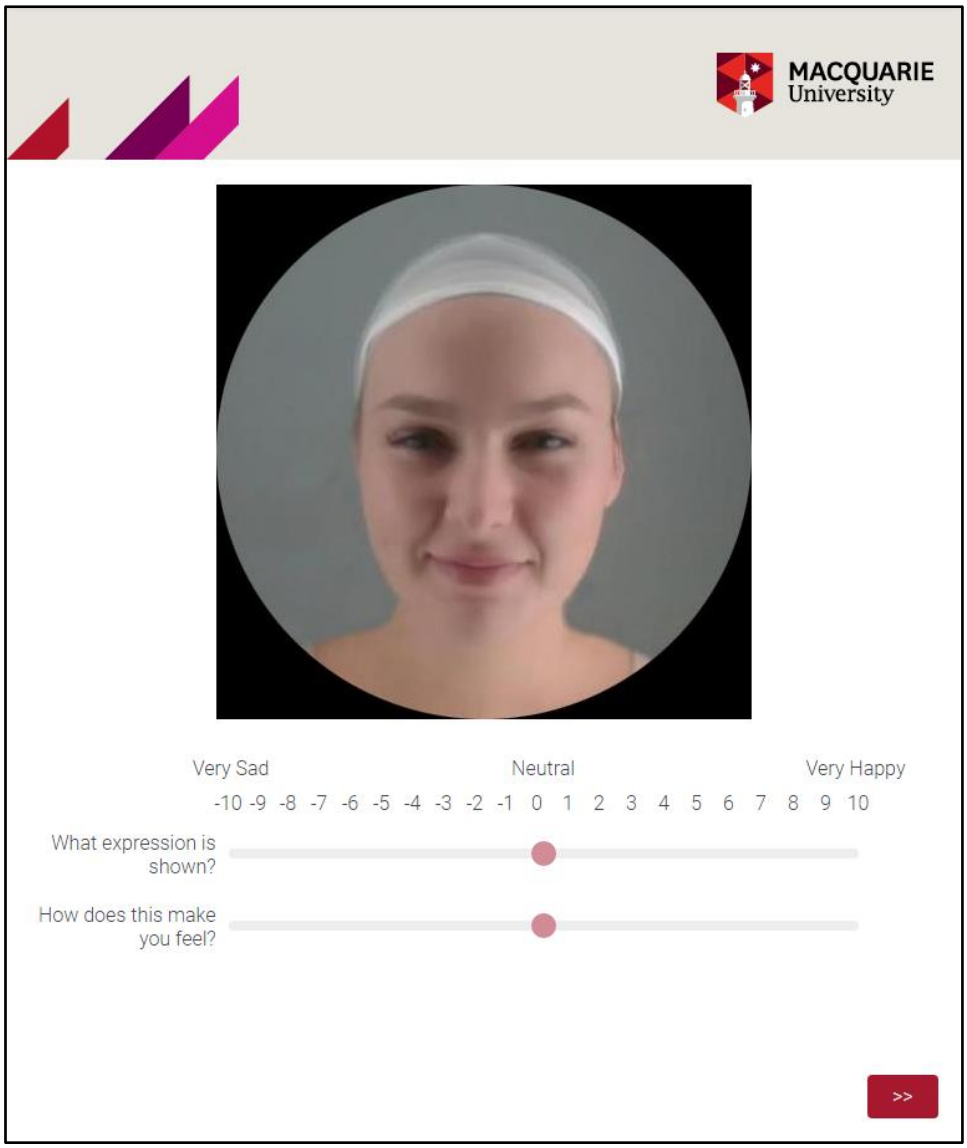


Figure 7. A question within the first block of the experiment which required rating happy and sad similar composites and happy and sad dissimilar composites on a slider between very sad and very happy for the expression and how the face makes the participant feel.

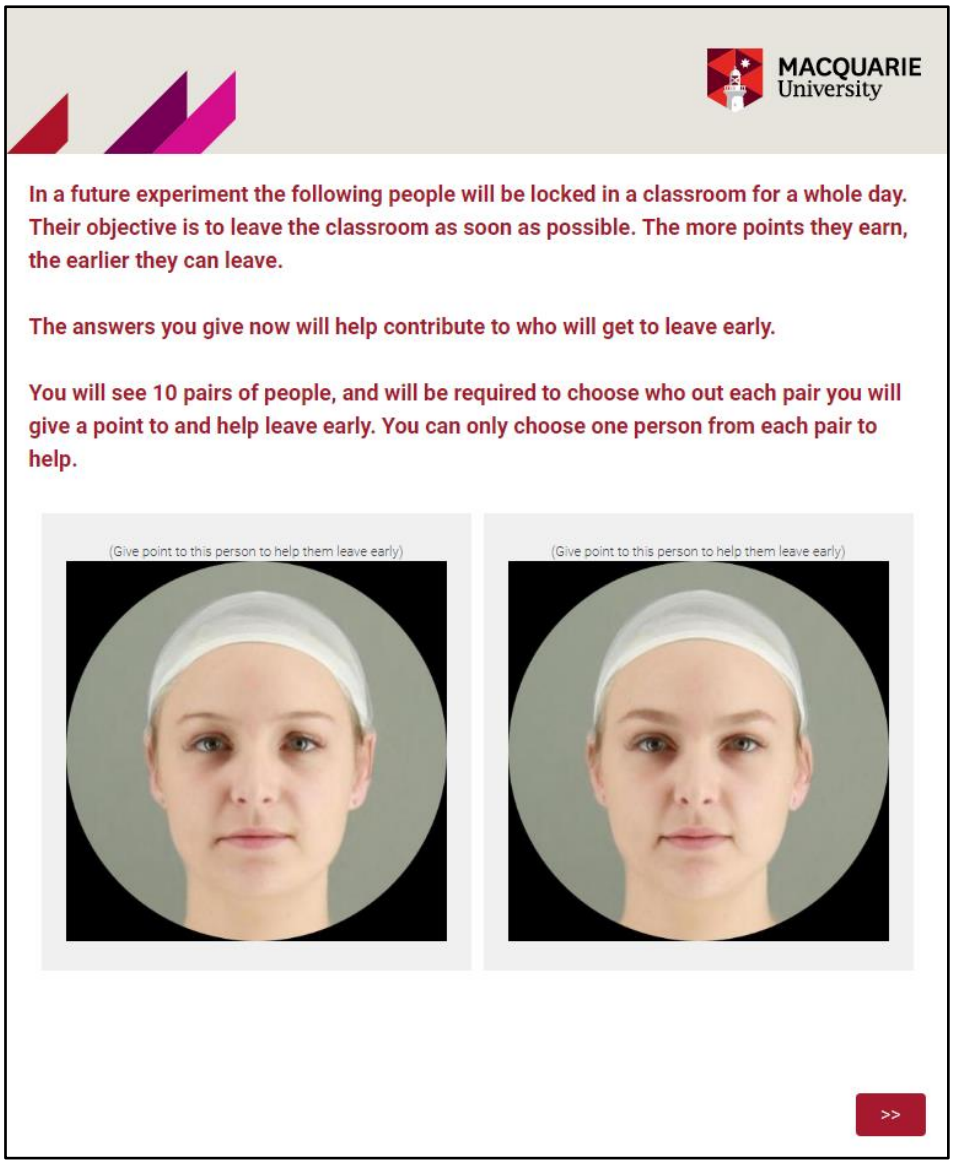


Figure 8. A question within the second block of the experiment which required choosing to help a similar or dissimilar neutral composite face according to a scenario.

Phase Two: Results

An alpha level of .05 was set for all analyses.

Empathy Quotient

Between a minimum score of 0 and maximum score of 80, the total EQ score for men in phase two ranged between 14 and 62, while the total EQ score for women in phase two ranged between 31 and 70.

Assumptions

The man identified in phase one who scored at the 5th percentile of men's total EQ scores also took part in phase two, and remained in all analyses for the reasons mentioned in phase one. No other possible outliers were detected.Linearity for correlation between total EQ and average cognitive and emotional empathy was confirmed from visual inspection of scatter plots.Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.The assumption of homogeneity of variance was met (Levene's $F(1,31) = 0.12$, $p = .73$). Therefore, equal variance was confirmed.

Sex Difference in EQ

To test if there was a significant difference between men's total EQ score and women's total EQ score, an independent samples T-Test was carried out on the dependant variable 'total EQ scores' using a between-subject factor of sex (men, women). Women's total EQ score ($M = 51.50$, $SD = 11.11$) was greater than men's total EQ score ($M = 42.93$, $SD = 12.37$), and the results of the independent samples T-Test revealed that this difference of 8.57, 95% CI [0.23 to 16.90] was significant, $t(31) = 2.10$, $p = .044$, $d = 0.73$.

Relationship between EQ and Cognitive and Emotional Empathy

To test if there was a significant correlation between men and women's total EQ scores and their average cognitive and emotional empathy scores, a Pearson product-moment correlation was carried out between total EQ score and average cognitive empathy score, and between total EQ score and average emotional empathy score.

For men, the relationship between total EQ (M = 42.93, SD = 12.37) did not significantly correlate ($r(13) = -.34$, $p = .22$) with their average cognitive empathy score (M = 3.23, SD = 1.02), or with their average emotional empathy score (M = 1.24, SD = 0.71), $r(13) = .26$, $p = .35$.

Women showed the same pattern, as the relationship between total EQ (M = 51.50, SD = 11.11) did not significantly correlate ($r(16) = .22$, $p = .38$) with their average cognitive empathy score (M = 3.76, SD = 1.17), or with their average emotional empathy score (M = 2.10, SD = 1.55), $r(16) = -.16$, $p = .52$.

Self-composite Bias

Out of the 33 participants in phase two, the faces of 22 of them were used in the creation of the average composite faces. To confirm that participants were not biased in their scores towards the composites that they was averaged in compared to the other composites, paired sample T-Tests were used to compare self-composite vs other-composite scores in cognitive and emotional empathy, and helping.

Assumptions

Visual inspection of box and whisker plots confirmed no outliers. Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

Self-composite compared to Other-composite in Cognitive Empathy

To test if there was a significant difference between the average cognitive empathy score for composites of the same sex that a participant was averaged in and the average cognitive empathy score for composites of the same sex that a participant was not averaged in, a paired samples T-Test was carried out on the average cognitive empathy scores for self-composites and the average cognitive empathy scores for other-composites. Although the average cognitive empathy score for self-composites (M = 3.86, SD = 1.87) was slightly less than the average cognitive empathy score for other-composites (M = 3.91, SD = 1.60), the results of the paired samples T-Test revealed that this difference of 0.45, 95% CI [-0.68 to 0.59] was not significant, $t(21) = 0.15$, $p = .88$.

Self-composite compared to Other-composite in Emotional Empathy

To test if there was a significant difference between the average emotional empathy score for composites of the same sex that a participant was averaged in and the average emotional empathy score for composites of the same sex that a participant was not averaged in, a paired samples T-Test was carried out on the average emotional empathy scores for self-composites and the average emotional empathy scores for other-composites. Although the average emotional empathy score for self-composites (M = 2.20, SD = 1.91) was slightly greater than the average emotional empathy score for other-composites (M = 1.99, SD = 1.41), the results of the paired samples T-Test revealed that this difference of 0.21, 95% CI [-0.51 to 0.93] was not significant, $t(21) = 0.61$, $p = .55$.

Self-composite compared to Other-composite in Helping

To test if there was a significant difference between the percentage of similar composites helped for composites of the same sex that a participant was averaged in and percentage of similar composites helped for composites of the same sex that a participant was not averaged in, a paired samples T-Test was carried out on the percentage of similar

composites helped for self-composites and the percentage of similar composites helped for other-composites. Although the percentage of similar composites helped for self-composites (M = 0.45, SD = 0.51) was slightly less than the percentage of similar composites helped for other-composites (M = 0.52, SD = 0.34), the results of the paired samples T-Test revealed that this difference of 0.068, 95% CI [-0.30 to 0.16] was not significant, $t(21) = 0.61$, $p = .55$.

Empathy

Tests for Hypothesis 1A

Similar faces were predicted to be cognitively and emotionally empathised with more than dissimilar faces. To test this prediction, paired-sample T-Tests compared the average cognitive and emotional empathy scores in similar and dissimilar faces. Average scores for cognitive and emotional empathy were each created by converting sad face scores into positive values, adding them to the happy face scores, and dividing by the number of faces.

Assumptions

Visual inspection of box and whisker plots confirmed no outliers.Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

Cognitive Empathy for all Similar Composites compared to all Dissimilar Composites

To test if there was a significant difference between the average cognitive empathy score for similar composites and the average cognitive empathy score for dissimilar composites, a paired samples T-Test was carried out on the average cognitive empathy scores for similar composites and the average cognitive empathy scores for dissimilar composites.

Although the average cognitive empathy score for similar composites (M = 3.52, SD = 1.31) was slightly greater than the average cognitive empathy score for dissimilar composites

(*M* = 3.51, *SD* = 1.15), the results of the paired samples T-Test revealed that this difference of 0.01, 95% CI [-0.36 to 0.38] was not significant, $t(32) = 0.07$, $p = .95$, as shown in Figure 9.

Similarly, the data analysed split by the sex of the participants did not find a significant difference between the average cognitive empathy score for similar composites (Men *M* = 3.09, *SD* = 1.12; Women; *M* = 3.88, *SD* = 1.33) and the average cognitive empathy score for dissimilar composites (Men *M* = 3.37, *SD* = 1.01; Women; *M* = 3.63, *SD* = 1.28) for either men ($t(14) = 1.23$, $p = .24$), or women ($t(17) = 0.96$, $p = .35$). There were also no significant results from analysing men and women's scores for the male and female composites separately, as shown in Table 2.

Emotional Empathy for all Similar Composites compared to all Dissimilar Composites

To test if there was a significant difference between the average emotional empathy score for similar composites and the average emotional empathy score for dissimilar composites, a paired samples T-Test was carried out on the average emotional empathy scores for similar composites and the average emotional empathy scores for dissimilar composites.

Although the average emotional empathy score for similar composites (*M* = 1.70, *SD* = 1.42) was slightly greater than the average emotional empathy score for dissimilar composites (*M* = 1.60, *SD* = 1.22), the results of the paired samples T-Test revealed that this difference of 0.97, 95% CI [-0.14 to 0.33] was not significant, $t(32) = 0.83$, $p = .41$.

Similarly, the data analysed split by the sex of participants did not find a significant difference between the average emotional empathy score for similar composites (Men *M* = 1.22, *SD* = 0.77; Women; *M* = 2.15, *SD* = 1.61) and the average emotional empathy score for dissimilar composites (Men *M* = 1.26, *SD* = 0.71; Women; *M* = 2.05, *SD* = 1.54) for either men ($t(14) = 0.44$, $p = .66$), or women ($t(17) = 0.72$, $p = .48$). There were also no significant

results from analysing men and women's scores for the male and female composites separately, as shown in Table 2.

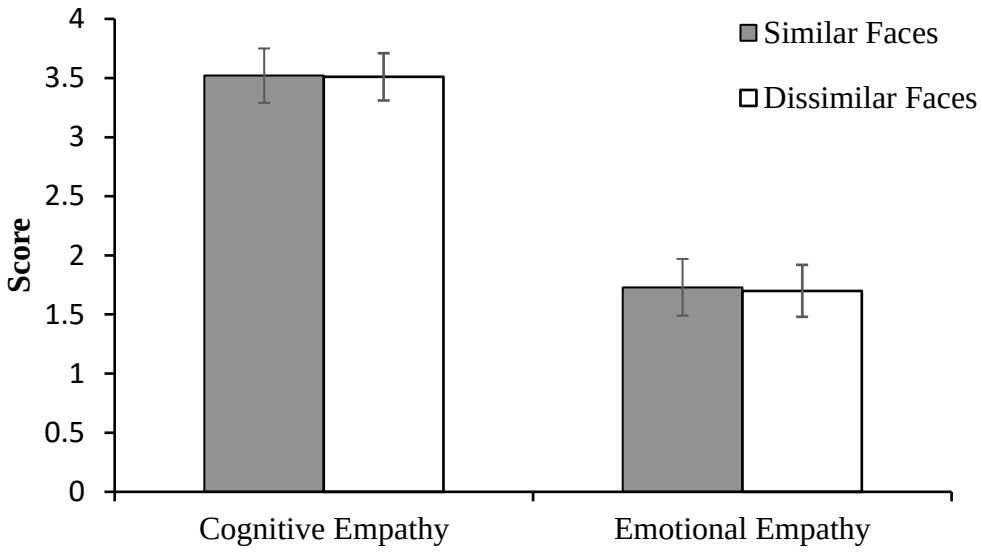


Figure 9. Mean cognitive and emotional empathy scores for similar and dissimilar faces. Values represent the intensity of the emotion recognised in the faces (cognitive empathy) or the intensity of those emotions felt by the participants (emotional Empathy). Error bars dictate one standard error around the mean.

Table 2

Paired sample T-Test results comparing cognitive and emotional empathy scores for similar compared to dissimilar composites. Split by composite's sex, and participant's sex.

Results for:	Similar composites		Dissimilar composites		Significance
	Mean	SD	Mean	SD	*p* =
Cognitive Empathy for					
Male composites					
All participants	3.49	1.26	3.43	1.11	.75
Men	3.07	0.99	3.29	0.97	.30
women	3.84	1.38	3.54	1.23	.30
Cognitive Empathy for					
Female composites					
All participants	3.56	1.53	3.59	1.27	.88
Men	3.12	1.60	3.45	1.14	.36
Women	3.92	1.41	3.71	1.40	.52
Emotional Empathy for					
Male composites					
All participants	1.70	1.42	1.60	1.22	.41
Men	1.15	0.86	1.11	0.67	.77
women	2.15	1.64	2.01	1.42	.43
Emotional Empathy for					
Female composites					
All participants	1.76	1.37	1.78	1.42	.85
Men	1.28	0.79	1.42	0.93	.40
Women	2.16	1.62	2.08	1.70	.70

Note. There were no significant results at the $p < .05$ level (2-tailed).

Tests for Hypothesis 1b

The difference in empathy for similar over dissimilar faces was predicted to be greater in emotional empathy than cognitive empathy. To test this prediction, paired sample T-Tests compared the self-preference for cognitive empathy and self-preference for emotional empathy. Self-preference scores for cognitive empathy and emotional empathy were each calculated by subtracting the dissimilar from the similar cognitive empathy scores, and subtracting the dissimilar from the similar emotional empathy scores.

Assumptions

Visual inspection of box and whisker plots confirmed no outliers. Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

Self-preference for Cognitive Empathy compared to Self-preference for Emotional Empathy

To test if there was a significant difference between the self-preference for cognitive empathy and the self-preference for emotional empathy, a paired samples T-Test was carried out on the similar minus dissimilar average cognitive empathy scores and the similar minus dissimilar average emotional empathy scores.

Although the self-preference for cognitive empathy (M = 0.01, SD = 1.05) was slightly greater than the self-preference for emotional empathy (M = 0.04, SD = 0.53), the results of the paired samples T-Test revealed that this difference of 0.02, 95% CI [-0.30 to 0.25] was not significant, $t(32) = 0.18$, $p = .86$, as shown in Figure 10.

Similarly, the data analysed split by the sex of participants did not find a significant difference between the self-preference for cognitive empathy (Men M = -0.28, SD = 0.88; Women; M = 0.26, SD = 1.13) and the self-preference for emotional empathy (Men M = -0.05, SD = 0.41; Women; M = 0.11, SD = 0.62) for either men ($t(14) = 1.25$, $p = .23$), or

women ($t(17) = 0.81$, $p = .43$). There were also no significant results from analysing men and women's scores for the male and female composites separately, as shown in Table 3.

Table 3

Paired sample T-Tests comparing self-preference for cognitive empathy and self-preference for emotional empathy. Split by composite's sex, and participant's sex.

Results for:	Cognitive Empathy		Emotional Empathy		Significance
	Mean	SD	Mean	SD	p =
Self-preference in Male composites					
All participants	0.06	1.06	0.10	0.67	.83
Men	-0.23	0.82	0.05	0.60	.28
women	0.30	1.13	0.14	0.74	.50
Self-preference in Female composites					
All participants	-0.04	1.37	-0.02	0.71	.94
Men	-0.33	1.35	-0.14	0.62	.46
Women	0.21	1.37	0.72	0.79	.48

Note. There were no significant results at the $p < .05$ level (2-tailed).

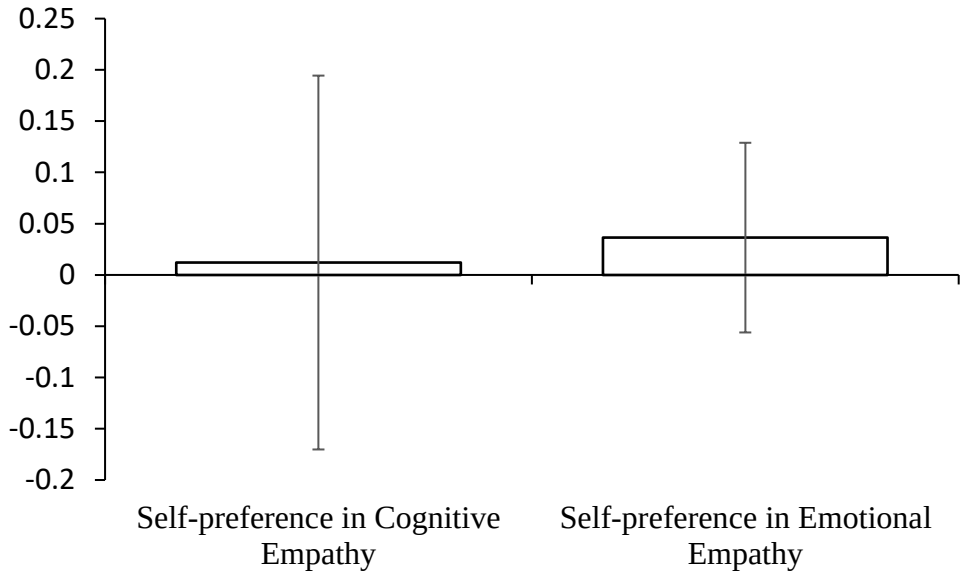


Figure 10. Self-preference scores in cognitive empathy and emotional empathy. For Self-preference in cognitive empathy, the values represent the intensity of the emotion recognised in the similar faces minus the dissimilar faces. For self-preference in emotional empathy, the values presentment the intensity of those emotions felt by the participants in the similar faces minus the dissimilar faces. Error bars dictate one standard error around the mean.

Helping

Testing Hypothesis 2

Similar faces were predicted to receive more help than dissimilar faces. To test this prediction, one sample T-Tests compared the percentage of help received by similar faces and 50% of the total amount of help received by all faces.

Assumptions

Visual inspection of box and whisker plots confirmed no outliers.

Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

Percentage of Help given to Similar Composites compared to 50% of the total amount of help given to all Composites.

To test if there was a significant difference between the percentage of help given to similar composites and 50% of the total amount of help given to all composites, a one sample T-Test was carried out between the percentage of help given to similar composites and a test value of 50.

Although the participants chose to help the similar composites (M = 50.61%, SD = 28.82) slightly more than 50%, the results of a one sample T-Test revealed that this difference of 0.61, 95% CI [-9.61 to 10.83] was not significant, $t(32) = 0.12$, $p = .91$.

Similarly, the data analysed split by the sex of the participants did not find a significant difference between the percentage of help given to similar composites (Men M = 46%, SD = 33.97; Women; M = 54.44%, SD = 24.06) and 50% of the total amount of help given to all composites for either men ($t(14) = 0.46$, $p = .66$), or women ($t(17) = 0.78$, $p = .44$). There were also no significant results from analysing men and women's scores for the male and female composites separately, as shown in Table 4.

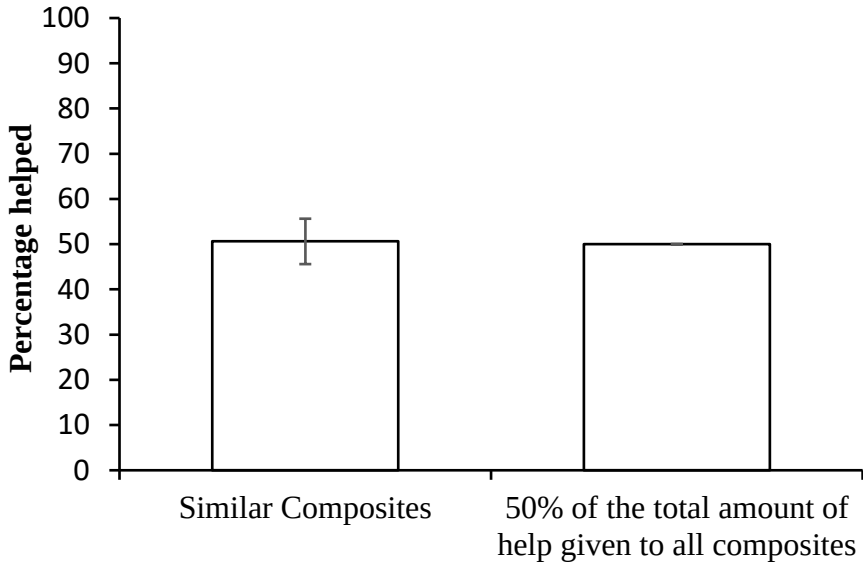


Figure 11. Percentage of help given to similar composites compared to 50% of the total amount of help given to all composites. Error bars dictate one standard error around the mean.

Table 4

One sample T-Tests comparing the percentage of help similar composites received to 50% of the total help received by all composites. Split by composite's sex, and participant's sex.

Results for:	Percentage of help given to similar composites		Significance
	Mean	SD	*p* =
Male composites			
All participants	49.09%	34.67	.88
Men	48.00%	35.29	.83
women	50.00%	35.15	1.00
Female composites			
All participants	52.12%	34.98	.73
Men	44.00%	40.85	.58
Women	58.89%	28.67	.21

Note. There were no significant results at the $p < .05$ level (2-tailed).

Testing Hypothesis 3

A self-preference in cognitively and emotionally empathising was expected to predict a self-preference in helping. To test this hypothesis, multiple linear regression was carried out to predict a self-preference in helping based on a self-preference in cognitive empathy and a self-preference in emotional empathy.

Assumptions

Visual inspection of box and whisker plots confirmed no outliers.

Linearity for correlation between cognitive and emotional empathy and was confirmed from visual inspection of scatter plots.

Multicollinearity was not evident as values for tolerance Variance Inflation Factor were close to 0 (Tolerance ≥ 0.36; VIF ≤ 3.45).

Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

Predicting the Percentage of Help given to Similar Composites from a Self-preference in Cognitive and Emotional Empathy

To test if the self-preference in cognitive empathy and the self-preference in emotional empathy predict the percentage of help given to similar composites, a multiple linear regression using the forced entry method was carried out to predict the dependant variable 'percentage of help given to similar composites' based on the similar minus dissimilar average cognitive empathy scores and the similar minus dissimilar average emotional empathy scores.

The results of this multiple linear regression revealed that the self-preference in cognitive empathy (B = –5.04 ± 6.99, β = -0.18) accounted for 13% of the variance in helping similar composites, and was not a significant predictor of helping similar composites (p = .48). Similarly, the self-preference in emotional empathy (B = – 5.04 ± 13.77, β = 0.01) accounted for only 0.5% of the variance in helping similar composites, and was also not a significant predictor of helping similar composites (p = .98). Consequently, the model (Percentage of help given to similar composites = 50.66 – 5.04 + 0.40) was not significant ($F(2, 30) = 0.49$, p = .62, R^2 = .03), as shown in Figure 12.

Similarly, the model split by sex of the participants also did not significantly predict helping behaviour for either men ($F(2, 12) = 0.72$, p = .51, R^2 = .11) or women ($F(2, 15)$ = 0.48, p = .63, R^2 = .06). For men, self-preference in cognitive empathy (Men: B = – -15.45 ±

13.01, β = -0.40; Women: B = –0.36 ± 8.02, β = -0.02) accounted for 32% of the variance in helping similar composites, while for women it accounted for 0.1% of the variance in helping similar composites. Neither were significant predictors of helping similar composites (Men: *p* = .26; Women: *p* = .097). For men, self-preference in emotional empathy (Men: B = 16.12 ± 28.16, β = 0.19; Women: B = -9.00 ± 14.67, β = -0.23) accounted for 16% of the variance in helping similar composites, while for women it accounted for 15% of the variance in helping similar composites. Neither were significant predictors of helping similar composites (Men: *p* = .58; Women: *p* = .55). There were also no significant results from analysing men and women's scores for the male and female composites separately, as shown in Table 5.

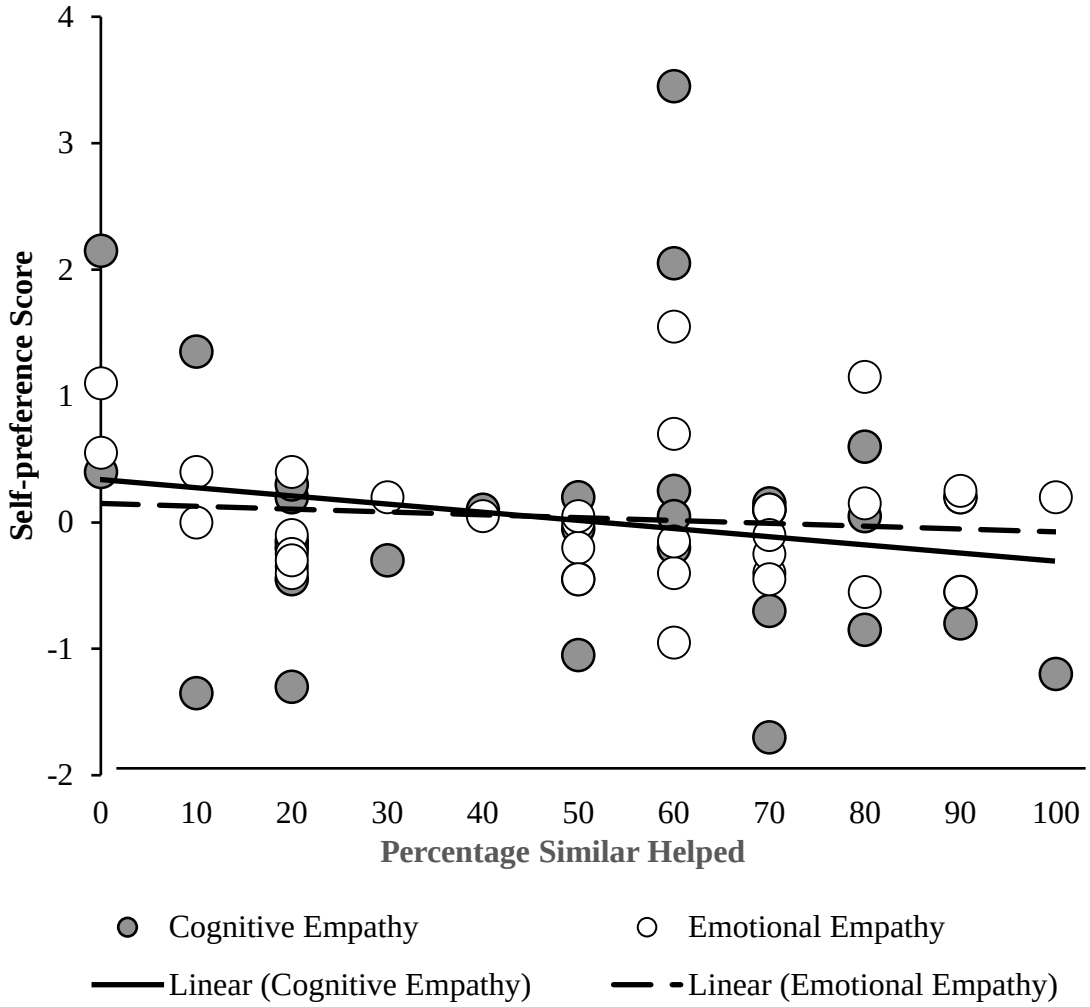


Figure 12. The relationship between the percentage of help given to similar composites and self-preference in cognitive and emotional empathy. For Self-preference in cognitive empathy, the values represent the intensity of the emotion recognised in the similar faces minus the dissimilar faces. For self-preference in emotional empathy, the values presentment the intensity of those emotions felt by the participants in the similar faces minus the dissimilar faces. Lines of best fit are included for each model.

Table 5

Multiple Regression predicting the percentage of help given to similar composites from a self-preference in cognitive and emotional empathy. Split by composite's sex, and participant's sex.

Results for:	Cognitive Empathy			Emotional Empathy				
	B	β	s-p r^2	B	β	s-p r^2	R^2	p =
Predicting help in male composites								
All participants	2.59 ± 6.53	.079	7%	3.64 ± 10.34	0.70	6%	.02	.79
Men	-4.53 ± 12.20	-0.11	10%	12.73 ± 16.63	0.22	22%	.05	.73
women	8.08 ± 8.85	0.28	23%	-5.59 ± 14.37	-0.12	1%	.05	.66
Predicting help in female composites								
All participants	-7.23 ± 7.35	-0.28	17%	-2.48 ± 14.35	-0.05	3%	.11	.19
Men	-12.69 ± 11.70	-0.42	30%	1.47 ± 25.33	0.02	2%	.16	.34
Women	-3.98 ± 9.23	-0.19	10%	-8.13 ± 16.07	0.22	12%	.16	.28

Note. There were no significant results at the $p < .05$ level (2-tailed).

Sex Differences in Empathy

Assumptions

Visual inspection of box and whisker plots confirmed no outliers. Normality was confirmed from visual inspection of histograms and Q-Q plots, and skew and kurtosis values within acceptable ranges.

The assumption of homogeneity of variance was met for all cognitive empathy variables (Levene's $F(31) = \leq 3.22$, $p \geq .08$), but not for any emotional empathy variables (Levene's $F(31) = \leq 10.65$, $p \geq .003$). Therefore, since the assumption of homogeneity of variance was violated for emotional empathy variables, their degrees of freedom were adjusted, and tests were used in which equal variances was not assumed.

Men compared to Women in Cognitive and Emotional Empathy among all Composites

To test if there was a significant difference between men's average cognitive empathy score and women's average cognitive empathy score, an independent samples T-Test was carried out on the dependant variables 'average cognitive empathy' and 'average emotional empathy' using a between-subject factor of sex (men, women).

Although women's average cognitive empathy score (M = 3.76, SD = 1.17) was greater than men's average cognitive empathy score (M = 3.23, SD = 1.02), the results of the independent samples T-Test revealed that this difference of 0.52, 95% CI [-1.31 to 0.27] was not significant, $t(31) = 1.35$, $p = .19$. Conversely, women's average emotional empathy score (M = 2.10, SD = 1.55) was greater than men's average emotional empathy score (M = 1.24, SD = 0.71), and the results of the independent samples T-Test revealed that this difference of 0.86, 95% CI [0.41 to -1.70] was significant, $t(25) = 2.11$, $p = .046$. For the results of male, female, happy, and sad composites analysed separately, refer to Table 6.

Table 6

Independent samples T-Tests comparing men to women in cognitive and emotional empathy for male composites, female composites, happy composites, and sad composites.

Results for:	Men		Women		Significance	Effect size
	Mean	SD	Mean	SD	*p* =	*d*
Average cognitive empathy						
Male composites	3.18	0.89	3.69	1.17	.17	-
Female composites	3.29	1.22	3.82	1.23	.22	-
Happy composites	3.60	1.54	4.02	1.67	.47	-
Sad composites	2.87	0.77	3.49	1.08	.07	-
Average emotional empathy						
Male composites	1.13	0.71	2.08	1.49	.02*	0.86
Female composites	1.35	0.80	2.12	1.61	.09	-
Happy composites	1.57	0.91	2.38	1.95	.13	-
Sad composites	2.87	0.77	3.49	1.08	.03*	0.67

Note. * Significant at the $p < .05$ level (2-tailed).

Effect size computed only for significant differences.

Phase Two: Discussion

The aim of phase two was for phase one participants to complete an experiment in which they made empathy and helping judgements on self-similar and self-dissimilar faces.

Hypotheses

Hypothesis 1A predicted that similar faces will be cognitively and emotionally empathised with more than dissimilar faces. However, the results of paired sample T-Tests have not confirmed this hypothesis, as there was no significant difference for either cognitive or emotional empathy in similar compared to dissimilar faces. This non-significant result was found for all composites combined, for male and female composites separately, by both sexes combined, and by males and females separately. This means that men and women were not only able to recognise the emotional expressions of faces that looked dissimilar to themselves equally as well as faces that looked similar to themselves, but that they also felt the emotions expressed by the faces that looked dissimilar to them as intensely as the faces that looked similar to them. This suggests that empathy towards others is not affected by facial self-similarity.

Hypothesis 1b predicted that the difference in empathy for similar over dissimilar faces will be greater in emotional empathy than cognitive empathy. However, the results of paired sample T-Tests have not confirmed this hypothesis, as there was no significant difference in cognitive compared to emotional empathy for the difference of similar over dissimilar faces. This non-significant result was found for all composites combined, for male and female composites separately, by both sexes combined, and by males and females separately. This means that men and women were not feeling the emotions expressed in the faces that looked similar to themselves more than how well they could recognise the emotion. This suggests that the notion of emotional empathy driving prosocial and cooperative

behaviour (Edele, Dziobek, & Keller, 2013; Klimecki, Jusyte, Scheeff, Schönenberg 2016) may not be accurate.

Hypothesis 2 predicted that similar faces will receive more help than dissimilar faces. However, the results of one sample T-Tests have not confirmed this hypothesis, as there was no significant difference in the help given to similar faces compared to 50% of the total amount of help given to all faces. This non-significant result was found for all composites combined, for male and female composites separately, by both sexes combined, and by males and females separately. This means that when men and women were choosing which person to help, they did not feel compelled to help people who looked similar to themselves, nor were they considering how similar the person's face was to their own in the decision making process. However, since empathy was not found to be greater for similar faces compared to dissimilar faces, the results for hypothesis two are consistent with the literature on empathy and helping, which suggests that empathy contributes to helping, and in its absence people are less likely to help (Eisenberg and Miller, 1987).

Hypothesis 3 predicted that self-preference in cognitively and emotionally empathising will predict a self-preference in helping. However, the results of a multiple linear regression has not confirmed this hypothesis, as neither a self-preference in cognitive empathy nor a self-preference in emotional empathy significantly predicted a self-preference in helping, as shown in Figure 12. This non-significant result was found for all composites combined, for male and female composites separately, by both sexes combined, and by males and females separately. This suggests that prosocial behaviour towards kin may be separate to prosocial behaviour motivated by empathy.

Self and other composites

The faces of 15 males and 15 females from phase one were averaged into composites. Of these, 22 were involved in phase two, so to ensure that these 22 participants were not biased

in their empathy and helping judgements towards the composites that their faces were averaged into, paired sample T-Tests were carried out to compare self-composite vs other-composite scores in cognitive empathy, emotional empathy, and helping. The results showed that that was no significant difference in cognitive empathy scores, emotional empathy scores, or helping between the composites that a participant was averaged in and the other composites of the same sex. Therefore, the participants were not biased in empathy and helping judgements towards the composites that they were a part of.

Sex differences in Empathy

The results of independent samples T-Tests revealed that women gave higher cognitive and emotional empathy scores than men, although only emotional empathy was significantly higher for women than men. Further inspection showed that men and women were only significantly different in emotional empathy felt for male composites not female composites. However despite a large effect size, caution should be taken when interpreting that result as significant, because it does not remain significant after conducting a Bonferroni correction, which may indicate a false-positive from running multiple T-Tests. Women were emotionally empathising the same as they were for female composites, but men were emotionally empathising much less with other male composites than they were with the female composites. This is in line with previous research that found in situations where men and women are in need of help, men have been shown to feel less sympathy towards other men compared to women (Eagly & Crowley, 1986).

The results of an independent samples T-Test revealed that there was no significant difference in the cognitive scores for happiness and sadness between men and women, nor was there a significant difference in the amount of emotional scores for happiness between men and women. However, there was a significant difference in the emotional sadness scores between men and women. Despite a medium effect size, caution should be taken when

interpreting that result as significant, because it does not remain significant after conducting a Bonferroni correction, which may indicate a false-positive from running multiple T-Tests. Women were emotionally empathising the same as they were for happy composites, but men were emotionally empathising significantly less with the sad composites than they were with the happy composites. This is consistent with an experiment by Han, Chen, and Zhang (2016) who also found that women were affected more by sad faces than men, whose emotional response to them appeared blunt in comparison to the women. Previous research has also found that men express less sadness than women (Chaplin, 2015).

Empathy Quotient

Even with only 72% of participants returning for phase two, EQ scores were relatively similar. Thus, just as in phase one, the results of an independent samples T-Test revealed a significant medium sized sex difference, as women had a significantly higher EQ than men. On the other hand, the results of a Pearson product-moment correlation between men and women's total EQ scores and their average cognitive and emotional empathy did not significantly correlate. In this study cognitive empathy was measured through recognising the emotional expressions of happy and sad faces, while emotional empathy was measured through feeling the emotion of those emotional expressions. The main difference is that the measures of cognitive and emotional empathy in this study focused only on the emotions that faces confer, while the EQ would measure a more general ability in understanding and reacting emotionally to others emotions that are not necessarily visual. This discrepancy may explain why the scales did not correlate.

Methodology Considerations

The colour profile for happy and sad images was slightly different to the colour profile of the neutral images. However, this should not have had an effect on the creation of the stimuli as only the shape of the faces were transformed not the colour or texture, and it should not have

had an effect on the results, as the happy and sad faces were presented in different blocks and not compared to each other. The happy and sad faces were used in the first block for judgments of empathy, while the neutral faces were used as force choice options for helping judgements in the second block.

Although the average composites were validated as all correctly expressing happy, sad, and neutral facial expressions, the final similar and dissimilar faces were not validated as looking similar and dissimilar to each participant. It would be infeasible to do such a task for each participant who had their own unique set of stimuli. However, confidence in the validity of the final stimuli remains strong as this study used the same transformation method previously used in research by Krupp, DeBruine, and Lalumière (2012) which found that people can perceive positive and negative relatedness.

General Discussion

The literature on kin selection in humans currently demonstrates that people can recognise kin visually through facial cues, and that people help kin more than non-kin. The results of this study add to the growing body of literature which counter this statement. In line with the finding that helping is not affected by self-resemblance supports a similar novel finding by Giang, Bell, and Buchner (2012) which found that facial resemblance does not enhance cooperation. They tested how self-resemblance affected cooperation during a social cooperation game where the players worked together to profit maximally by investing in each other or acting selfishly and try to cheat the opponent out of their money and keep their money for only themselves. The study used similar and dissimilar morphed faces as opponent players. The results showed that the participants did not cooperate significantly more with faces that were similar to themselves compared to faces that were dissimilar to themselves. Therefore, this supports the notion that human helping behaviour is not based on a preference for self-resembling faces.

Empathy may not be used in kin selection as the results of this study suggest that people do not cognitively or emotionally empathise with others who look similar to themselves more than they cognitively and emotionally empathise with others who look dissimilar to themselves. In fact, there was a small but non-significant trend in the opposite direction presented within the scatterplot of the multiple regression analyses, which meant that people who were empathising more with the similar faces were helping more dissimilar faces, but the people who were empathising more with the dissimilar faces were helping more similar faces. This suggests that helping based on empathy and helping based on self-resemblance are two competing strategies that a person may use when choosing who or whether to help someone.

Perceived trustworthiness has been demonstrated to be higher in similar faces compared with dissimilar faces (DeBruine, 2002; DeBruine, 2005; Sofer, Dotsch, Wigboldus, & Todorov, 2015), and even trustworthy people have been shown to be perceived as more physically similar to oneself (Farmer, McKay, & Tsakiris, 2014). The evolutionary benefit of trusting similar faces to oneself was thought to be connected to prosocial regard and lead to preferential helping behaviour as a form of kin selection. Although this study did not measure trust for similar and dissimilar faces, the results of this study suggest that people do not help others who look similar to themselves more than they help others who look dissimilar to themselves. This means that self-resemblance increases trust but does not increase helping. Therefore this suggests that trust and prosocial regard do not necessarily lead to helping, and trust within a kin recognition context may have a different benefit than helping.

The results of this study are specific to within-race differences in facial similarity. People may still help others who look similar to them compared to dissimilar to them when the differences are very large such as between different racial groups. In fact the literature show an overwhelming amount of research in which people are more like to help and cooperate with other people within their own race than other races (Stürmer & Siem, 2017). Empathy neural responses have even been found to be different for own race compared to other races (Xu, Saucier, Miller, & Doucet, 2005; Zuo, Wang, & Han, 2009), which may be the link between helping and racial group differences since empathy been established as a powerful link to helping behaviour. Contrary to this, (Dovidio, Gaertner, & Abad-Merino, 2017) suggests that people do not help others based on racial group but instead help others based on whether that person is a part of their own in-group or a part of an out-group, and that helping based on racial groups is reflecting an in-group bias not a racial bias within itself. Additionally, Balliet, Wu, and De Dreu (2014) conducted a meta-analysis of cooperative games including but not limited to, prisoners dilemma, dictator games, and public goods

games. Their results revealed that people were more likely to cooperate with in-group members than out-group members. This suggests that it's not the physical racial similarities that contribute to people helping others, but I identifying within the same group.

One of the main mechanisms behind in-group out-group preferences in helping is learning. Learning may also be a fundamental way that humans recognise kin, not through learning phenotypes and matching them against other which this study tested, but instead through directly learning who we are related to from abilities unique to humans (Krupp, DeBruine & Jones, 2011). This includes being taught directly through language and labels, as humans are commonly told from others who their close relatives, and each relative is given a label such as brother or cousin.

Alternatively, people may be learning indirectly who our kin are which are not unique to humans. These include spatially-based recognition and imprinting. Proximity to other people during infancy and childhood has been shown to lead to unconscious associations of those people as family members, and may serve as a major way in which humans recognise kin. Spatially-based recognition refers to when a person develops unconscious associations of people who live close or with them as family members. It is a primitive form of learning known as imprinting, which is the rapid learning of an attitude of behaviour during a critical period that becomes engrained and highly immune to change. For example the Westermarck effect asserts that the people you grow up with are more likely to be sexually avoided (Westermarck, 1891). This effect may have evolved as people who usually lived together were probably related, but this is not always the case which means that in some circumstances where non-kin live together during infancy then they may be sexually avoided like kin (Krupp, et al., 2011). For example, Wolf (1966) found that in Northern Taiwanese marriages where young girls live with their future husband and his family, problems such as sexual dysfunction were rife leading to most couples breaking up.

In contrast to kin selection, other processes and strategies may be more important when deciding who to preferentially help. For example, after being debriefed, one of the participants explained that they used a strategy based on fairness by alternating between selecting to help the face on the left and the face on the right, which lead to feeling less guilt about not helping one of each pair of faces based on superficial physical traits. This suggests that cognitive strategies that are most likely unique to humans may exist which people use extensively when deciding to help others, compared to other animals such as chimpanzees which predominately rely on visual cues to recognise kin (Parr & de Waal, 1999), and been found to help kin based on physical similarity (Mitani, Merriwether, & Zhang, 2000). It may be the case that learning and higher order cognition may contribute to helping others more than through recognising visual cues of relatedness in humans. In this current study, the participants were all Caucasian, almost all residents of Australia, young adults between 18 and 29, and educated at high school to university level. It may also be possible that these people are more likely to rely on such learning and cognitive strategies compared to relying simply on visual cues of relatedness, of which the results of this study may reflect.

Limitations

The EQ-40 was administered unaltered which presented strongly agree and slightly agree on the left while strongly disagree and slightly disagree was on the right. This layout is the opposite of the layout in almost all other questionnaires, and may have caused some confusion in participants which led to answers in the opposite direction. The reason for presenting them that way, is that it was the intended format of the questionnaire according to Baron-Cohen, and Wheelwright (2004). However, this may be a possible explanation to why Baron-Cohen, and Wheelwright (2004) found that a small proportion of non-autistic participants were scoring in similar ways as the autistic participants. This may further explain why a man in the main study, and a women in the stimuli validation study both scored a low

score of 14 out of 80, but provided normal scores for all other variables. Overall the EQ was not associated with empathising with the facial expressions of this experiment, and since the facial stimuli were validated as expressing each emotion correctly, it's likely that the EQ measured a general ability to empathise which is different to the specific to empathise with facial expressions. Additionally, the EQ was a self-report measure of empathy which may represent the participant's belief about their ability to empathise, and the task asking the participants how much they felt the emotion displayed in different faces may have been a more accurate measure of their true ability to empathise.

The results may also be due to differences in the facial stimuli and real faces. By presenting the similar and dissimilar faces together within the helping scenario, it is unlikely participants would have believed the scenario to be real and that their choices had real outcomes helping real people. In a future experiment, by not having the similar and dissimilar pairs of composites presented next to each other, the participants may help in a way that better reflects the real world and this may lead to a difference in helping self-similar compared to self-dissimilar composites."

Furthermore, faces that have been averaged into a composite have been shown to be more attractive than non-composite faces (Little, 2014). Thus, when transforming between an average face and the participant's face, the transformation towards the average face would look more attractive than the transformation towards the participants face. Furthermore, attractive people are more likely to be empathised with (Müller, Leeuwen, Baaren, Bekkering, & Dijksterhuis, 2013), and helped (Chen et al., 2012; Wilson, 1978) which may bias results in favour of the dissimilar composites which are transformed closer to average than to the participant. However, several experiments by Little, DeBruine, and Jones (2014), have shown that this may not be the case, and attractiveness differs along a separate dimension that that of averageness. Regardless, it may have been possible that the

participants were responding within a sexual context and emotionally empathised with dissimilar opposite sex participants more than similar opposite sex participants, assuming the sample was predominantly heterosexual. However the results show that this was only slightly evident in males empathising, and not enough to reach significance in either empathising or helping. Therefore, even if some participants were responding within a sexual context, according to the results these effects were cancelled out, possibly due to the other participants responding within a non-sexual context.

A power analyses showed that there was enough statistical power to find medium to large effects, but not small effects. A greater number of participants may have improved the chances of finding a significant effect, as the dropout rate between phase one and two was significantly greater than expected. Nevertheless, the differences between similar and dissimilar scores for empathy and helping were so small that even with more power the effects are likely to be extremely small.

Instead of forcing a standardised testing environment by having participants come into the lab, the participants were told to complete the experiment in a quiet room where they would be free from distractions and able to concentrate for 15 minutes. The helping task was low risk but high value, yet the overall the cost to the participant for helping was low, and self-preference may have a stronger altruistic effect in high cost scenarios (Stewart-Williams, 2007), such as a task in which a participant wouldn't usually help just a random person. Thus, future research should explore kin selection within these high risk scenarios.

Conclusion

This study has found no support for within-race self-resemblance affecting empathy or helping. Conversely, several sex differences in empathy were found that were consistent with the literature, such as women having higher EQ than men, men showed less emotional empathy towards other men than to women, and women empathised with sad facial

expressions much more than men did. This study's findings offers an alternative way that kin selection functions in humans, as visual cues of relatedness which were once thought to increase helping, have been shown to have no effect. In future research, the forced emotion photos seen could be used to see if there was a difference in empathy between forced and real emotions, which was beyond the scope of this current study. However, it is more likely that if a self-preference in empathy and helping exists, then it would only be present from the real emotions which this current study used. Future research should also be directed at understanding the particular evolutionary benefits trustworthiness gives in absence of helping, as this study's results suggest that trust may not be a precursor of cooperation and helping in the context of kin selection.

References

Alvergne, A., Perreau, F., Mazur, A., Mueller, U., & Raymond, M. (2014). Identification of visual paternity cues in humans. *Biology letters*, *10*(4), 20140063.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Ang, R. P., & Goh, D. H. (2010). Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender. *Child Psychiatry & Human Development*, *41*(4), 387-397.

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: A meta-analysis.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*(2), 163-175.

Batson, C. D., Batson, J. G., Griffitt, C. A., Barrientos, S., Brandt, J. R., Sprengelmeyer, P., & Bayly, M. J. (1989). Negative-state relief and the empathy—altruism hypothesis. *Journal of Personality and Social Psychology*, *56*(6), 922.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of personality and social psychology*, *45*(3), 706.

Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of economic Literature*, *14*(3), 817-826.

Bellard M., McAuliffe K., Burch R.L. (2006). From the child's point of view: The effect of paternal resemblance in investment. *Presented at the Meeting of the Human Behavior and Evolution Society*. Philadelphia, PA, June 7–11, 2006.

Bendapudi, N., Singh, S., & Bendapudi, V. (1996). Enhancing helping behavior: An integrative framework for promotion planning.

Benenson, J., Markovits, H., Fitzgerald, C., Geoffroy, D., Flemming, J., Kahlenberg, S., & Wrangham, R. (2009). Males' Greater Tolerance of Same-Sex Peers Psychological Science, 20 (2)

Bishop, D. I., Meyer, B. C., Schmidt, T. M., & Gray, B. R. (2009). Differential investment behavior between grandparents and grandchildren: The role of paternity uncertainty. *Evolutionary Psychology*, *7*(1), 147470490900700109.

Brandt, H., Hauert, C., & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1519), 1099-1104.

Brody, N., & Vangelisti, A. L. (2016). Bystander intervention in cyberbullying. *Communication Monographs*, *83*(1), 94-119.

Buss, D. M. (1996). Paternity uncertainty and the complex repertoire of human mating strategies.

Chaplin, T. M. (2015). Gender and emotion expression: A developmental contextual perspective. *Emotion Review*, *7*(1), 14-21.

Chaudhuri, A., & Gangadharn, L. (2003). *Gender Differences in Trust and Reciprocity* (No. 875). The University of Melbourne.

Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience letters*, *517*(2), 65-70.

Cialdini, R. B., & Kenrick, D. T. (1976). Altruism as hedonism: a social development perspective on the relationship of negative mood state and helping. *Journal of personality and social psychology*, *34*(5), 907.

Croson, R., & Buchan, N. (1999). Gender and culture: International experimental evidence from trust games. *The American Economic Review*, *89*(2), 386-391.

Danielsbacka, M., Tanskanen, A. O., Jokela, M., & Rotkirch, A. (2011). Grandparental child care in Europe: Evidence for preferential investment in more certain kin. *Evolutionary Psychology*, *9*(1), 147470491100900102.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. Barnes and Noble Publishing Inc.: New York.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113-126.

Dawkins, R. (1979). Twelve misunderstandings of kin selection. *Ethology*, *51*(2), 184-200.

Davis, M. H. (1983). The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of personality*, *51*(2), 167-184.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1498), 1307-1312.

DeBruine, L. M. (2004). Facial resemblance increases the attractiveness of same–sex faces more than other–sex faces. *Proceedings of the Royal Society of London B: Biological Sciences*, *271*(1552), 2085-2090.

DeBruine, L. M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1566), 919-922.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). University of North Carolina, Chapel Hill: Sage publications.

Dovidio, J. F., Gaertner, S. L., & Abad-Merino, S. (2017). Helping behaviour and subtle discrimination. In *Intergroup Helping* (pp. 3-22). Springer International Publishing.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological bulletin*, *100*(3), 283.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological bulletin*, *100*(3), 283.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results*, *1*, 1061-1073.

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and individual differences*, *24*, 96-102.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological bulletin*, *101*(1), 91.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological bulletin*, *101*(1), 91.

Euler, H. A., & Weitzel, B. (1996). Discriminative grandparental solicitude as reproductive strategy. *Human Nature*, *7*(1), 39-59.

Fabes, R. A., & Eisenberg, N. (1998). Meta-analyses of age and sex differences in children's and adolescents' prosocial behavior. *Handbook of Child Psychology*, *3*.

Farmer, H., McKay, R., & Tsakiris, M. (2014). Trust in me: Trustworthy others are seen as more physically similar to the self. *Psychological science*, *25*(1), 290-292.

Fatkin, J. M., & Lansdown, T. C. (2017). Group size and personalization's effect on Facebook message response rates. *Information Technology & People*, *30*(1), 71-80.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer program]. *Bonn, FRG: Bonn University, Department of Psychology*.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785-791.

Gardner, A., West, S. A., & Wild, G. (2011). The genetical theory of kin selection. *Journal of Evolutionary Biology*, *24*(5), 1020-1043.

Giang, T., Bell, R., & Buchner, A. (2012). Does facial resemblance enhance cooperation?. *PloS one*, *7*(10), e47809.

Gong, L. (2007). Is happy better than sad even if they are both non-adaptive? Effects of emotional expressions of talking-head interface agents. *International Journal of Human-Computer Studies*, *65*(3), 183-191.

Haldane, J. B. S. (1955). Population genetics. *New Biology.* 18, 34-51

Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, *26*(3), 245-256.

Hall, J. A., Gunnery, S. D., & Andrzejewski, S. A. (2011). Nonverbal emotion displays, communication modality, and the judgment of personality. *Journal of Research in Personality*, *45*(1), 77-83.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of theoretical biology*, *7*(1), 17-52.

Hamilton, W. D., & Axelrod, R. (1981). The evolution of cooperation. *Science*, *211*(27), 1390-1396.

Han, Y., Chen, B., & Zhang, X. (2016, November). Sex Difference of Saccade Patterns in Emotional Facial Expression Recognition. In *International Conference on Cognitive Systems and Signal Processing* (pp. 144-154). Springer, Singapore.

Henrich, J., & Fehr, E. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In P. Hammerstein (Ed.), Genetic and cultural evolution of cooperation (pp. 55 – 82). Cambridge, MA7 MIT Press.

Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of personality and social psychology*, *69*(2), 280.

Iredale, W., Van Vugt, M., & Dunbar, R. (2008). Showing off in humans: Male generosity as a mating signal. *Evolutionary Psychology*, *6*(3), 147470490800600302.

Johnson, R. C. (1996). Attributes of Carnegie medalists performing acts of heroism and of the recipients of these acts. *Ethology and Sociobiology*, *17*(5), 355-362.

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and violent behavior*, *9*(5), 441-476.

Klimecki, O. M., Mayer, S. V., Jusyte, A., Scheeff, J., & Schönenberg, M. (2016). Empathy promotes altruistic behavior in economic interactions. *Scientific reports*, *6*, 31961.

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, *32*(6), 1134.

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology*, *74*(3), 686.

Krupp, D. B., Debruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior*, *29*(1), 49-55.

Krupp, D. B., DeBruine, L. M., & Jones, B. C. (2011). Cooperation and conflict in the light of kin recognition systems. *The Oxford handbook of evolutionary family psychology*, 345-362.

Krupp, D. B., DeBruine, L. M., Jones, B. C., & Lalumière, M. L. (2012). Kin recognition: evidence that humans can perceive both positive and negative relatedness. *Journal of evolutionary biology*, *25*(8), 1472-1478.

Krupp, D. B., DeBruine, L. M., Jones, B. C., & Lalumière, M. L. (2012). Kin recognition: evidence that humans can perceive both positive and negative relatedness. *Journal of evolutionary biology*, *25*(8), 1472-1478.

Kumar, S., Calvo, R., Avendano, M., Sivaramakrishnan, K., & Berkman, L. F. (2012). Social support, volunteering and health around the world: Cross-national evidence from 139 countries. *Social science & medicine*, *74*(5), 696-706.

Lipps, T. (1903). Einfühlung, inner Nachahmung, und Organ-umpfindungen [Empathy, inner imitations, and sensations]. *Archiv für die gesamte Psychologie, 2*, 185–204

Little, A. C. (2014). Facial attractiveness. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 621-634.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2014). Sex differences in attraction to familiar and unfamiliar opposite-sex faces: Men prefer novelty and women prefer familiarity. *Archives of Sexual Behavior*, *43*(5), 973-981.

Marre, D. & Bestard, J. (2009). The family body: persons, bodies and resemblance. J. Edwards & C. Salazar eds. European kinship in the age of biotechnology. London & New York: Berghahn Books, chapter 4, pp. 64 - 78.

Mateo, J. M. (2015). Perspectives: Hamilton's legacy: mechanisms of kin recognition in humans. *Ethology*, *121*(5), 419-427.

McAndrew, F. T. (2002). New evolutionary perspectives on altruism: Multilevel-selection and costly-signaling theories. *Current Directions in Psychological Science*, *11*(2), 79-82.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of personality*, *40*(4), 525-543.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of personality*, *40*(4), 525-543.

Meier, S., & Stutzer, A. (2008). Is volunteering rewarding in itself?. *Economica*, *75*(297), 39-59.

Michalski, R. L., & Shackelford, T. K. (2005). Grandparental investment as a function of relational uncertainty and emotional closeness with parents. *Human Nature*, *16*(3), 293-305.

Mitani, J. C., Merriwether, D. A., & Zhang, C. (2000). Male affiliation, cooperation and kinship in wild chimpanzees

Müller, B. C., Leeuwen, M. L., Baaren, R. B., Bekkering, H., & Dijksterhuis, A. (2013). Empathy is a beautiful thing: Empathy predicts imitation only for attractive others. *Scandinavian journal of psychology*, *54*(5), 401-406.

Nowell, C., & Tinkler, S. (1994). The influence of gender on the provision of a public good. *Journal of Economic Behavior & Organization*, *25*(1), 25-36.

Parr, L. A., & de Waal, F. B. (1999). Visual kin recognition in chimpanzees. *Nature*, *399*(6737), 647-648.

Penton-Voak, I. S., Perrett, D. I., & Peirce, J. W. (1999). Computer graphic studies of the role of facial similarity in judgements of attractiveness. *Current Psychology*, *18*(1), 104-117.

Piper, G., & Schnepf, S. V. (2008). Gender differences in charitable giving in Great Britain. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *19*(2), 103-124.

Platek, S. M., & Kemp, S. M. (2009). Is family special to the brain? An event-related fMRI study of familiar, familial, and self-face recognition. *Neuropsychologia*, *47*(3), 849-858.

Pollet, T. V., Nettle, D., & Nelissen, M. (2006). Contact frequencies between grandparents and grandchildren in a modern society: Estimates of the impact of paternity uncertainty. *Journal of Cultural and Evolutionary Psychology*,*4*(3-4), 203-213.

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of personality assessment*, *93*(1), 84-95.

Robinson III, E. M., & Curry, J. R. (2005). Promoting altruism in the classroom. *Childhood Education*, *82*(2), 68-73.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21(2), 95-103

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and cognition*, *67*(2), 162-167.

Salamon, L. M., & Sokolowski, W. (2001). Volunteering in cross-national perspective: Evidence from 24 countries.

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and society*, *7*(1), 58-92.

Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in helping whites and blacks: A meta-analysis. *Personality and Social Psychology Review*, *9*(1), 2-16.

Schnall, S., Roper, J., & Fessler, D. M. (2010). Elevation leads to altruistic behavior. *Psychological science*, *21*(3), 315-320.

Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: cross-cultural and multimethod studies. *Journal of personality and social psychology*, *89*(6), 1010.

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, *6*(1), 38-47.

Servátka, M. (2009). Separating reputation, social influence, and identification effects in a dictator game. *European Economic Review*, *53*(2), 197-209.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617-627.

Simmons, C. H., & Sands-Dudelczyk, K. (1983). Children helping peers: Altruism and preschool environment. *The Journal of psychology*, *115*(2), 203-207.

Smith, J. M. (1964). Group selection and kin selection. *Nature*, *201*(4924), 1145-1147.

Sofer, C., Dotsch, R., Wigboldus, D. H., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science*, *26*(1), 39-47.

Stewart-Williams, S. (2007). Altruism among kin vs. nonkin: effects of cost of help and reciprocal exchange. *Evolution and human behavior*, *28*(3), 193-198.

Stürmer, S., & Siem, B. (2017). A Group-Level Theory of Helping and Altruism Within and Across Group Boundaries. In *Intergroup Helping* (pp. 103-127). Springer International Publishing.

Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards realism in facial image transformation: Results of a wavelet mrf method. In *Computer Graphics Forum* (Vol. 24, No. 3, pp. 449-456). Blackwell Publishing, Inc.

Titchener, E. (1909). *Elementary psychology of the thought processes*. New York: Macmillan.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, *46*(1), 35-57.

Westermarck, E. (1891). *The history of human marriage*. Macmillan: London.

Wilson, D. W. (1978). Helping behavior and physical attractiveness. *The Journal of Social Psychology*.

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of personality and social psychology*, *50*(2), 314.

Wolf, A. P. (1966). Childhood association, sexual attraction, and the incest taboo: A Chinese case. *American Anthropologist*, *68*(4), 883-898.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, *29*(26), 8525-8529.

Appendix A (page 1 of 2)

Human Research Ethics Committee Approval Form page 1/2.

Office of the Deputy Vice-Chancellor
(Research)

Research Office
Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 4459
http://www.research.mq.edu.au/
ABN 90 952 801 237

MACQUARIE University
SYDNEY·AUSTRALIA

20 April 2017

Dear Dr Stephen

Reference No: 5201700157

Title: *The effect of facial similarity on empathy and helping*

Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Human Sciences & Humanities)).

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted by:

- Macquarie University

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007 – Updated May 2015) (the *National Statement*).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the *National Statement*, which is available at the following website:

http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.

3. All adverse events, including events which might affect the continued ethical and scientific acceptability of the project, must be reported to the HREC within 72 hours.

4. Proposed changes to the protocol and associated documents must be submitted to the Committee for approval before implementation.

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 9850 4194 or by email ethics.secretariat@mq.edu.au

Appendix A (page 2 of 2)

Human Research Ethics Committee Approval Form page 2/2.

The HREC (Human Sciences and Humanities) Terms of Reference and Standard Operating Procedures are available from the Research Office website at:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics

The HREC (Human Sciences and Humanities) wishes you every success in your research.

Yours sincerely

Dr Karolyn White
Director, Research Ethics & Integrity,
Chair, Human Research Ethics Committee (Human Sciences and Humanities)

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007) and the *CPMP/ICH Note for Guidance on Good Clinical Practice.*

Appendix B (page 1 of 2)

The happy videos presented during phase one.

Happy Cat

The happy cat video played for 41 seconds. It involved cats ringing bells to receive food, however when one cat stops receiving food it becomes jealous of the other cat.

https://www.youtube.com/watch?v=xEMjTtaTzB8



Happy Dog

The happy dog video played for 1:15 minutes. It involved a guilty dog with a bin lid stuck on his head whose owner came home to find surrounded by trash.

https://www.youtube.com/watch?v=PuA6ZjpEJys

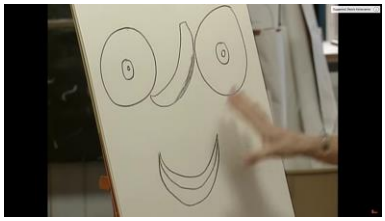


Happy Mr Bean

The happy Mr Bean video played for 2:50 minutes. It was a clip from Mr Bean in which he attempts to take part in an art class.

https://www.youtube.com/watch?v=lAnOylx4J3E





Appendix B (page 2 of 2)

The sad videos presented during phase one.

Sad Breakup

The sad breakup video played for 2:36 minutes. It was a clip from Dawson's Creek involving one of the characters finding out about being cheated on which leads to a relationship break up.

https://www.youtube.com/watch?v=IPf7r6nyhac



Sad Refugee

The sad refugee video played for 3:05 minutes. It involved a Caucasian English girl becoming a refugee when England goes to war, as a way to make refugees more relatable to the Caucasian people of developed nations.

https://www.youtube.com/watch?v=Aa3o7vWG93w



Sad Dog

The sad dog video played for 5:15 minutes. It was a series of clips from Futurama in which one of the characters has a chance to bring his dog back to life, but doesn't because he thinks the dog lived a full life without him, however it is revealed that the dog actually spent his whole life waiting for his owner to return until he died waiting.

https://www.dropbox.com/s/y90lyg216sfwzm9/sad%203%20dog.mp4?dl=0



Appendix C (page 1 of 4)

The Empathy Quotient Questionnaire (EQ-40) instructions.

How to fill out the questionnaire

Below are a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree with it by circling your answer. There are no right or wrong answers, or trick questions.

IN ORDER FOR THE SCALE TO BE VALID, YOU MUST ANSWER EVERY QUESTION.

Examples

E1. I would be very upset if I couldn't listen to music every day.	strongly agree	(slightly agree)	slightly disagree	strongly disagree
E2. I prefer to speak to my friends on the phone rather than write letters to them.	strongly agree	slightly agree	slightly disagree	(strongly disagree)
E3. I have no desire to travel to different parts of the world.	(strongly agree)	slightly agree	slightly disagree	strongly disagree
E4. I prefer to read than to dance.	strongly agree	slightly agree	(slightly disagree)	strongly disagree

Appendix C (page 2 of 4)

The Empathy Quotient Questionnaire (EQ-40). Question 1 to 13.

1. I can easily tell if someone else wants to enter a conversation.	strongly agree	slightly agree	slightly disagree	strongly disagree
2. I find it difficult to explain to others things that I understand easily, when they don't understand it first time.	strongly agree	slightly agree	slightly disagree	strongly disagree
3. I really enjoy caring for other people.	strongly agree	slightly agree	slightly disagree	strongly disagree
4. I find it hard to know what to do in a social situation.	strongly agree	slightly agree	slightly disagree	strongly disagree
5. People often tell me that I went too far in driving my point home in a discussion.	strongly agree	slightly agree	slightly disagree	strongly disagree
6. It doesn't bother me too much if I am late meeting a friend.	strongly agree	slightly agree	slightly disagree	strongly disagree
7. Friendships and relationships are just too difficult, so I tend not to bother with them.	strongly agree	slightly agree	slightly disagree	strongly disagree
8. I often find it difficult to judge if something is rude or polite.	strongly agree	slightly agree	slightly disagree	strongly disagree
9. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.	strongly agree	slightly agree	slightly disagree	strongly disagree
10. When I was a child, I enjoyed cutting up worms to see what would happen.	strongly agree	slightly agree	slightly disagree	strongly disagree
11. I can pick up quickly if someone says one thing but means another.	strongly agree	slightly agree	slightly disagree	strongly disagree
12. It is hard for me to see why some things upset people so much.	strongly agree	slightly agree	slightly disagree	strongly disagree
13. I find it easy to put myself in somebody else's shoes.	strongly agree	slightly agree	slightly disagree	strongly disagree

Appendix C (page 3 of 4)

The Empathy Quotient Questionnaire (EQ-40). Question 14 to 26.

14. I am good at predicting how someone will feel.	strongly agree	slightly agree	slightly disagree	strongly disagree
15. I am quick to spot when someone in a group is feeling awkward or uncomfortable.	strongly agree	slightly agree	slightly disagree	strongly disagree
16. If I say something that someone else is offended by, I think that that's their problem, not mine.	strongly agree	slightly agree	slightly disagree	strongly disagree
17. If anyone asked me if I liked their haircut, I would reply truthfully, even if I didn't like it.	strongly agree	slightly agree	slightly disagree	strongly disagree
18. I can't always see why someone should have felt offended by a remark.	strongly agree	slightly agree	slightly disagree	strongly disagree
19. Seeing people cry doesn't really upset me.	strongly agree	slightly agree	slightly disagree	strongly disagree
20. I am very blunt, which some people take to be rudeness, even though this is unintentional.	strongly agree	slightly agree	slightly disagree	strongly disagree
21. I don’t tend to find social situations confusing.	strongly agree	slightly agree	slightly disagree	strongly disagree
22. Other people tell me I am good at understanding how they are feeling and what they are thinking.	strongly agree	slightly agree	slightly disagree	strongly disagree
23. When I talk to people, I tend to talk about their experiences rather than my own.	strongly agree	slightly agree	slightly disagree	strongly disagree
24. It upsets me to see an animal in pain.	strongly agree	slightly agree	slightly disagree	strongly disagree
25. I am able to make decisions without being influenced by people's feelings.	strongly agree	slightly agree	slightly disagree	strongly disagree
26. I can easily tell if someone else is interested or bored with what I am saying.	strongly agree	slightly agree	slightly disagree	strongly disagree

Appendix C (page 4 of 4)

The Empathy Quotient Questionnaire (EQ-40). Question 27 to 40.

27. I get upset if I see people suffering on news programmes.	strongly agree	slightly agree	slightly disagree	strongly disagree
28. Friends usually talk to me about their problems as they say that I am very understanding.	strongly agree	slightly agree	slightly disagree	strongly disagree
29. I can sense if I am intruding, even if the other person doesn't tell me.	strongly agree	slightly agree	slightly disagree	strongly disagree
30. People sometimes tell me that I have gone too far with teasing.	strongly agree	slightly agree	slightly disagree	strongly disagree
31. Other people often say that I am insensitive, though I don't always see why.	strongly agree	slightly agree	slightly disagree	strongly disagree
32. If I see a stranger in a group, I think that it is up to them to make an effort to join in.	strongly agree	slightly agree	slightly disagree	strongly disagree
33. I usually stay emotionally detached when watching a film.	strongly agree	slightly agree	slightly disagree	strongly disagree
34. I can tune into how someone else feels rapidly and intuitively.	strongly agree	slightly agree	slightly disagree	strongly disagree
35. I can easily work out what another person might want to talk about.	strongly agree	slightly agree	slightly disagree	strongly disagree
36. I can tell if someone is masking their true emotion.	strongly agree	slightly agree	slightly disagree	strongly disagree
37. I don't consciously work out the rules of social situations.	strongly agree	slightly agree	slightly disagree	strongly disagree
38. I am good at predicting what someone will do.	strongly agree	slightly agree	slightly disagree	strongly disagree
39. I tend to get emotionally involved with a friend's problems.	strongly agree	slightly agree	slightly disagree	strongly disagree
40. I can usually appreciate the other person's viewpoint, even if I don't agree with it.	strongly agree	slightly agree	slightly disagree	strongly disagree

Thank you for filling this questionnaire in.

Appendix D (page 1 of 6)

The happy male composite faces.

Happy Male Light

Happy Male Mixed

Happy Male Dark1

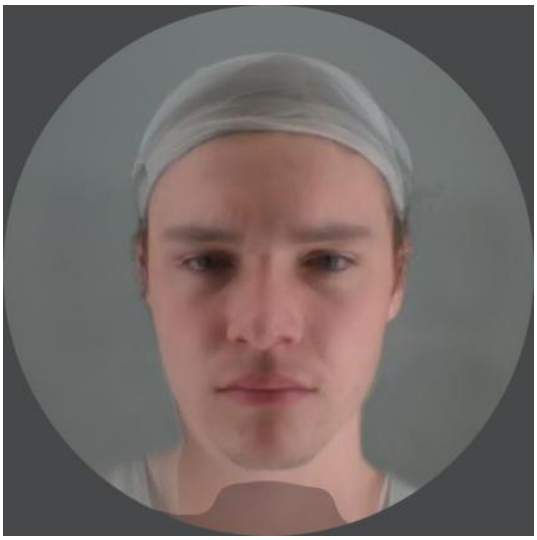
Happy Male Dark2

Happy Male Dark3

Appendix D (page 2 of 6)

The sad male composite faces.

Sad Male Light



Sad Male Mixed

Sad Male Dark1

Sad Male Dark2

Sad Male Dark3

Appendix D (page 3 of 6)

The neutral male composite faces.

Neutral Male Light

Neutral Male Mixed

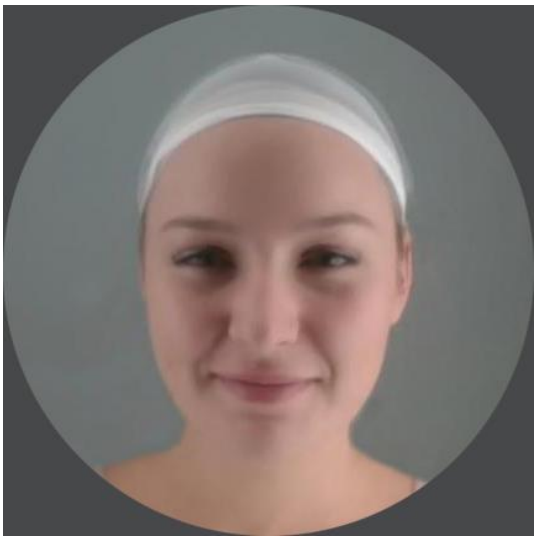
Neutral Male Dark1

Neutral Male Dark2

Neutral Male Dark3

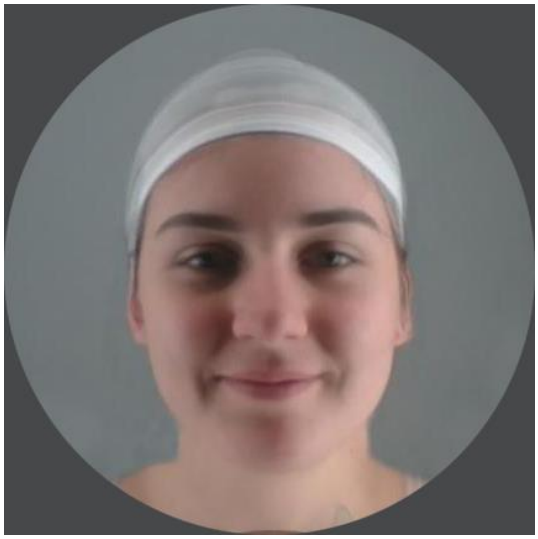
Appendix D (page 4 of 6)

The happy female composite faces.

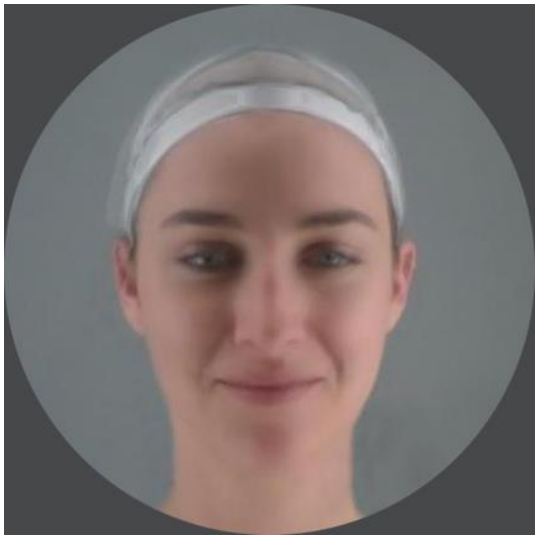


Happy Female Light

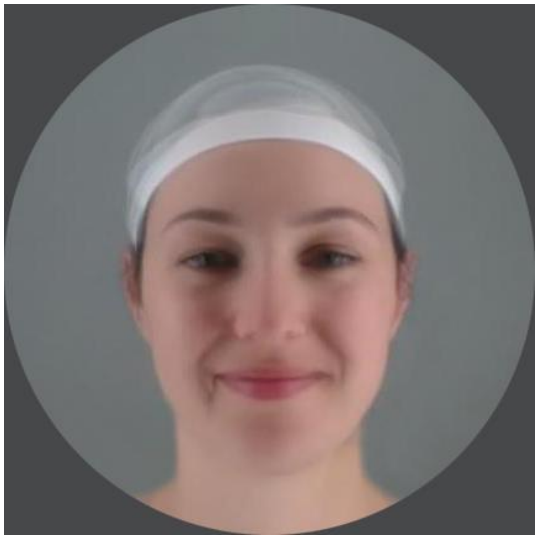
Happy Female Mixed



Happy Female Dark1



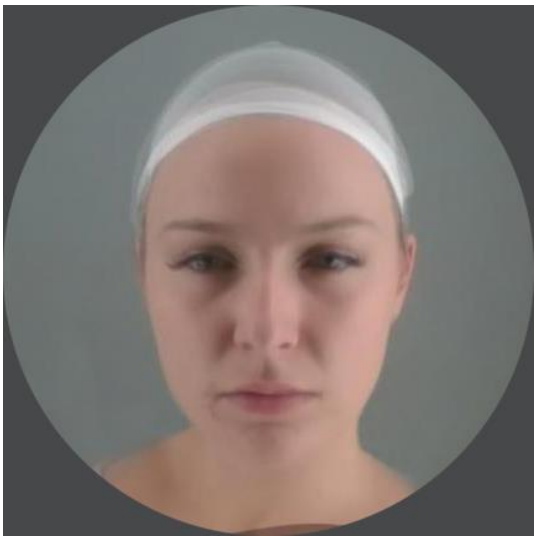
Happy Female Dark2



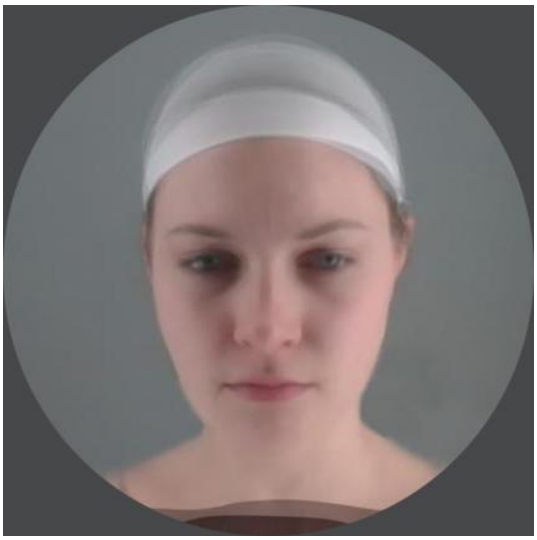
Happy Female Dark3

Appendix D (page 5 of 6)

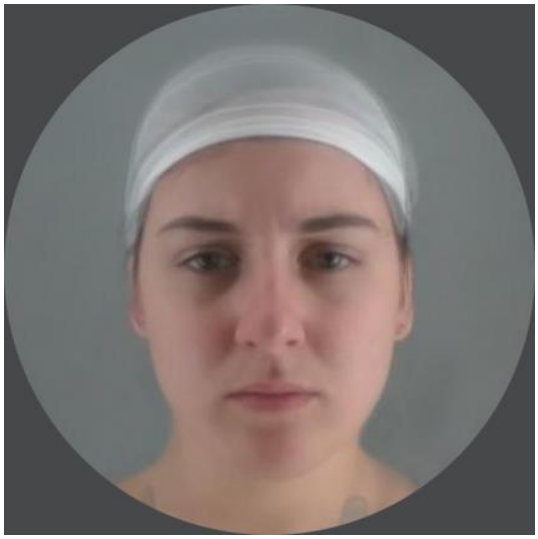
The sad female composite faces.



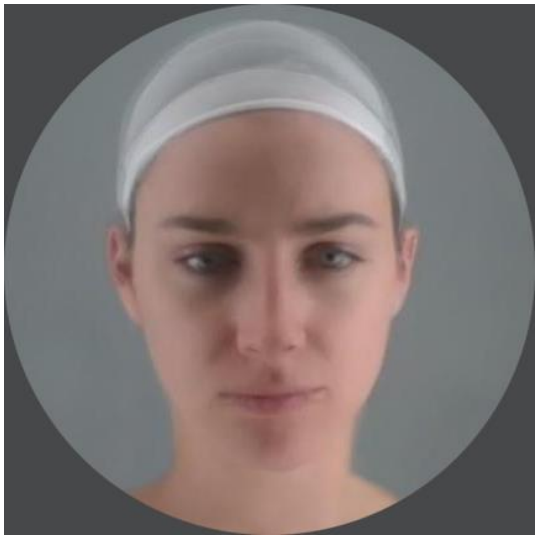
Sad Female Light



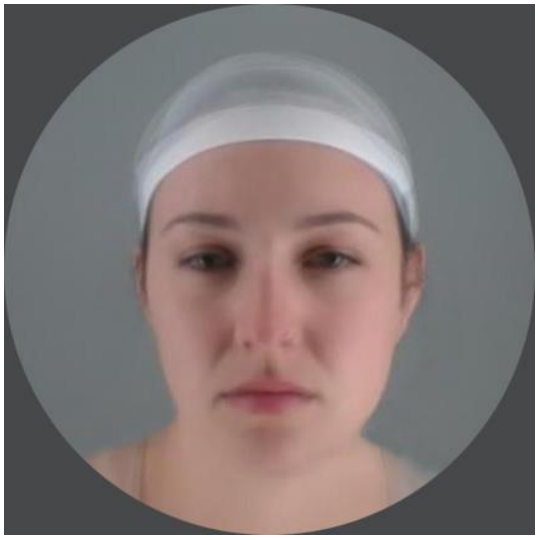
Sad Female Mixed



Sad Female Dark1



Sad Female Dark2



Sad Female Dark3

Appendix D (page 6 of 6)

The neutral female composite faces.

Neutral Female Light

Neutral Female Mixed

Neutral Female Dark1

Neutral Female Dark2

Neutral Female Dark3