Social Perception in Group Scenes:

Social context modulates perceptions of facial attractiveness

Daniel J. Carragher

(B. Psych Hons)

A dissertation presented to the

College of Education, Psychology, and Social Work of Flinders University

for the degree of Doctor of Philosophy in the field of Psychology

19th December 2018

Table of Contents

Thesis Summary

Despite the cautionary reminder to never "judge a book by its cover", we regularly judge others based upon their facial appearance. Far from meaning that we are all terribly judgmental, these trait impressions occur automatically. Even though they are often not accurate, the trait judgments that we make about others can influence our own decision making. The candidate with the more "competent" face wins approximately 70% of national elections, and criminals with "untrustworthy" faces receive longer prison sentences for the same crimes than those with "trustworthy" faces.

Trait impressions have been the focus of research in the field of *social perception* since the earliest days of experimental psychology. While these studies have undoubtedly improved our understanding of the way that trait judgments are made from faces, the vast majority of these studies have been conducted by presenting observers with a single face at a time. However, we often meet people for the first time when they are surrounded by others, perhaps at a café or a bar. Consequently, very little is known about the way we make trait impressions about an individual face that is seen among a group of other faces. Within this thesis, I aimed to improve our understanding of the way that the facial attractiveness and trustworthiness of an individual is evaluated when they are seen among a group of other faces, compared to when they are seen alone.

In the introduction to this thesis, I discuss the factors that influence the trustworthiness judgments that are made from the face, as well as the characteristics of faces that are perceived to be attractive. Then, I describe the way that the visual system processes complex visual scenes, such as a group of faces, using a process called ensemble coding. Bringing together these lines of research, I discuss "the cheerleader effect", a phenomenon that is said to occur when the same face is perceived to be more attractive in a group compared to alone.

The research in this thesis significantly advances our understanding of the cheerleader effect. My findings show that all individuals are perceived to be approximately 1.5-2% more

attractive in a group than they are alone, regardless of how attractive they are, or how attractive the other group members are. I also show that the cheerleader effect does not extend to judgments of trustworthiness. Crucially, my findings are also inconsistent with the hierarchical encoding mechanism that was initially proposed to cause the cheerleader effect, which was related to the way that the visual system uses ensemble coding to summarise groups of faces. Based on the results contained within this thesis, I offer an alternative explanation for the cheerleader effect, which suggests that the effect might be related to the socially desirable characteristics that are attributed to individuals in groups. In conclusion, my findings demonstrate that social context reliably influences perceptions of facial attractiveness, and suggests that the field of social perception must be expanded to consider the influence of social context on the trait impressions that are made from the face.

Acknowledgements

Professor Mike Nicholls – Thank you for welcoming me into the Brain and Cognition Lab. I have learned more from you about psychology, academia, watches and fine whisky, than I could ever put into words. Thank you for your enthusiasm, encouragement, persistence and honesty over the past four years. I will miss having lunch in the lab every day, at noon sharp.

Dr Nicole Thomas – I could not have imagined a better or more supportive supervisor. From taking me to conferences, on lab visits, and everything in between, you have truly gone above and beyond. I feel very lucky to have been your student. Thank you for everything.

The Brain and Cognition Laboratory – To the senior members of the lab, Dr Megan Bartlett, Dr Owen Churches, Dr Oren Griffiths, Dr Scott Gwinn, and Dr Blake Lawrence. All of you have gone out of your way to help me during my time in the lab. I am incredibly grateful to have had such a fantastic team to work with, learn from, and look up to.

To my fellow PhDs, Ali, Ancret, Ellie, Liz, Nathan, Shannon, Sophie and Steph. I have had the most fantastic four years as a member of the Brain and Cognition Lab, and that is because of you. Your friendship provided endless amounts of joy and laughter, and for that I will always be grateful. I could not imagine having done this without you. I will miss each of you terribly.

Family – To Mum, Dad, and Ethan. I would not be the person I am today without your love and support. Thank you for always encouraging me to pursue my goals, no matter how crazy they might seem. This thesis would not have been possible without you. To Katie, Tara, and Ian. Thank you for your kindness and support over these past years. Finally, to Jessica. Thank you for your unconditional love, support, understanding, and truly endless patience.

I would also like to acknowledge the Australian Government for awarding me an Australian Government Research Training Program Scholarship.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that, to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text.

Research from this thesis has been published in the following peer-reviewed articles:

Carragher, D. J., Lawrence, B. J., Thomas, N. A., & Nicholls, M. E. R. (2018). Visuospatial asymmetries do not modulate the cheerleader effect. *Scientific Reports*, *8*(1), 2548, doi: 10.1038/s41598-018-20784-5

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. (2017). Is trustworthiness lateralized in the face? Evidence from a trust game. *Laterality: Asymmetries of Body, Brain and Cognition*, 20-38, doi: 10.1080/1357650x.2017.1298120

Research from this thesis is currently under review at the following journals:

Carragher, D. J., Thomas, N. A., Gwinn, O. S., & Nicholls, M. E. R. (Resubmitted). The cheerleader effect occurs in the absence of hierarchical encoding. *Journal of Experimental Psychology: General.*

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. (Submitted). The dissociable influence of group presence on evaluations of attractiveness and trustworthiness. *The Quarterly Journal of Experimental Psychology.*

Research from this thesis has been presented at the following conferences:

Carragher, D. J. 2018. *First impressions: The curious case of The Cheerleader Effect.* Presentation delivered at the Higher Degree Research Conference, Flinders University, June 19th.

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. 2018. *I get more attractive with a little help from my friends: Dual mechanisms underlie the cheerleader effect.* Poster presented at the 18th meeting of the Vision Sciences Society, Florida, U.S.A, May 18-23.

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. 2018. *Why are faces more attractive in a group? Multiple mechanisms underlie the cheerleader effect.* Presentation delivered at the 44th Experimental Psychology Conference, Hobart, Australia, April 4-7.

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. 2017. *I get more attractive with a little help from my friends.* Presentation delivered at the 44th Experimental Psychology Conference, Shoal Bay, Newcastle, Australia, April 19-22.

Carragher, D. J., Churches, O., Thomas, N. A., & Nicholls, M. E. R. 2016. *Turn Either Cheek: Trustworthiness is not lateralised in the human face.* Poster presented at the 5th North Sea Laterality Conference, Groningen, the Netherlands, September 1-3.

Research from this thesis has also been communicated in the following seminars:

Carragher, D. J. 2018. *Why am I more attractive in a group? Understanding the mechanisms of the cheerleader effect.* Invited presentation delivered to the Vision and Memory, and the Multisensory Integration and Perception Labs, University of California, San Diego, May 15.

Carragher, D. J. 2018. *Why am I more attractive with my friends? Understanding the cheerleader effect.* Invited presentation delivered to the Mind, Body and Cognition Colloquium, College of Education, Psychology and Social Work, Flinders University, April 30.

Carragher, D. J. 2017. *I get more attractive with a little help from my friends; Understanding the cheerleader effect.* Invited presentation delivered to the School of Psychology, Victoria University of Wellington, New Zealand, July 27.

Signed

____

Daniel J. Carragher

Author Note

I have used the pronoun "we" instead of "I" when referring to the actions taken during the course of research in each of the four experimental chapters contained within this thesis. Although I have taken the primary role in conceptualising, designing, programming, conducting, analysing and writing the experiments contained within this thesis, I have had valuable help from the co-authors of each publication, who are listed at the start of each experimental chapter. Therefore, the pronoun "we" appropriately conveys the collaborative nature of this work, and the role of that all listed authors had in seeing each project through to completion. I have used the pronouns "I" and "my" sparingly in the introduction and general discussion sections, primarily when describing the outline of the thesis, or summarising the results of each experimental chapter.

Finally, all four experimental chapters within this thesis have been published, or submitted for publication, in their own right. As such, each chapter has been written to stand alone as a unique publication. Although I have made every effort the avoid the unnecessary repetition of ideas between chapters, the similar line of research carried through Chapters 4, 5 and 6 make some overlap between the content of these chapters unavoidable.

Chapter 1: First Impressions of Faces

1.1 Introduction Overview

Despite the remarkable breadth of research in the field of social perception, very little is known about the way we make trait impressions for faces that are encountered in a social context. The overarching aim of this thesis was to investigate the changes that occur to judgments of facial attractiveness and trustworthiness when the same individual face is seen in a group with other faces, compared to when they are seen alone. Specifically, my research sought to extend our understanding of "*the cheerleader effect*", a little known phenomenon whereby the same face is perceived to be more attractive in a group compared to alone (Walker & Vul, 2014).

The introduction to this thesis contains two chapters. To begin, Chapter 1 outlines the importance of understanding the trait impressions that we make from faces, and discusses how judgments of trustworthiness and attractiveness are made from facial appearance. Chapter 1 concludes by summarising the few existing investigations that have examined how the attractiveness of an individual is evaluated in a group. Chapter 2 describes how the visual system processes complex visual scenes, such as a group of faces, through a process known as ensemble coding. Furthermore, Chapter 2 discusses the way that complex group scenes are stored in memory, and how this process affects the way that individual items from the group are remembered. Finally, I bring these two lines of research together to address the cheerleader effect.

1.2 First Impressions

The face contains a remarkable amount of information about a person. From the face, we can tell someone's age, identity, gender, ethnicity and even the emotion they are feeling. But, can we also tell their personality from their face? Aristotle's *Physiognomica*, published around 300 BC, is the oldest surviving text that describes physiognomy, the art of judging character from the face. These early physiognomic teachings were based upon attributing the perceived personality characteristics of animals to the individuals whose appearance resembled that animal. Thus, those who were fortunate enough to resemble a lion were seen to be courageous, whereas those who were already unfortunate enough to resemble a cow, were thought to be lazy (Della Porta, 1586 as cited in Todorov, 2017). Despite initially sounding relatively trivial, there were serious consequences to the popularity of physiognomy in the 19th century (Todorov, 2017). Cesare Lombroso, the founder of criminal anthropology, believed that criminals could be identified by the shape of their face, and offered his "expert" testimony in numerous criminal trials (Oosterhof & Todorov, 2008). Physiognomy also came incredibly close to changing the course of scientific discovery. Charles Darwin's theory of evolution was based upon the observations he made during his voyage to South America aboard *The Beagle* in 1831. However, Darwin was nearly refused his place aboard *The Beagle*, because of the shape of his nose. According to Darwin, the captain of the ship, a "disciple of Lavater…. doubted whether anyone with my nose could possess sufficient energy and determination for the voyage" (p. 72, Darwin, 1973).

Although the idea that the shape of someone's nose can tell us anything about their personality might sound ridiculous today, other physiognomic beliefs remain surprisingly common. Even recently, 75% of individuals surveyed indicated that it was possible to accurately identify personality traits from the face (Hassin & Trope, 2000). Perhaps the belief that first impressions are accurate endures because we are all so familiar with having made them. After all, first impressions occur spontaneously (Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002), and importantly, they influence our decision making and behaviour towards others (Olivola, Funk, & Todorov, 2014; Rezlescu, Duchaine, Olivola, & Chater, 2012; Wilson & Eckel, 2006). The winning candidate in national elections can be predicted by judgments of facial competence (Antonakis & Dalgas, 2009; Ballew & Todorov, 2007; Todorov, Mandisodza, Goren, & Hall, 2005), and judgments of facial trustworthiness predict the severity of criminal sentences given to convicted felons (Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015, 2016). Of course, the ability to make these impressions rapidly would be remarkably useful, if they were accurate (Olivola et al., 2014; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Todorov & Porter, 2014).

Recently, many studies have reported that there is "a kernel of truth" to the trait judgments that we make from faces (for review see Tskhay & Rule, 2013). Extraversion (Penton-Voak, Pound, Little, & Perrett, 2006), intelligence (Zebrowitz, Hall, Murphy, & Rhodes, 2002), trustworthiness (Verplaetse, Vanneste, & Braeckman, 2007), sexual orientation (Rule & Ambady, 2008; Rule, Ambady, & Hallett, 2009), religious denomination (Andrzejewski, Hall, & Salib, 2009) and even political affiliation (Olivola & Todorov, 2010a; Rule & Ambady, 2010) have all been found to be identified from the face at above chance levels. Although debating the accuracy of first impressions is beyond the scope of the present work, others have urged that caution is needed when interpreting the practical significance of these findings (Olivola & Todorov, 2010b; Todorov et al., 2015). Even in unrealistically favourable conditions (Olivola & Todorov, 2010b; Todorov et al., 2015), accuracy is typically only just above chance levels (55-65%, Tskhay & Rule, 2013). Furthermore, judgments about stable personality traits vary substantially between different photographs of the same person, strongly suggesting that these trait judgments are highly dependent on the image selected, rather than any enduring personality characteristics of the individual (Rule, Krendl, Ivcevic, & Ambady, 2013; Todorov & Porter, 2014). Yet, regardless of whether these impressions are accurate, there is very strong agreement between observers in the trait impressions that are made from faces (Olivola & Todorov, 2010b; Todorov, 2008; Todorov et al., 2015). So, if these impressions are not reliably accurate, why do we make them at all?

To understand why first impressions are made from the face, we first need to consider the role that the face has in social signalling. The face is used to display emotional expressions, which have a vital role in non-verbal communication (Batty & Taylor, 2003; Elfenbein & Ambady, 2002). Happiness, sadness, anger, disgust, fear, and surprise are perceived accurately both within and between cultures, demonstrating that the expression and recognition of these emotions is universal (Ekman, 1971; Ekman & Friesen, 1971; Ekman et al., 1987; Elfenbein & Ambady, 2002, 2003; however, see Russell, 1994). Importantly, emotional expressions are *transient* displays that occur in response to a stimulus. Secord (1958) first proposed that personality impressions occurred due to a process he called *temporal extension*, whereby "the perceiver regards a momentary characteristic of the person as if it were an enduring attribute" (p. 307). According to Secord's hypothesis, someone who is smiling is not only happy, but is also judged to be a kind and friendly person. Knutson (1996) tested this temporal extension hypothesis by asking observers to rate the perceived dominance and affiliation of individuals that were photographed showing the six basic emotional expressions[1]. Faces displaying angry expressions were judged to be high on dominance and low on affiliation, whereas those showing happiness were high on dominance and on affiliation (Knutson, 1996), suggesting that observers do infer stable personality traits from temporary emotional expressions (Secord, 1958). However, temporal extension alone cannot account for all of the trait impressions that we make, because trait impressions also occur rapidly for faces without emotional expressions (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006).

Emotional expressions are important social signals that forecast the behavioural intentions of others (Adams, Ambady, Macrae, & Kleck, 2006; Adams & Kleck, 2005; Marsh, Ambady, & Kleck, 2005). Accurate perception of anger is critical for survival, because the expression of anger signals the approach intent of an individual (Adams et al., 2006). Conversely,

---

[1] Wiggins, Phillips, and Trapnell (1989) showed that verbal personality descriptions could be plotted in a circular space arranged around two orthogonal trait axes; affiliation (cold-hearted/warm-agreeable) and dominance (assured-dominant/unassured-submissive).

seeing an angry expression facilitates avoidance responses in the observer, which is presumed to allow us to avoid aggressive individuals and dangerous situations (Marsh et al., 2005; Stins et al., 2011; Willis, Palermo, & Burke, 2011). McArthur and Baron (1983) proposed that the adaptive benefit of avoiding angry individuals is so great that we have come to "see" emotional expressions in neutral faces. For example, if someone's resting face has slightly upturned lips they might be perceived to be happy, whereas someone whose brow is naturally furrowed might appear angry. Consistent with this emotion overgeneralisation hypothesis, Montepare and Dobish (2003) found that observers attributed the same personality traits to neutral faces that merely resembled happy or angry expressions, as they did to faces that were photographed showing happiness and anger (Knutson, 1996). Taken together, spontaneous trait impressions are likely caused by a tendency to infer stable personality traits from temporary emotional expressions (Knutson, 1996; Secord, 1958), and a disposition to perceive emotions in neutral faces that structurally resemble emotional expressions (Montepare & Dobish, 2003; Said, Sebe, & Todorov, 2009b; Todorov, 2008; Todorov et al., 2015; Zebrowitz & Montepare, 2006; Zebrowitz & Montepare, 2008). This process of emotion overgeneralisation can explain why trait impressions occur automatically, even though they are not inherently accurate (Todorov, 2008).

With general agreement among researchers as to why trait impressions occur, Oosterhof and Todorov (2008) turned their attention to identifying the attributes of the face that underlie these trait impressions. This question is complicated by the fact that trait impressions from natural faces are highly correlated with one another (Oosterhof & Todorov, 2008). To address this issue, Oosterhof and Todorov (2008) used a data driven approach, wherein observers were asked to make spontaneous impressions about a set of randomly generated artificial faces, which all had neutral expressions. Remarkably, the observers gave more than 1,100 different descriptions of the faces, which were reduced to 15 different traits using a principle component analysis (PCA)[2]. Then, a new sample of participants judged the same neutral faces using only these 15 traits, with the subsequent ratings submitted to a second PCA to identify relationship between the structural characteristics of each face and the judgments that were made. This second PCA identified two underlying factors in the trait judgments; emotional valence, which accounted for 63% of the variance in trait judgments, and perceived dominance, which accounted for 18% of the variance (Oosterhof & Todorov, 2008)[3]. Interestingly, trustworthiness judgments had the strongest correlation with the valence factor (r = .92) but were unrelated to the dominance factor (r = -.10), whereas (unsurprisingly) judgments of dominance were highly correlated with the dominance factor (r = .87) but not with the valence factor (r = -.20). Oosterhof and Todorov's (2008) model shows that spontaneous trait impressions vary reliably as a combination of the valence and dominance of a face in a two-dimensional trait space.

1.3 Trustworthiness

Of the many personality traits that are spontaneously evaluated from the face, trustworthiness judgments are perhaps the most important (Todorov, 2008). Trustworthiness judgments had the strongest correlation with the valence dimension of trait space, which suggests that trustworthiness judgments alone can be used to approximate the general valence of a face (Oosterhof & Todorov, 2008; Todorov, 2008). These findings are consistent with the intuitive notion that trustworthiness judgments are related to positive expressions (Oosterhof & Todorov, 2008), as smiling faces are perceived to be more trustworthy than faces without smiles (Krumhuber et al., 2007; Scharlemann, Eckel, Kacelnik, & Wilson, 2001). The relationship between trustworthiness and positive expressions was further supported by Oosterhof and Todorov (2008), who used their computer modelling technique to create a continuum of faces

[2] Oosterhof and Todorov (2008) reported that the 13 impressions to show reliable interrater agreement were: attractive, weird, mean, trustworthy, aggressive, caring, emotionally stable, unhappy, responsible, sociable, dominant, confident and intelligent. Two traits, egoistic and boring, showed poor interrater reliability, and were excluded from the second PCA.

[3] Using naturalistic faces, Sutherland et al. (2013) replicated the trustworthiness (approachability) and dominance dimensions in trait space. However, Sutherland and colleagues also identified an additional dimension of youthfulness-attractiveness, which might have emerged due to the greater variability of natural faces.

that varied only on trustworthiness. Even though the model was created using the trustworthiness judgments given to faces without expressions, Oosterhof and Todorov (2008) found that extremely untrustworthy faces (i.e., those -8 SD from the mean), had developed angry expressions, while extremely trustworthy faces (+8 SD) had developed smiles. These findings suggest that when evaluating facial trustworthiness, observers are highly sensitive to cues in neutral faces that resemble positive and negative expressions (Engell, Todorov, & Haxby, 2010).

The strong relationship between trustworthiness and emotional valence, specifically expressions of happiness and anger (Engell et al., 2010; Oosterhof & Todorov, 2008), suggests that trustworthiness judgments might be used in approach avoidance decision making (for review see Todorov, 2008). Notably, viewing expressions of happiness or anger facilitate approach and avoidance responses, respectively, within an observer (Adams et al., 2006; Marsh et al., 2005; Stins et al., 2011; Willis et al., 2011). Furthermore, trustworthiness judgments occur rapidly, which is necessary for survival responses (Bar et al., 2006; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Willis and Todorov (2006) found a strong correlation between trustworthiness judgments for faces that were presented for just 100 ms and those made without time constraint, while Todorov et al. (2009) found that trustworthiness judgments were made reliably for faces presented for just 33 ms, but not 17 ms. From this minimum threshold of 33 ms, the reliability of trustworthiness judgments continued to increase until presentation times of 167 ms; after which time the reliability of the trustworthiness judgments plateaued, suggesting that they had become firmly established (Todorov et al., 2009). Interestingly, Bar et al. (2006) found that judgments of threat, but not intelligence, could be made reliably for faces presented for 39 ms, which raises the possibility that trustworthiness impressions are akin to the survival related judgments that occur before higher-order trait impressions.

The perceived trustworthiness of unfamiliar faces also automatically activates the amygdala (Engell et al., 2007; Said, Baron, & Todorov, 2009a; Todorov & Engell, 2008; Winston et al., 2002), a neural structure that is implicated in the processing of fearful and threatening stimuli (Öhman, 2005). In an fMRI study, Winston et al. (2002) presented unfamiliar faces to observers, who were either asked to judge the trustworthiness or age of each face. Regardless of whether observers were engaged in the trustworthiness rating task or the age estimation task, amygdala activity increased as the trustworthiness of the faces decreased, indicating that the amygdala automatically evaluates the trustworthiness of unfamiliar faces (Engell et al., 2007; Todorov, Baron, & Oosterhof, 2008). The activation of the amygdala in response to facial trustworthiness is non-linear, such that compared to faces of average trustworthiness, greater activation occurs in response to both very untrustworthy and very trustworthy faces (Said et al., 2009a; Todorov, 2008; Todorov et al., 2008). Notably, amygdala activation is greater still for untrustworthy than trustworthy faces (Said et al., 2009a), which might suggest that we prioritise the need to detect and avoid untrustworthy individuals over the need to approach trustworthy individuals (Todorov, 2008).

Trustworthiness judgments are critically important for the development of social cooperation (Cosmides & Tooby, 1992). Gunnthorsdottir, McCabe, and Smith (2002) defined trust as the "voluntary transfer of a good or favor to someone else, with future reciprocation expected but not guaranteed" (p. 50). Choosing a trustworthy partner that reciprocates trust can be beneficial to an individual, but incorrectly trusting a cheater can be extremely costly (Chang, Doll, van 't Wout, Frank, & Sanfey, 2010; Cosmides & Tooby, 1992). Cosmides and Tooby (1992) hypothesised that the need to accurately identify cheaters from co-operators would lead to the evolution of an adaptive cheater-detection mechanism. In support of this proposition, observers are better than chance at detecting cheaters in economic games (Verplaetse et al., 2007) and can distinguish trustworthy Nobel Peace Prize laureates from untrustworthy convicted felons (Porter, England, Juodis, ten Brinke, & Wilson, 2008), suggesting we can accurately identify untrustworthy individuals from just their facial appearance (Cosmides & Tooby, 1992). Moreover, individuals with "dishonest" faces are more likely to volunteer for studies that involve deceiving others (Bond, Berry, & Omar, 1994), and are more likely to deceive others in a cooperative economic game (Slepian & Ames, 2015; Stirrat & Perrett, 2010), suggesting that those with untrustworthy faces might truly be less trustworthy. However, accurate trustworthiness detection is not always observed. Zebrowitz, Voinescu, and Collins (1996) found no relationship between trustworthy facial appearance and clinical assessments of honesty, and Rule et al. (2013) found that observers could not correctly identify current business executives from convicted white-collar criminals. Although untrustworthy individuals can sometimes be identified from their facial appearance, it is clear that trustworthiness judgments are often unreliable (Cosmides & Tooby, 1992).

Even though there is limited evidence for the accuracy of trustworthiness judgments from the face, observers consistently choose to trust others that have a trustworthy appearance (Chang et al., 2010; Rezlescu et al., 2012; van 't Wout & Sanfey, 2008). Using an economic game, in which participants share money with virtual partners, Rezlescu et al. (2012) found that participants sent 42% more money to partners that had a trustworthy appearance, compared to partners that had untrustworthy faces. Interestingly, this advantage remained even when participants were given accurate summaries of the partner's past behaviour; when given identical histories of cheating, partners with trustworthy faces still received 6% more money than those with untrustworthy faces. Children also send more money to individuals that are rated to be highly trustworthy than those that are untrustworthy, suggesting that trustworthy appearance is used to guide cooperative decision-making from an early age (Ewing, Caulfield, Read, & Rhodes, 2015). Despite the questionable accuracy of trustworthiness judgments made from the face, observers are willing to pay money within the economic game to reveal a photograph of their partner (Eckel & Petrie, 2011; Ewing et al., 2015), which suggests that they believe that the face contains valid cues to the trustworthiness of their partner. Although the accuracy of these trustworthiness judgments remains an open question, it is clear that facial trustworthiness is a powerful social cue that can shape our cooperative behaviour toward others.

1.4 Attractiveness

Facial attractiveness is closely related to judgments of trustworthiness. Both are positive traits that are related to the valence dimension of trait perception (Oosterhof & Todorov, 2008), and are highly correlated with each other (Todorov, 2008). Despite their similarities, there is a key distinction to be made between these two judgments. Whereas trustworthiness is a judgment about an inferred personality trait, attractiveness is a judgment that is made about the aesthetic properties of the face. Interestingly, even though it is a judgment about the aesthetic qualities of the face, attractiveness is still associated with many positive stereotypes. In their seminal paper, Dion, Berscheid, and Walster (1972) found that attractive individuals were thought to be happier, have more socially desirable personalities, more prestigious jobs, and be more likely to marry than unattractive individuals. The attribution of favourable characteristics to attractive individuals is variously referred to as the "what is beautiful is good" or "attractiveness halo" stereotype, and is evident in the treatment of attractive people in the real world (Dion et al., 1972; Eagly, Ashmore, Makhijani, & Longo, 1991; Langlois et al., 2000). Attractive individuals receive more lenient criminal sentences (Sigall & Ostrove, 1975; Stewart, 1980), better academic grades (Landy & Sigall, 1974), more job offers (Gilmore, Beehr, & Love, 1986), higher salaries (Hamermesh & Biddle, 1993), more votes in elections (Stockemer & Praino, 2015), and go on more dates than unattractive people (Langlois et al., 2000). Even attractive children who misbehave are evaluated more positively than unattractive children who commit the same act (Dion, 1972). Attractiveness is clearly a valuable social commodity. But, what makes a face attractive in the first place?

In contrast to the conventional wisdom that "beauty is in the eye of the beholder", decades of research have shown that there is considerable agreement between individuals when rating facial attractiveness (Langlois et al., 2000). Males and females make very similar ratings of facial attractiveness, for both male and female faces (Langlois et al., 2000; Rhodes, 2006). Interestingly, even though identity (Chiroro & Valentine, 1995; Meissner & Brigham, 2001) and emotion recognition (Elfenbein & Ambady, 2003) are often reduced for other ethnicity faces, observers from different cultures nonetheless agree on facial attractiveness (Coetzee, Greeff, Stephen, & Perrett, 2014; Cunningham, Roberts, Barbee, Druen, & Wu, 1995; Langlois et al., 2000), suggesting that facial attractiveness is judged using similar standards around the world (Little, Jones, & DeBruine, 2011b). Furthermore, at just 70 hours old, infants spend longer gazing at faces that adults find highly attractive, compared to those rated to be unattractive (Langlois, Ritter, Roggman, & Vaughn, 1991; Slater et al., 1998). Crucially, these infants have not been exposed to cultural standards of attractiveness, which could potentially influence the ratings of adults, which strongly suggests that we are born with an innate understanding of the characteristics that make faces attractive. Taken together, the innate (Langlois et al., 1991; Slater et al., 1998), and universal nature of facial attractiveness (Cunningham et al., 1995; Langlois et al., 2000) strongly suggest that attractiveness has a biological origin (Little et al., 2011b; Thornhill & Gangestad, 1999).

Facial attractiveness is used by both male and female observers when selecting a potential mate (Buss, 1989; Buss & Barnes, 1986; Feingold, 1990; Rhodes, Simmons, & Peters, 2005). Because observers discriminate between potential mates based on their physical appearance, Thornhill and Gangestad (1999) hypothesised that mate choices have evolved to become sensitive to the attributes that serve as reliable cues to the genetic quality of a potential mate. Choosing a healthy mate has potential benefits not only for the individual (i.e., remaining parasite free), but also for the potential offspring that can inherit good genes and a strong immune system (Gangestad, Thornhill, & Yeo, 1994; Lie, Rhodes, & Simmons, 2008; Thornhill & Gangestad, 1999). Individuals that use honest cues in the face to identify mates of high genetic quality will gain a reproductive advantage over rivals, and will produce more offspring for the next generation (Thornhill & Gangestad, 1999). Over time, sexual selection will come to favour those traits that are valuable for mate selection (Little et al., 2011b; Thornhill & Gangestad, 1999). Although many idiosyncratic factors might make any particular face attractive, attributes that relate to the health of a potential mate are generally perceived to be desirable (Rhodes, 2006). Evolutionary psychologists have identified three cues in the face: *symmetry*, *sexual dimorphism*, and *averageness*, which signal the health of a potential mate and are perceived to be highly attractive (for comprehensive reviews see Little et al., 2011b; Rhodes, 2006).

Facial symmetry refers to how closely the left and right halves of the face resemble one another (Little et al., 2011b). Random deviations from symmetry, known as fluctuating asymmetries, occur throughout development due to random mutations or exposure to environmental stressors such as pathogens and toxins (Møller & Thornhill, 1998; Rhodes, 2006; Thornhill & Gangestad, 1999). Therefore, facial symmetry is suggested to be an indicator of the genetic quality of a potential mate, with greater symmetry indicating a stronger immune system that has proven resistant to environmental stressors through development (Møller & Thornhill, 1998; Thornhill & Gangestad, 1999). As hypothesised, facial symmetry shows negative correlations with self-reported instances of respiratory infection (Thornhill & Gangestad, 2006) and lifetime instances of serious disease in males (Waynforth, 1998), which suggests that symmetry is an honest cue to the potential health of a mate. Symmetry is positively correlated with attractiveness in natural faces (Mealey, Bridgstock, & Townsend, 1999; Penton-Voak et al., 2001; Scheib, Gangestad, & Thornhill, 1999), and faces that have been manipulated to be symmetrical are preferred to their original, naturally asymmetric images (Grammer & Thornhill, 1994; Perrett et al., 1999; Rhodes, 2006; Rhodes, Proffitt, Grady, & Sumich, 1998).

Despite containing the same structural features (e.g., two eyes, a nose, a mouth), male and female faces often look markedly different. Starting at puberty, the increase in levels of sex hormones leads to the development of *sexually dimorphic* traits in the face. Masculine facial features, such as a large jaw, brow, and facial hair, develop with the exposure to testosterone (Penton-Voak & Chen, 2004; Thornhill & Gangestad, 1999). Conversely, oestrogen inhibits the development of these characteristics, and feminine appearance is characterised by a small chin, high cheek bones, and large lips (Fink & Penton-Voak, 2002; Thornhill & Gangestad, 1999;

Thornhill & Møller, 1997). Therefore, one reason that sexually dimorphic appearance is hypothesised to be attractive is because these traits signal sexual maturity in a potential mate (Thornhill & Gangestad, 1999). Consistent with this prediction, feminine facial appearance is highly attractive in female faces (Perrett et al., 1998; Rhodes, 2006; Rhodes, Chan, Zebrowitz, & Simmons, 2003), and higher levels of oestrogen are positively correlated with feminine facial appearance, facial attractiveness and reproductive health (Smith et al., 2006). On the other hand, masculinity is attractive in natural male faces (O'Toole et al., 1998; Penton-Voak et al., 2001; Rhodes, 2006; Scheib et al., 1999), and is correlated with health during adolescence (Rhodes et al., 2003) and testosterone levels in adulthood (Penton-Voak & Chen, 2004). Curiously, however, sometimes a preference for feminised male faces is also found (Little & Hancock, 2002; Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000)[4]. This inconsistency has been attributed to the fact that masculine faces can be perceived as dominant and aggressive (Oosterhof & Todorov, 2008; Rhodes, 2006), whereas feminised male faces are perceived to be caring, honest, and better parents (Perrett et al., 1998).

Finally, in contrast to the common assumption that attractive faces must be distinct or unusual, faces that have *average* characteristics are perceived to be attractive (Langlois & Roggman, 1990; Rhodes, 2006; Rhodes & Tremewan, 1996; Rhodes et al., 2001; Valentine, Darling, & Donnelly, 2004). Perceptions of averageness are determined by how closely a face resembles a stored mental prototype, which represents the average of all faces previously encountered by the observer (Rhodes, 2006; Valentine, 1991). Averageness is commonly manipulated using face morphing techniques, whereby a series of individual faces are averaged together, to create a single composite face that has the average physical shape and colouration of the individual faces that it is made from. These digitally averaged faces are highly attractive, such

[4] As noted by Rhodes (2006) in her meta-analysis on the attractiveness of sexually dimorphic appearance, the methodology used to create the masculine face stimuli also has a large effect on the perceived attractiveness of masculine features in male faces. Studies using natural male faces tend to find a preference for masculinity (e.g., Penton-Voak et al., 2001), whereas studies that create masculine faces through digital average processes tend to find a preference for femininity (e.g., Perret et al., 1998).

that they are often rated to be more attractive than each of the individual faces from which they are composed (Langlois & Roggman, 1990; Rhodes, Sumich, & Byatt, 1999; Rhodes et al., 2001). Importantly, the attractiveness of average faces is not a result of the digital averaging process itself (Alley & Cunningham, 1991), as natural faces that have average characteristics are also found to be attractive (Rhodes et al., 1999; Rhodes & Tremewan, 1996; Valentine et al., 2004). Like facial symmetry, average facial features might signal that the individual possesses a heterogeneous immune system, which could offer any potential offspring resistance to a diverse array of diseases and parasites (Thornhill & Gangestad, 1993; Thornhill & Gangestad, 1999). Although average faces can be highly attractive (Rhodes, 2006), recent findings suggest that they are not maximally attractive. For example, despite being less average (i.e., closer to the population level prototype), an average face that is made from 15 highly attractive faces is perceived to be more attractive than an average face made from 60 random faces (DeBruine, Jones, Unger, Little, & Feinberg, 2007; Perrett, May, & Yoshikawa, 1994; Said & Todorov, 2011).

In conclusion, symmetry, sexual dimorphism and averageness are physical characteristics in the face that both signal the health of an individual, and are perceived to be attractive, which is consistent with the notion that facial attractiveness has been shaped by sexual selection (Gangestad et al., 1994; Grammer & Thornhill, 1994; Little et al., 2011b; Rhodes, 2006; Thornhill & Gangestad, 1999).

1.5 Attractiveness in Group Scenes

The vast majority of facial attractiveness research, and social perception research in general, has been conducted by presenting observers with a single face at a time (Barrett, Mesquita, & Gendron, 2011; Phillips, Weisbuch, & Ambady, 2014). Perhaps faces are presented in isolation because the cues that are used to guide attractiveness judgments (i.e., symmetry, sexual dimorphism and averageness), are physical qualities of the face that is being evaluated. However, in the real world, very rarely do we encounter faces in situations that are devoid of context. Often, we meet others for the first time in social settings, whether it be in a café, a classroom, or a bar. Early in the history of social perception, researchers understood that any useful model of social perception would need to describe how the trait judgments that are made for an individual change in response to social context, such as when a face is surrounded by other faces (Levy & Richter, 1963). However, enthusiasm for this idea was not widely shared, and as such, the existing social perception literature offers few answers about whether the same trait judgments are made for a face that is seen among a group of other faces. Interestingly, those few studies that have investigated this question, suggest that perceptions of facial attractiveness are modulated by social context.

The influence of social context on attractiveness judgments is demonstrated by "mate choice copying", a mate selection strategy that is prevalent in both human (Jones, DeBruine, Little, Burriss, & Feinberg, 2007; Place, Todd, Penke, & Asendorpf, 2010; Waynforth, 2007) and non-human species (Vakirtzis, 2011). Mate choice copying occurs when an individual perceives a potential mate to be more desirable upon learning that the potential mate is desirable to others (Gouda-Vossos, Nakagawa, Dixson, & Brooks, 2018; Little, Jones, DeBruine, & Caldwell, 2011c; Vakirtzis, 2011). Mate choice copying is used as a shortcut to identify high quality mates, by acting on the assumption that the potential partner must have desirable qualities if their current partner has chosen to be with them (Rodeheffer, Proffitt Leyva, & Hill, 2016). Although primarily investigated in females, who desire mate qualities that are not readily observed in the face (e.g., education, kindness; Buss & Barnes, 1986), mate choice copying is displayed by both male and female observers (Little, Burriss, Jones, DeBruine, & Caldwell, 2008; Place et al., 2010). Female observers find male faces to be significantly more desirable when shown with an attractive female compared to an unattractive female (Little et al., 2008; Little, Caldwell, Jones, & DeBruine, 2011a; Waynforth, 2007), while males display a preference for females shown with an attractive male partner (Little et al., 2008). Interestingly, Rodeheffer et al. (2016) found that females engaged in mate choice copying because they attributed more favourable personality characteristics to a male when he was seen with a female partner compared to alone. These findings are consistent with the suggestion that mate choice copying is used to infer the unobservable qualities (e.g., trustworthiness, kindness) that a potential mate might possess (Rodeheffer et al., 2016).

Although mate choice copying demonstrates that the attractiveness of the same face is modulated by social context, these findings are restricted to opposite-sex mate choices (Little et al., 2011c). However, early social perception research suggests that the influence of social context on attractiveness judgments is much more generalised. For example, Sigall and Landy (1973) investigated whether the positive stereotypes about attractive individuals would "radiate" from an attractive individual to somebody who is seen beside them. The participants arrived in the waiting room of the laboratory and were seated opposite a male and a female confederate. The female confederate was made to appear either attractive or unattractive. After passively observing the confederates for a short time, the experimenter retrieved the participant from the waiting room and asked about their impressions of the male confederate. Participants gave more favourable impressions of the male confederate when the female was highly attractive, compared to when she was unattractive. Similarly, Bar-Tal and Saxe (1976) reported that when male and female image pairs were described as married couples, impressions of the male were raised or lowered to match the attractiveness of the female. These findings were interpreted as showing an effect of attractiveness assimilation, wherein the attractiveness of the target face is raised or lowered to match the attractiveness of the second face.

However, assimilation effects are not always observed. In a creative design, Kenrick and Gutierres (1980) sent research assistants around the campus dormitories to find individuals who were watching *Charlie's Angels*, a television show that featured three highly attractive female lead characters. These participants were then asked to rate the attractiveness of a single female face, which was known to the researchers to be of average attractiveness. Compared to participants who were watching other television shows at the time, those watching *Charlie's Angels* gave the lowest attractiveness ratings to the test face. Kenrick and Gutierres's (1980) findings were consistent with an effect of contrast, wherein the perceived attractiveness of a face is biased away from the attractiveness of a previously seen face. Melamed and Moss (1975) speculated that the conflicting accounts of assimilation and contrast might be due to the perceived association between the images. Consistent with their hypothesis, assimilation was found when the two faces were presented as friends, whereas a contrast effect occurred when the same two faces were presented as strangers (Kernis & Wheeler, 1981; Melamed & Moss, 1975).

A second possible explanation for the conflicting accounts of contrast and assimilation relates to the many methodological limitations of these early studies (Geiselman, Haight, & Kimata, 1984). Often these studies relied on a very small number of trials or stimuli, increasing the likelihood that idiosyncrasies in the stimulus pairs could unduly influence the results (Bar-Tal & Saxe, 1976; Kenrick & Gutierres, 1980). Remarkably, other researchers chose to manipulate the attractiveness of real confederates over the course of weeks or months, which raises questions about the consistency of such a manipulation (Kernis & Wheeler, 1981; Sigall & Landy, 1973). To clarify the conditions under which contrast and assimilation effects occurred, Geiselman et al. (1984) presented target faces of average attractiveness between two distractor faces, which could be highly attractive, of average attractiveness or highly unattractive. Unlike previous investigations, the attractiveness of the faces had been pre-rated, and the attractive and unattractive distractor faces were equally distant in their attractiveness from the faces of average attractiveness. Regardless of the stated association between the faces in the group (friends or strangers), Geiselman et al. (1984) found an effect of assimilation, whereby target faces were significantly more attractive in the attractive group condition, and significantly less attractive in the unattractive group. Geiselman et al. (1984) did, however, show that assimilation occurred when the faces were presented simultaneously, but that a contrast effect occurred when the faces were presented sequentially (Wedell, Parducci, & Geiselman, 1987). Although assimilation effects seem to occur during simultaneous presentations, debate is still ongoing about whether the attractiveness of previously seen faces in sequential presentation paradigms leads to contrast

(Pegors, Mattar, Bryan, & Epstein, 2015) or assimilation effects (Xia, Leib, & Whitney, 2016). Despite the conflicting accounts of contrast and assimilation, these findings clearly demonstrate that the attractiveness of an individual is modulated by social context.

Of the few researchers to consider the effect of social context on attractiveness judgments, most have investigated how the attractiveness of an individual is changed by social context. Conversely, Willis (1960) investigated how observers judged the average attractiveness of a whole group of faces (i.e., "how attractive is this group?"). The groups were filled with individuals who were of similar attractiveness levels, such that the groups themselves were attractive or unattractive (Willis, 1960). Interestingly, rather than averaging the attractiveness of the individual group members (Anderson, 1965), observers gave more extreme ratings to the groups; attractive groups were rated to be more attractive than the mathematical average attractiveness of the individuals, whereas unattractive groups were rated to be less attractive (Willis, 1960). Recently, van Osch, Blanken, Meijs, and van Wolferen (2015) asked observers to rate either the attractiveness of the individual members of a group, or the average attractiveness of the whole group. Unlike Willis (1960), van Osch et al. (2015) found that observers only consistently overestimated the average attractiveness of the whole group, relative to the mathematical average of the attractiveness ratings given to the individual group members. van Osch and colleagues called this phenomenon the "*group attractiveness effect*", and demonstrated that it occurred for female, male, and mixed gender groups. Using eye-tracking, van Osch et al. (2015) showed that observers were preferentially attending to the most attractive faces in each group, which likely caused them to overestimate the average attractiveness of the whole group. Taken together, these findings demonstrate that social context not only influences the attractiveness of an individual, but the attractiveness of the group in which they appear.

1.6 Summary

Stable personality traits are inferred from the face spontaneously (Engell et al., 2007; Winston, O'Doherty, Kilner, Perrett, & Dolan, 2007; Winston et al., 2002), and are based on the similarities between the structure of the face and emotional expressions (Montepare & Dobish, 2003). Even though they are not highly accurate (Olivola & Todorov, 2010b), these impressions influence our decision making (Olivola et al., 2014; Rezlescu et al., 2012; Wilson & Rule, 2015, 2016). Trustworthiness judgments are closely related to the perceived emotion of a face (Oosterhof & Todorov, 2008), and are used to facilitate approach avoidance decision making (Todorov, 2008). Despite being an aesthetic judgment about the face, facial attractiveness is associated with many positive stereotypes (Dion et al., 1972; Langlois et al., 2000), and is crucial for mate selection (Buss, 1989; Buss & Barnes, 1986; Rhodes et al., 2005). The consistency with which facial attractiveness is judged between genders and cultures (Langlois et al., 2000), suggests that facial attractiveness has a biological basis. Facial attractiveness has likely been shaped by sexual selection to favour cues such as symmetry, sexual dimorphism and averageness, which are honest signals of the fitness of a potential mate (Little et al., 2011b; Rhodes, 2006; Thornhill & Gangestad, 1999).

In addition to the physical features of the face itself, recent findings have shown that facial attractiveness is influenced by social context (Geiselman et al., 1984; Jones et al., 2007; Little et al., 2011c; Waynforth, 2007). Even though these findings, at times, can be contradictory, this line of research clearly demonstrates that the perceived attractiveness of an individual is modulated by social context. At the most basic level, these findings show that even though the observer is asked to evaluate the attractiveness of an individual face in a group, the irrelevant faces around the target are also encoded, and interfere with the judgment that is made. This notion is consistent with the way that the visual system summarises complex visual scenes for efficient processing, in a process known as ensemble coding (Whitney, Haberman, & Sweeny, 2014; Whitney & Yamanashi Leib, 2018).

Chapter 2: Ensemble Perception

2.1 Complex Visual Scenes

Vast and intricately detailed, the visual world around us contains more detail than we could ever hope to process (Whitney et al., 2014). Subjectively, we have a rich visual experience; however, our visual system is remarkably limited. "Change blindness" occurs when we are unable to identify dramatic changes in visual scenes, which suggests that although we can see the full scene, very little of the visual environment is fully attended at any one time (Simons & Levin, 1997; Simons & Rensink, 2005). Visual working memory, the system used to actively hold information while we complete cognitive tasks (Baddeley, 1992; Brady, Konkle, Alvarez, & Oliva, 2008), has a capacity limit of approximately four items (Luck & Vogel, 1997). Moreover, object recognition in the peripheral visual field is inhibited by "crowding", which occurs when an individual item that could be easily identified alone is no longer distinguishable when surrounded by other items (Bulakowski, Post, & Whitney, 2011; Levi, 2008; Pelli, 2008; Whitney & Levi, 2011). The contradiction between our subjectively rich visual experience and the limitations of our visual system has led some to describe our visual experience as an "illusion of completeness" (Chong & Treisman, 2003; Noë, Pessoa, & Thompson, 2000). So, how does the visual system overcome these capacity limitations to provide us with a seemingly complete visual experience?

One way that the visual system overcomes these limitations is by relying upon accurate "gist" representations of the visual environment (Alvarez, 2011; Alvarez & Oliva, 2008). Despite containing an inordinate amount of visual detail, much of the visual information in our environment is structured and redundant (Alvarez, 2011; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018). Every blade of grass in a lawn is extremely similar to every other blade of grass; as are the leaves on a tree, or the trees in a forest. Although our visual system is capable of individuating the blades of grass in a lawn, often such a task offers little benefit to the observer (Haberman & Whitney, 2009). Instead, the visual system takes advantage of structure and redundancy in the environment and compresses information about similar objects or textures into a single percept, through a process known as ensemble coding[5] (e.g., blades of grass become a lawn, while individual trees become a forest; Alvarez, 2011; Whitney et al., 2014). By creating average textures, the cognitive demand associated with encoding vast numbers of individual stimuli is significantly reduced, and many of the limitations of the visual system can be bypassed (Alvarez, 2011; Whitney et al., 2014). Importantly, although the information about individual items can be lost during ensemble coding (Ariely, 2001; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), observers retain access to precise statistical information about the scene (Alvarez, 2011; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018).

2.2 Low-Level Stimuli

Ensemble coding was first demonstrated for low-level visual features, such as orientation, motion and size, which are processed in the earliest stages of the visual stream. In their seminal study, Parkes et al. (2001) presented sets of nine gabor patches to the observer's peripheral vision and asked them to identify the orientation of a single tilted target gabor in the display. Observers could accurately report the orientation of the target when it was presented alone; however, they were unable to do so when it was encircled by eight distractor gabor patches, which is consistent with the deleterious effect of crowding (Whitney & Levi, 2011). Remarkably, despite being unaware of the orientation of the crowded target gabor, Parkes et al. (2001) found that observers were able to accurately report the *average* orientation of the group. Crucially, because the target gabor displayed the greatest tilt, accurately identifying the average orientation of the group could only be achieved by successfully integrating the orientation of the crowded target with that of the distractors. Parkes and colleagues' (2001) findings demonstrated that observers retain access to information about the crowded item, and suggest that summary statistics are extracted automatically from the visual scene.

---

[5] Ensemble coding is described by many names in the literature, including ensemble perception, global processing, and summary statistics (Alvarez, 2011). The product of ensemble coding is also referred to by various names, including the ensemble average, ensemble representation, ensemble percept and summary representation.

In the same year, Ariely (2001) investigated whether observers were able to accurately perceive average size. First, Ariely (2001) created a set of four uniquely sized circles that were evenly distributed in size, such that two circles were smaller than the average size, while two were larger. Observers were presented with sets of 4, 8, 12, or 16 circles, which were created by presenting repetitions of the same 4 circles (see Figure 2.1). Observers were initially presented with a set of circles for 500 ms, which was immediately followed by a memory probe that consisted of a single circle. Based on size, observers had to indicate whether the probe circle had been presented in the previous set (i.e., yes/no). Curiously, instead of accurately identifying the sizes of the four circles that had been presented, observers were most likely to report "recognising" circles when they matched the average size of the set. Observers could also accurately report whether the probe circle was smaller or larger than the *average* size circle from the preceding set, demonstrating that ensemble coding is precise. Remarkably, observers could identify the average size of a set of 16 circles with the same precision as set size 4, indicating that sensitivity to the mean was not adversely influenced by increasing set size. Yet, despite the precision of their average size judgments, observers were no better than chance when asked to identify which of two circles had been shown in the previous set, suggesting that they had little awareness of the actual size of the individual circles in the group. Ariely's (2001) findings suggest that even when instructed to focus on the individual circles within the group, observers implicitly encode the mean size of the set.

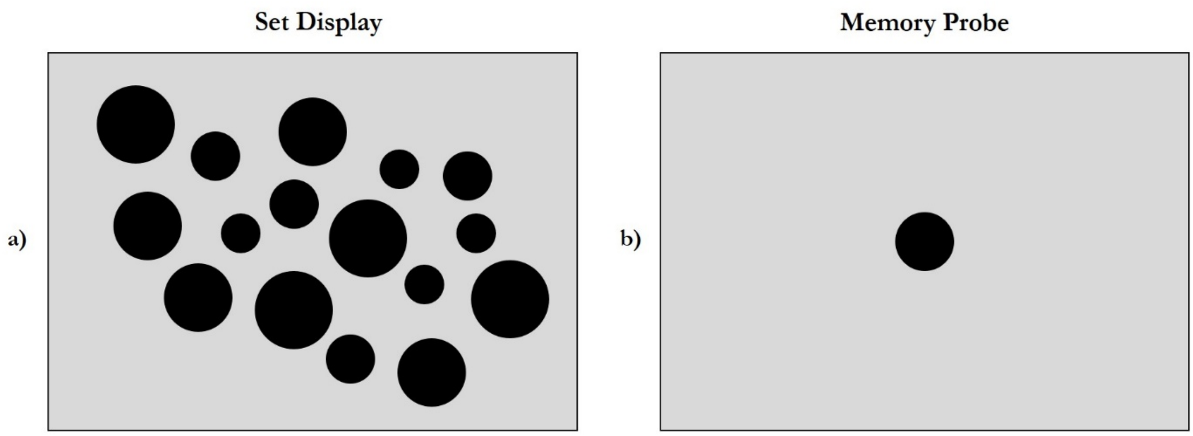

Figure 2.1. An example of Ariely's (2001) ensemble coding paradigm a) a set of 16 circles b) a memory probe, showing the mean size of the set. Figure adapted from Ariely (2001).

Through ensemble coding, observers rapidly extract the summary statics from a visual scene (Ariely, 2001; Parkes et al., 2001). These summary extractions are highly efficient, as the precision of average size judgments is not affected by viewing time, presentation method (i.e., simultaneous or sequential), or the distribution of circle sizes within the set (Chong & Treisman, 2003, 2005b). Moreover, Alvarez and Oliva (2008) showed that summary statistics could be extracted from items in the display that were outside the focus of attention. These displays contained four targets and four distractor shapes that moved independently from one another, which were presented on opposite sides of the display. Observers were less accurate to identify the final position of the individual distractor items compared to the individual targets, confirming that less attention was allocated to the individual distractors; however, observers could identify the centre point of the four distractor items with the same precision that they achieved for the four targets, suggesting that even with reduced attention, summary statistics are extracted from across the visual scene (Alvarez & Oliva, 2008, 2009).

To date, studies have shown that summary statistics are gleaned across a vast array of low-level visual domains, including size (Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b; Robitaille & Harris, 2011), orientation (Dakin & Watt, 1997; Parkes et al., 2001; Solomon, 2010), spatial location (Alvarez & Oliva, 2008, 2009), colour (Maule, Witzel, & Franklin, 2014; Webster, Kay, & Webster, 2014) and motion (Sweeny, Haroz, & Whitney, 2013; Watamaniuk & Duchon, 1992). Crucially, ensemble coding also occurs for sets of faces (Haberman & Whitney, 2007, 2009).

2.3 Ensemble Coding of Faces

Haberman and Whitney (2007) were the first to examine whether ensemble perception extended beyond lower-level visual stimuli to faces, which are higher-level visual representations that are processed later in the visual stream (Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006). Haberman and Whitney (2007) investigated whether observers could extract the average emotion from a set of faces. A continuum of emotional faces was created by morphing between two photographs of the same individual, one showing extreme happiness and one showing extreme sadness, and extracting a sequence of 50 faces that differed from one another in equal "emotional units"[6] (see Figure 2.2a). Observers were presented with a set of 4 emotional faces for 2000 ms, after which time a single probe face appeared. Observers then indicated whether the probe was happier or sadder than the *average* expression of preceding group. To establish baseline performance, observers were first presented with sets of identical emotional faces. Observers reached 75% accuracy when the probe differed from the homogenous group by 3-4 emotional units. Then, observers were presented with sets of 4 faces that differed in emotion from one another by 6 emotional units. Remarkably, observers reached 75% accuracy when the probe differed from the mean emotion of the set by 4 emotional units, replicating the precision of the baseline condition. These findings suggest that observers can identify the mean emotion of a group with the same accuracy as they can identify the emotion of a single face. In their seminal paper, Haberman and Whitney (2007) also found that observers can identify the mean

[6] As noted by Haberman & Whitney (2007, 2009), the term "emotional units" does not mean that each face represented a distinct emotional expression. Rather, the continuum of 50 faces showed the possible expressions in the transition between the happiness and sadness.

gender of a group of faces, demonstrating that the ensemble coding of faces is not limited to expressions.

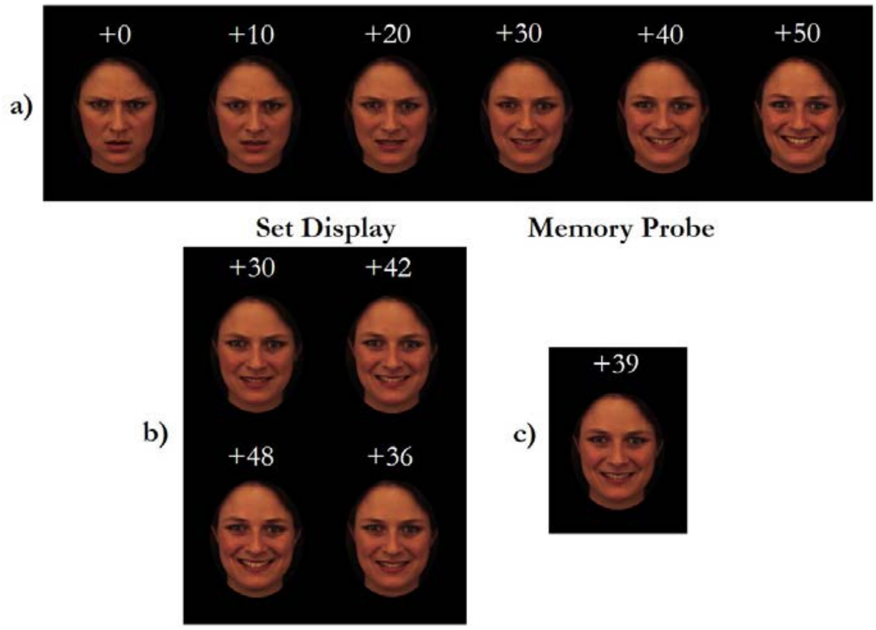

Figure 2.2. An example of Haberman and Whitney's (2007, 2009) ensemble coding paradigm a) a sequence of emotional faces created by morphing between an individual showing anger (+0) and happiness (+50) b) a set of four emotional faces c) a memory probe, showing the mean expression of the set. Figure adapted from Haberman and Whitney (2009).

Haberman and Whitney (2009) further investigated the ensemble coding of emotion using a paradigm reminiscent of Ariely (2001). Observers were asked to focus on the emotions shown by individual faces that were presented in sets of 4, 8, 12, and 16. Like Ariely (2001), increasing set size was achieved by presenting repetitions of the same four emotional faces. When presented with a single probe face, observers responded to whether the probe was a member of the preceding group. Observers were most likely to endorse faces that showed the mean emotion of the set, even though no such face was presented, and were no better than chance at identifying which of two faces was previously presented in the group, suggesting that they had little awareness of the individual faces in the group (Haberman & Whitney, 2007, 2009).

Sensitivity to the average emotion of the set was not affected by increasing set size or decreasing stimulus presentation time. The extraction of average emotion is not limited to simultaneous presentations paradigms, as Haberman, Harp, and Whitney (2009) found that regardless of the number of faces in the set (4, 12, or 20), observers could accurately identify the mean expression of a set of faces that was presented sequentially. As was the case for simultaneous presentations (Haberman & Whitney, 2009), the accuracy of mean emotion extraction was equivalent for sets of heterogeneous and homogenous faces that were presented sequentially, further demonstrating that ensemble coding of emotion is highly accurate.

Rapidly extracting the average emotion from a crowd might offer adaptive advantages to an individual (Haberman & Whitney, 2009). Rather than needing to evaluate the angry emotion on each face in a crowd, an observer can rapidly summarise that the average expression of the crowd is one of anger, and quickly escape from what is actually an angry mob. Yet, crowds are dynamic, and until recently, ensemble coding of faces had only been demonstrated for static emotional faces (Haberman & Whitney, 2007, 2009). Elias, Dyer, and Sweeny (2016) investigated whether ensemble coding could also be used to average dynamic facial expressions in crowds that changed emotion synchronously or asynchronously. Observers could accurately identify the average emotion of the crowd when the faces changed emotion synchronously; however, they were less precise when the faces changed emotions asynchronously, suggesting that the coordinated behaviour of the crowd can affect the precision of the ensemble average. Similarly, Haberman, Lee, and Whitney (2015b) demonstrated that observers could accurately identify the variance of emotional expressions in a crowd, but became less accurate as the variance in the crowd increased. Sensitivity to the variance of emotion in a crowd might allow an observer to quickly gauge whether the whole crowd is angry, or whether the crowd only contains a single angry individual (Haberman et al., 2015b). Interestingly, when a set of faces contains a single emotional outlier, observers appear to exclude it from their ensemble percept, possibly to ensure that the summary representation more accurately reflects the average composition of the group (Haberman & Whitney, 2010).

In addition to displaying emotional expressions, the face is also vitally important for identity recognition. The recognition of emotion and identity are based on fundamentally different cues; emotion recognition depends on the perception of temporary facial movements, whereas identity recognition relies on viewpoint invariant information (Bruce & Young, 1986; Haxby & Gobbini, 2011). de Fockert and Wolfenstein (2009) were the first to investigate whether observers could extract the average identity from sets of unfamiliar faces (see Figure 2.3). After being presented with a group consisting of four different identities, observers were shown a single probe face, which could be: a face from the previous group (set exemplar), a face not from the group (a foil), the average face of the group (set average), or the average face of a different group (foil average). Observers showed greater recognition for set exemplars than for foils, suggesting that they retained awareness of the individual faces in the group; however, they also mistakenly "recognised" the set average as having been a member of the group. Importantly, the foil average was rarely endorsed, indicating that the increased recognition of set averages was not due to the general familiarity of average faces (Langlois, Roggman, & Musselman, 1994). Instead, these findings suggest that observers summarise the average identity of the set, and mistake it for a face that has previously been seen (de Fockert & Wolfenstein, 2009).

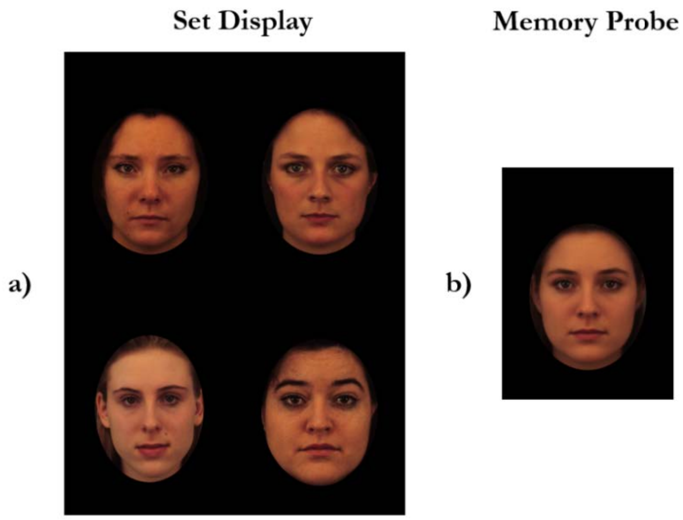

Figure 2.3. An example of de Fockert and Wolfenstein's (2009) ensemble identity coding paradigm a) a set consisting of four unique identities b) a memory probe, showing the average face of the set. Figure adapted from de Fockert and Wolfenstein (2009).

The ensemble coding of identity is not limited to unfamiliar faces (de Fockert & Wolfenstein, 2009), as observers also mistakenly recognise the average face for groups consisting of famous faces (Neumann, Schweinberger, & Burton, 2013) and groups composed of different photographs of the same person (Kramer, Ritchie, & Burton, 2015). Observers are remarkably sensitive to the average identity of the set, such that when instructed to identify whether the probe was an *image* from the previous group, they endorse the average face of the set, but not an average face created from different photographs of the set members (Kramer et al., 2015; Neumann et al., 2013). Like emotion, average identity can also be extracted from sets of faces presented sequentially (Yamanashi Leib et al., 2014). In their sequential presentation paradigm, Yamanashi Leib et al. (2014) presented observers with sets of faces that had be photographed to show the model posing with their face oriented 22.5° or 90° to the left or right of centre; however, the memory probe only showed faces that were directly facing the camera. Even though none of the individual faces had been seen facing forward, observers could accurately recognise the average identity of the group from forward facing memory probes. This finding strongly suggests that observers are able to integrate identity across viewpoints and images to form a single accurate ensemble representation of identity (de Fockert & Wolfenstein, 2009; Neumann et al., 2013; Yamanashi Leib et al., 2014).

In contrast to the advantage offered by summarising facial expressions in crowds, the benefit of averaging identity is less clear. Rather, averaging identity might even be detrimental to the recognition of familiar individuals (Neumann et al., 2013). One possibility is that averaging identity facilitates identity learning (Burton & Jenkins, 2011; Jenkins & Burton, 2011; Kramer et al., 2015), as average faces that are composed of different photographs of the same person are recognised with greater accuracy than individual exemplars (Burton, Jenkins, Hancock, & White, 2005). In particular, ensemble coding of identity might help children learn new faces rapidly, at a time when their individual face recognition skills are poor (Germine, Duchaine, & Nakayama, 2011). Ensemble coding abilities emerge early in development, with children as young as 4 years of age being able to accurately average size (Sweeny, Wurnitsch, Gopnik, & Whitney, 2015). By the age of 8, children also show the ability to average across identity, endorsing both set exemplars and the set average as members of the previously presented group (Rhodes et al., 2017). Rhodes et al. (2017) also found a positive correlation between the age of the observer and ensemble coding ability, even when individual face recognition ability was controlled, indicating that ensemble coding of identity might develop separately from individual face recognition.

Very recently, attention has turned toward investigating whether ensemble coding extends beyond emotion and identity recognition to influence the other judgments that can be made from faces. Starting by morphing between an extremely unattractive and an extremely attractive face, to create a sequence of faces that varied on attractiveness, Luo and Zhou (2018) investigated whether observers could extract the average attractiveness from a group of faces. Observers were found to extract the average attractiveness of a set of four heterogeneous faces with the same precision that they demonstrated for sets of four identical faces (Luo & Zhou, 2018). Moreover, the accuracy of ensemble coding was unaffected by the variance of attractiveness between the individual faces in the group. Furthermore, Phillips, Slepian, and Hughes (2018) found that observers could also accurately identify the average variance for sets of faces along dimensions of ethnicity, gender, and dominance. Together, these findings demonstrate that ensemble coding can be used to summarise a diverse array of judgments from faces.

Even though observers extract precise averages of expression (Haberman et al., 2009; Haberman & Whitney, 2007, 2009), identity (de Fockert & Wolfenstein, 2009; Neumann et al., 2013; Yamanashi Leib et al., 2014), and attractiveness (Luo & Zhou, 2018) from sets of faces, they are often unable to identify the individual faces that were presented in the group (Haberman & Whitney, 2007, 2009). Haberman and Whitney (2011) employed a change blindness paradigm (Simons & Levin, 1997), to investigate whether awareness for the individual faces in the group was required for accurate ensemble coding. The same set of 16 emotional faces was presented to observers twice in rapid succession. In the second set, however, four faces had changed expression, which also changed the average emotion of the group. Observers were asked to indicate whether the first or second set was happier, and then to identify one face in the second set that had changed expression. Remarkably, even when observers could not identify a single face that had changed expression, they could still accurately identify the happier group; demonstrating that observers could display accurate ensemble coding while simultaneously experiencing change blindness (Haberman & Whitney, 2011). Similarly, in a paradigm reminiscent of Parkes et al. (2001), Fischer and Whitney (2011) found that even when observers could not identify the expression of an individual face due to crowding, they could nonetheless accurately report the mean emotion shown by the group. Together, these findings suggest that summary statistics are automatically extracted from sets of faces, even when observers cannot attend to the individual faces in the group (Fischer & Whitney, 2011; Haberman & Whitney, 2007, 2009, 2011).

Even though observers have little awareness for the faces in the group (Fischer & Whitney, 2011; Haberman et al., 2009; Haberman & Whitney, 2009, 2011), summary statistics are likely calculated by averaging many faces from the entire set (Haberman & Whitney, 2010; Whitney & Yamanashi Leib, 2018; Yamanashi Leib et al., 2014). In a meta-analysis across low-level and high-level paradigms, Whitney and Yamanashi Leib (2018) found that number of items integrated into the ensemble representation was approximately equal to the square root of the total number of items in the display. Consistent with this estimate, Yamanashi Leib et al. (2014) showed that observers were averaging at least 4 faces to extract the average identity from a set of 18 faces ($\sqrt{18} = 4.24$). Remarkably, the precision of summary statistics often *increases* with set size, such that observers can more accurately identify the mean of the set than they can identify an individual face (Ariely, 2001; Chong & Treisman, 2003; Elias et al., 2016; Haberman & Whitney, 2009, 2010; Robitaille & Harris, 2011; Sweeny et al., 2013; Yamanashi Leib et al., 2014). While initially appearing counterintuitive, the greater precision of ensemble coding relative to individual coding is consistent with the benefits of averaging (Alvarez, 2011; Ariely, 2001). There is random noise associated with the encoding of any individual face (Haberman et al., 2015b); however, this noise is uncorrelated between the encoding of different individual faces. Averaging cancels out the uncorrelated noise associated with encoding the individual exemplars, which results in a summary representation that is more precise than any encoded individual exemplar (Alvarez, 2011).

The extraction of summary statistics occurs rapidly, even for large sets of faces (Haberman et al., 2009; Yamanashi Leib et al., 2014). While ensemble coding paradigms typically present sets of faces simultaneously for 1500 to 2000 ms, accurate ensemble representations can be formed when individual faces are presented sequentially for as little as 23 ms (Haberman et al., 2009; Yamanashi Leib et al., 2014), which is less time than required to make a reliable trait impression for a single face (Bar et al., 2006; Todorov et al., 2009). Simultaneous presentation paradigms have shown that ensemble coding of emotion can occur after a set of 16 faces is presented for only 50 ms, which is far below the time required to encode each face individually (Crouzet, Kirchner, & Thorpe, 2010). Although average representations are less precise at such short durations, they are still more precise than individual exemplar recognition (Haberman & Whitney, 2009; Li et al., 2016; c.f. Neumann, Ng, Rhodes, & Palermo, 2017). Recognition accuracy for the set average and individual exemplars continues to increase until presentation times are between 800 and 1600 ms (Li et al., 2016; Neumann et al., 2017). For set presentations of greater than 1600 ms, a dissociation occurs, wherein individual exemplar recognition continues to increase, but endorsement of the set average declines (Li et al., 2016; Neumann et al., 2017), suggesting that summary statistics might be discarded once individual representations are established.

The ensemble coding of faces shares many of the same characteristics as the ensemble coding of low-level visual stimuli; both are precise, rapid, unaffected by set size, and occur without awareness of the individual items in the set. However, several converging lines of evidence suggest that, rather than relying on low-level image cues, the ensemble coding of faces relies upon face specific processing mechanisms. When asked to indicate whether a probe face is the same *identity* as a face from the previous set, observers endorse both the set average, and an average face made from different photographs of the identities in the set (Neumann et al., 2013). Observers cannot rely on averaging low-level image features to endorse set averages made from different photographs, because those photographs had not previously been seen (Yamanashi Leib et al., 2014). Second, observers are less accurate to identify the average emotion of a crowd when the set of faces is inverted (Elias et al., 2016; Haberman & Whitney, 2009; Yamanashi Leib et al., 2012). Although the low-level features of the image remain the same when inverted, inversion disrupts the holistic processing of faces (Farah, Tanaka, & Drain, 1995; Tanaka & Farah, 1993), which significantly inhibits the recognition of faces relative to other visual stimuli (Valentine, 1988; Yin, 1969).

Further evidence that the ensemble coding of faces involves face specific processing comes from examining the ensemble coding abilities of clinical samples that traditionally have difficulty with face recognition (Rhodes, Neumann, Ewing, & Palermo, 2015; Robson, Palermo, Jeffery, & Neumann, 2018; Yamanashi Leib et al., 2012). Congenital prosopagnosia is characterised by face recognition deficits (Palermo et al., 2011; Susilo & Duchaine, 2013), while autism spectrum disorder is a neurodevelopmental condition that often results in poor face recognition abilities (Dawson, Webb, & McPartland, 2005). Although individuals with these conditions demonstrate ensemble coding by endorsing the average identity of the group as a set member, they do so at a significantly lower rate than neurotypical controls (Rhodes et al., 2015; Robson et al., 2018; c.f. Yamanashi Leib et al., 2012). The ensemble coding deficits shown by individuals with congenital prosopagnosia or autism spectrum disorder were directly related to their impaired ability to accurately identify individual faces (Rhodes et al., 2015; Robson et al., 2018), suggesting that efficient ensemble coding might be predicated on individual exemplar coding ability (Neumann et al., 2017).

Ensemble coding occurs during the processing of both low-level (Ariely, 2001; Chong & Treisman, 2003, 2005b) and high-level visual stimuli (Haberman & Whitney, 2007, 2009; Neumann et al., 2013). Haberman, Brady, and Alvarez (2015a) investigated whether the visual system operates a single generalised ensemble coding mechanism that performs ensemble coding across low- and high-level stimuli, or whether multiple ensemble processors are used. Observers completed four separate tasks: judging the orientation of an individual gabor patch, the mean orientation of a set of four gabors, the identity of an individual face, and the average identity for a set of four faces. Performance was strongly correlated within visual domains, such that better performance on the individual orientation task predicted better performance on the mean orientation task, and accuracy on the individual identity task predicted better performance on the mean identity task. However, performance on the orientation tasks did not predict performance on the identity tasks. Haberman et al. (2015a) also found that performance on different ensemble coding tasks with low-level visual features were highly correlated with one another (e.g., average orientation and average colour), as were tasks requiring high-level discriminations (e.g., average expression and average identity), but there was no correlation between performance on the low-level and high-level ensemble coding tasks. For this reason, Haberman and colleagues (2015a) suggest that the visual system operates at least two ensemble coding mechanisms, one for low-level stimuli, and one for high-level stimuli.

2.4 The Hierarchical Structure of Visual Working Memory

Regardless of the psychophysical method used to measure ensemble coding, ensemble coding tasks rely on the observer being able to recall the characteristics of the preceding set of stimuli. Thus, these tasks rely on observers engaging their visual working memory. Although there is still debate about the exact capacity of visual working memory, there is strong agreement that working memory is severely limited (Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 1997, 2013), particularly in comparison to the massive storage capacity of long-term memory (Brady et al., 2008). Whilst conventional models of working memory assume that individual items are encoded separately to one another (Zhang & Luck, 2008), recent findings suggest that visual working memory is constructive, and encodes the individual items of the group in reference to the summary statistics of a display (Brady & Alvarez, 2011, 2015a, 2015b; Brady & Tenenbaum, 2013; Griffiths, Rhodes, Jeffery, Palermo, & Neumann, 2018; Lew & Vul, 2015; Walker & Vul, 2014).

Brady and Alvarez (2011) were the first to investigate whether ensemble coding influenced how observers stored individual items in working memory. Observers saw a display of nine uniquely sized circles, three of which were blue, green, and red. On any one trial, the circles were grouped by colour and by size (e.g., small = red, medium = green, large = blue). To make the colour of the circles salient, the observers were instructed to study the sizes of the individual red and blue circles in the set, but to always ignore the green circles. Following set presentation, a probe circle was presented at the location of the individual circle that was to be recalled. Unbeknownst to the observer, each display was repeated twice in the experiment. These matched trials were identical in every way, except that the colour of the target circle had been switched (blue ↔ red). Consequently, observers estimated the size of the same target circle twice; once when it was the same colour as the small circles, and once as the same colour as the large circles. Brady and Alvarez (2011) found that the same circle was recalled as significantly larger when it was the same colour as the large circles, but smaller when it was the same colour as the small circles. In each condition, the same circle was remembered to be a size that was 15-20% closer to the mean size of the same colour circles than it truly was. Brady and Alvarez's (2011) findings demonstrate that visual working memory encodes the components of a scene in a hierarchical structure, whereby the individual items in the group are encoded in reference to the summary statistics of the display. Consequently, the hierarchical structure of visual working memory biases the recall of individual items from the scene, such that they are remembered as being more alike the ensemble average than they really were.

Although very similar, there is a critical distinction between ensemble coding and hierarchical encoding paradigms. In an ensemble coding paradigm, the observer compares the probe to the whole display (e.g., "was this face present in the previous display?") or makes a judgment about the average characteristics of the group (e.g., "does this face show the mean expression of the group?"). Conversely, in a hierarchical encoding paradigm, the observer compares the probe to a single item that was cued during the presentation of the group display (e.g., "was this face the previously cued target?"). Hierarchical encoding, therefore, specifically refers to the recall of a single item from the set. Notably, hierarchical encoding occurs because, in addition to encoding the individual items within a scene, visual working memory also encodes the summary statistics of a scene. Therefore, ensemble coding is required for hierarchical encoding to occur. Although it is clear that observers can very accurately extract the average emotion from a group through ensemble coding (Haberman & Whitney, 2007, 2009), very little is known about the way that hierarchical encoding influences the recall of an individual face from a group.

To date, only Griffiths et al. (2018) have investigated how hierarchical encoding influences the recall of the emotional expression shown by an individual in a group. From a morphed sequence of emotional faces, Griffiths et al. (2018) presented observers with sets of 4 emotional faces for 2000 ms. The emotional faces in the group varied around an average emotion that was not shown. After the set was removed from the screen, a cue was presented at the location of a single face in the group, which identified the target face to recall. A single probe face then appeared, and observers made a forced choice response to indicate whether the probe showed the same expression as the target. Griffiths et al. (2018) found that although observers could quite accurately remember the emotion shown by the target face, their responses were biased toward the mean emotion of the group. Faces showing expressions less intense than the group average were recalled as having shown more intense emotion, whereas faces showing more intense emotion were recalled as being less emotional. Griffiths and colleagues' (2018) findings are consistent with those of Brady and Alvarez (2011) and demonstrate that the hierarchical structure of visual working memory also biases the recall of an individual's emotional expression toward the mean emotion of the group.

2.5 The Cheerleader Effect

The literature discussed above provides a comprehensive review of the ensemble coding of faces. These studies have overwhelmingly investigated whether identity and emotion can be represented as summary statistics. Despite the significant attention that facial attractiveness has received in the social perception literature, only one study has investigated whether facial attractiveness is subject to ensemble coding (Luo & Zhou, 2018). Yet, several years before Luo and Zhou (2018) showed that ensemble coding occurs for facial attractiveness, Walker and Vul (2014) investigated whether a phenomenon referred to in popular culture as "*the cheerleader effect*", was an example of hierarchical encoding in attractiveness judgments. The cheerleader effect was first described on the television series *How I Met Your Mother* (Rashid & Fryman, 2008). The show's protagonist, Barney Stinson, claimed that women (at a bar) appeared to be much more attractive when they were sitting together in a group, compared to when they were sitting alone. Realising that the cheerleader effect might be consistent with an effect of hierarchical encoding, Walker and Vul (2014) investigated whether the cheerleader effect was indeed a real phenomenon[7].

Walker and Vul (2014) tested the existence of the cheerleader effect in a within-participants paradigm, wherein observers were required to make two attractiveness ratings for the same face; one when it was presented in a group, and one when it was presented alone. In the group presentation, observers were presented with a set of three faces from a group photograph (see Figure 2.4). After 2000 ms, a cue appeared around a single face in the group for an additional 1000 ms. The entire image then disappeared from the screen, and observers gave an attractiveness rating for the cued target face. During the alone trials, the same target face was presented in isolation for 2000 ms. The cheerleader effect was calculated by subtracting the attractiveness ratings given to the face when it was seen alone, from the rating given to the same face when it was seen in a group. Consistent with the cheerleader effect, Walker and Vul (2014) found that both male and female faces were perceived to be significantly more attractive in a group than when seen alone. The size of the cheerleader effect remained constant for faces shown in groups of 4, 9, and 16 faces. Furthermore, the cheerleader effect occurred for faces within naturally occurring group images, as well as for faces presented side by side to create a composite group (as in Figure 2.4).

[7] I am grateful to both Dr Drew Walker and Dr Ed Vul for sharing the story of how their cheerleader effect research came to be with me.

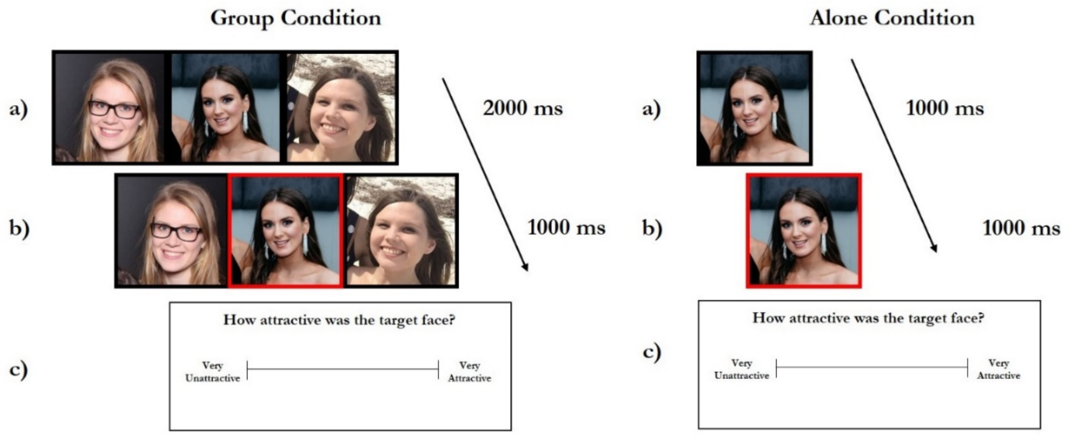

Figure 2.4. An example of Walker and Vul's (2014) cheerleader effect paradigm a) the group stimulus is initially presented un-cued b) the target face is cued from the group c) the attractiveness rating is made along a visual analogue scale.

Walker and Vul (2014) hypothesised that the cheerleader effect was the result of a complex interaction between ensemble perception, the attractive characteristics of average faces, and the hierarchical structure of visual working memory. When presented with a group of faces, the observer rapidly creates an ensemble average (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009; Neumann et al., 2013). Ordinarily, the ensemble average is an accurate representation of the average characteristics of the group (e.g., emotion, identity). However, averageness is highly attractive in faces (Rhodes, 2006), and digitally averaged faces are often perceived to be significantly more attractive than the individual faces from which they are composed (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001). Walker and Vul (2014) proposed that much like digitally averaged faces, the ensemble average is also perceived to be highly attractive, because it possesses the average characteristics of the faces in the group. Finally, due to the hierarchical structure of visual working memory, the observer recalls the individual face as being more similar to the ensemble average than it truly was (Brady & Alvarez, 2011; Griffiths et al., 2018). Walker and Vul (2014) hypothesise that because the ensemble average is more attractive than the individual faces in the group, each individual is recalled as being more attractive in a group than when they were alone, ultimately giving rise to the cheerleader effect. Crucially, although the cheerleader effect is consistent with such a mechanism, Walker and Vul (2014) did not directly test their proposed hierarchical encoding mechanism.

2.6 Current Aims

The overarching aim of this thesis was to build upon the cheerleader effect paradigm to explore the changes in social perception that occur when an individual face is presented in a group, compared to when they are seen alone (Walker & Vul, 2014). My first aim was simply to investigate whether the cheerleader effect could be replicated. Despite the considerable media attention that Walker and Vul's (2014) article received, neither a replication, nor an extension, of the cheerleader effect had been published since their initial findings[8]. Second, I investigated whether the cheerleader effect is subject to visuospatial asymmetries of attention, which are known to influence gaze behaviour and perceptions of attractiveness. Third, I investigated whether the cheerleader effect extends beyond aesthetic attractiveness judgments to trait judgments of trustworthiness. Finally, I investigated whether the cheerleader effect is, in fact, caused by the hierarchical encoding mechanism proposed by Walker and Vul (2014).

---

[8] One non-replication of the cheerleader effect was published by Ojiro et al. (2015) in *The Quantitative Methods for Psychology*. Briefly, Ojiro et al. (2015) attempted to replicate Experiment 4 of Walker and Vul (2014), which investigated whether the size of the cheerleader effect increased in sets of 1, 4, 9, and 16 faces. Walker and Vul (2014) found a main effect of group size, but reported that the size of the cheerleader effect did not differ between groups of 4, 9, and 16 faces. Following the method of Walker and Vul (2014), Ojiro et al. (2015) twice failed to replicate the main effect of group size (although in one experiment the effect neared significance, $p = .06$). Notably, Ojiro et al. (2015) failed to report the one-sample t-tests that are necessary to determine whether a significant increase in attractiveness has occurred in any group condition (although visual inspection of the figures suggests that the basic cheerleader effect was not significant). However, closer examination of the article history of Ojiro et al. (2015) shows that the article was published just one week after being received by the journal, raising the possibility that it was not subjected to peer-review. As such, the findings of Ojiro et al. (2015) have been omitted from our published work, and from this thesis.

Chapter 3: Is Trustworthiness Lateralised in the Face?

Citation

The entirety of this chapter has been published: Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. (2017). Is trustworthiness lateralized in the face? Evidence from a trust game. *Laterality: Asymmetries of Body, Brain and Cognition*, 23(1), 20-38, doi: 10.1080/1357650x.2017.1298120.

The article has been reproduced here in full. Authors NAT and MERN contributed to the conceptualisation of the current project and edited the manuscript.

Rationale

While the final aim of this thesis was to investigate the influence of social context on social perception, the two experiments reported in the following chapter reflect the original line of research that I had planned to conduct in this thesis. The dominance of the right hemisphere in the expression and perception of emotion (Borod et al., 1998), causes emotional expressions to be displayed asymmetrically on the face (Indersmitten & Gur, 2003; Lindell, 2013b), as greater movement occurs on the left side of the face than the right (Nicholls, Ellis, Clement, & Yoshino, 2004; Sackeim, Gur, & Saucy, 1978). The asymmetric expression of emotion can influence the perception of emotion, such that the left side of the face is perceived to be happier than the right (Zaidel, Chen, & German, 1995). Starting with trustworthiness, I planned to investigate whether these emotional asymmetries would influence the trait judgments that are made from the face.

To foreshadow the results of this chapter, I found no evidence that trustworthiness was lateralised in the face. Due to the robustness of these non-significant results, I reconsidered the viability of this original research project. Ultimately, this line of research was not continued, and I turned my attention toward understanding the nature of social perception in group scenes.

Abstract

A turn of the head can be used to convey or conceal emotion, as the left side of the face is more expressive than the right. As the left cheek moves more when smiling, the present study investigated whether perceived trustworthiness is lateralised to the left cheek, using a trust game paradigm. In Experiment 1, participants were asked to share money with male and female "virtual partners". Left-left or right-right composite faces were used to represent the partners. There were no differences in the amount shared based on composite face, suggesting trustworthiness is not lateralised in the face. However, there was a robust effect whereby female partners were perceived to be significantly more trustworthy than males. In Experiment 2, the virtual partners presented either the left or the right cheek prominently. As in Experiment 1, the amount shared with the partners did not change depending on the cheek presented. Interestingly, female partners were again sent significantly more money than males. We found no support for lateralised trustworthiness in the face, suggesting that asymmetries in the face are not large enough to influence trustworthiness judgments. Instead, more stable facial features, such as sex-typical characteristics appear to influence perceived trustworthiness.

Introduction

Despite the cautionary reminder not to "judge a book by its cover", we regularly make social judgments about individuals that are based solely on their facial appearance. These social judgments are used to infer the possible attributes that an individual may have, such as how competent or trustworthy they may be, and can subsequently influence our decision-making. For example, the perceived competence of politicians vying for election affects voting behaviour (Antonakis & Dalgas, 2009; Praino, Stockemer, & Ratis, 2014; Todorov et al., 2005). The attractiveness of criminals influences court room sentencing (Stewart, 1980), and the perceived trustworthiness of murderers predicts whether they receive a life sentence or are given the death penalty (Wilson & Rule, 2015, 2016). In reality, it is unlikely that competent appearance makes a better politician, or that individuals who appear untrustworthy actually commit more gruesome crimes. While the initial social judgments that we make are based upon a rapid first impression, these judgments are slowly updated over time to accurately reflect the observed behaviour of the individual (Chang et al., 2010).

First impressions occur rapidly, and trait impressions occur reliably after a face has been seen for just 39 ms (Bar et al., 2006; Willis & Todorov, 2006). Importantly, judgments of traits such as trustworthiness, attractiveness and dominance are consistent both within observers, and between individuals (Todorov et al., 2015). Judgments of trustworthiness incorporate stable facial attributes such as attractiveness and dominance (Oosterhof & Todorov, 2008), and also transient emotional expressions such as smiling (Krumhuber et al., 2007), suggesting that trustworthiness judgments represent a general valence approximation of the face, and act as an extension of approach-avoidance decision making (Todorov, 2008). Despite their automatic nature (Engell et al., 2007; Winston et al., 2002), the accuracy of trustworthiness judgments is only just better than chance (Kovacs-Balint, Bereczkei, & Hernadi, 2013; Verplaetse et al., 2007). Any reliable cues to trustworthiness in the face, such as facial width (Carré & McCormick, 2008; Stirrat & Perrett, 2010), are likely overlooked or over interpreted by perceivers (Olivola &

Todorov, 2010b). For instance, participants are willing to pay to reveal a photograph of their partner in a trust game (Eckel & Petrie, 2011; Ewing et al., 2015), suggesting that participants believe that facial appearance will improve the accuracy of trustworthiness judgments. Yet, when trustworthy and untrustworthy faces are given identical histories of untrustworthy past behaviour, participants still share more money with trustworthy faces (Rezlescu et al., 2012). This line of research suggests that even though trustworthiness judgments are not always accurate, trustworthy appearance continues to bias decision-making in spite of more reliable cues, such as reputation or previous behaviour.

Trustworthy appearance is influenced by smiling (Oosterhof & Todorov, 2008; Scharlemann et al., 2001), as faces showing either posed or genuine smiles are perceived as more trustworthy than those showing no expression (Krumhuber et al., 2007). Emotional expressions, such as smiles, are displayed asymmetrically on the face (Lindell, 2013b), which can have important implications for the perception of emotion. The left side of the face is more expressive than the right, and produces greater movement when displaying emotion (Borod, Kent, Koff, Martin, & Alpert, 1988; Lindell, 2013b; Nicholls et al., 2004; Sackeim et al., 1978). There is a well-documented preference for subjects to present the more expressive left cheek in photographs and painted portraits (Lindell, 2013b; McManus & Humphrey, 1973; Nicholls, Clode, Wood, & Wood, 1999b; Powell & Schirillo, 2009), suggesting that a turn of the head is used to control the communication of emotion. Indeed, Nicholls et al. (1999b) found participants presented their left cheek when asked to pose for an emotional photograph, but presented the right cheek when asked to conceal emotion. The findings of Nicholls et al. (1999b) indicate that we have an innate understanding of the communicative role of the left and right sides of the face, and deliberately present one cheek depending on our motivation.

Photographs that present the left cheek are judged to be more emotionally expressive than those showing the right (Nicholls, Wolfgang, Clode, & Lindell, 2002), which can have important implications for the social judgments that are made from first impressions of faces.

Asymmetries in the expression and perception of emotion are commonly investigated using composite faces, wherein one hemiface is mirror-reversed and fused with its original orientation, to create a symmetrical face that consists of two left or two right hemifaces (Indersmitten & Gur, 2003; Okubo, Ishikawa, & Kobayashi, 2013; Sackeim et al., 1978). For example, smiling left-left composite faces are rated as happier (Zaidel et al., 1995), while right-right composites are judged to be more attractive (Zaidel et al., 1995). This dissociation suggests that the left and the right hemifaces can influence the perception of emotions and stable traits differently. Although smiling and attractiveness have been shown to be lateralised in the face (Zaidel et al., 1995) and both contribute to impressions of trustworthiness (Oosterhof & Todorov, 2008), there have been few investigations into whether trustworthiness is lateralised within the face. Zaidel, Bava, and Reis (2003) asked participants to select the most trustworthy composite face, and found that trustworthiness was not lateralised to either cheek. However, this result must be interpreted with caution, as participants could indicate within a force-choice task that both faces were "the same" perceived trustworthiness, and selected the "same" option on approximately 25% of trials. While Zaidel et al. (2003) suggest trustworthiness might not be lateralised in the face, the inclusion of the "same" response affects the strength of the conclusions that can be drawn.

More recently, Okubo and colleagues (Okubo et al., 2013; Okubo, Ishikawa, Kobayashi, & Suzuki, 2016; Okubo, Kobayashi, & Ishikawa, 2012) investigated the lateralisation of trustworthiness in the face, specifically for objectively trustworthy and untrustworthy individuals. Okubo et al. (2016) asked male participants to pose for a photograph that would be shown to other individuals in a trust game, before partaking in the trust game themselves. The trust game is a commonly used paradigm to investigate trustworthiness, in which participants are asked to share money with one another (Berg, Dickhaut, & McCabe, 1995). Okubo et al. (2016) found that the most cooperative participants were equally likely to show either cheek in their photograph. In contrast, untrustworthy participants (i.e., males with the highest instances of deception in the game) were more likely to present the left cheek rather than the right.

Potentially, untrustworthy individuals are motivated to appear as trustworthy as possible to encourage cooperation from their partners. This suggestion implies that untrustworthy individuals present the left cheek because it appears more trustworthy than the right. In a separate task in which a new sample of participants rated the trustworthiness of the previously collected images, untrustworthy males were rated as significantly more trustworthy when showing the left cheek compared to the right (Okubo et al., 2016). Additionally, Okubo et al. (2013) found that left-left composite images of untrustworthy males were judged as significantly more trustworthy than right-right composites. In contrast, there was no trustworthiness asymmetry for images of trustworthy males where one cheek was featured more prominently (Okubo et al., 2016) or for composite faces (Okubo et al., 2013). Therefore, unlike Zaidel et al. (2003), research by Okubo et al. (2013; 2016) suggests that trustworthiness could be lateralised in the face.

Prompted by the findings of Okubo et al. (2016), we investigated whether trustworthy appearance is lateralised in the face using a trust game. Okubo and colleagues (2012; 2013; 2016) asked participants to rate the trustworthiness of images, whereas we used a trust game paradigm. Trust games are highly sensitive to changes in trustworthy appearance (Rezlescu et al., 2012), and provide a more ecologically valid means of assessing perceived trustworthiness than asking participants for trait ratings of stimuli. Additionally, Okubo et al. (2016) found that untrustworthy males presented their left cheek in a trust game, which suggests that asymmetries in posing direction are used as a form of communication during this task. We created composite face stimuli to investigate whether trustworthiness is lateralised in the face. The present study is the first to investigate whether lateralised trustworthiness influences trusting behaviour using a trust game paradigm. We hypothesised that left-left composite images would be sent more money than right-right composite faces.

To extend upon the work of Okubo et al. (2012; 2013; 2016), we also included images of female partners in the trust game. Females are more emotionally expressive (Nicholls, Clode, Lindell, & Wood, 2001) and consistently show stronger left posing biases than males (Lindell, 2013b). Therefore, it is possible that the emotional asymmetries that influence trustworthy appearance in males might be more apparent amongst females. As this factor has not been examined previously by Okubo and colleagues (2012; 2013; 2016), we did not make a directional hypothesis in relation to sex of the partner, but including it as an exploratory variable.

Experiment 1

Method

Participants

Forty participants (M_{age} = 21.4, *SD* = 7.03, 20 females) were recruited from Flinders University. Participants received course credit for their participation, as well as a "bonus" payment of $5.00 AUD (see *Trust Game* below). As participants were required to be right-handed as defined by the FLANDERS handedness survey (Nicholls, Thomas, Loetscher, & Grimshaw, 2013), one left-handed participant was excluded. An additional participant was excluded for using a predetermined strategy in the trust game (Rezlescu et al., 2012), where they declined to share money with the majority of partners (> 81% trials). The final sample consisted of 38 participants (M_{age} = 21.37, *SD* = 7.19, 20 females). The Flinders University Social and Behavioural Research Ethics Committee granted ethical approval.

Apparatus

The experiment was presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), interfaced with a 22" monitor with a refresh rate of 60Hz (1680 x 1050 px). Stimuli were viewed at a distance of 500 mm, and head movements were minimised by using a chin-rest. Responses were made via an accessory numerical keypad. The hand used to respond on the keypad was counterbalanced between participants. Participants were monitored while they completed the experiment through a live video feed in the testing room to ensure they stayed on task.

Stimuli

Composite faces. Composite faces are commonly used to investigate emotional asymmetries between the left and right sides of the face (Okubo et al., 2013; Sackeim et al., 1978). In the present study, we created composite faces to investigate whether trustworthiness was lateralised in one cheek. To create the composite faces, we first sourced original images from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist, Flykt, & Öhman, 1998). In the original images, both female and male models were asked to pose expressions of happiness, while positioned directly in front of the camera.

We created our composite stimuli by dividing each original face along the vertical midline, to separate the left and right hemifaces. Each hemiface was then copied and mirror-reversed, before being realigned with the original. This procedure resulted in a pair of symmetrical left-left and right-right composite faces, for each of the original KDEF models (Okubo et al., 2013; Sackeim et al., 1978). A grey oval-shaped mask was fitted to each composite image to occlude external facial features that may reveal obvious signs of symmetry. If either of the composite faces for a model appeared unrealistic, both composite images were excluded from the final stimulus set. This procedure ensured each identity had both a left-left and right-right composite image, so that any difference observed in the experiment was the result of the expressivity of the faces – rather than having different models as left-left and right-right composites.

The final stimulus set contained 88 unique faces: a left-left and a right-right composite image for 22 models of each gender. Each image was presented twice during the experiment, for a total of 176 trials. The composite stimuli were randomised into 4 blocks of 44 faces. The left-left and right-right composites of each model always appeared in separate blocks, and all 88 stimuli were seen once before being presented a second time. Within each block, the stimuli appeared in random order.

Trust Game. Berg et al. (1995) introduced the trust game paradigm (or Investment Game) to investigate both *trust* and *trustworthiness*. In the original paradigm, "Player A" is endowed with $10 and can send any amount to "Player B". The amount shared by Player A is multiplied by a fixed factor before Player B receives it. Player B can then choose to keep the entire sum, or to return any amount back to Player A. The amount sent from Player A to Player B is the measure of *trust*, while the amount returned to Player A from Player B is the measure of *trustworthiness*. As is often the case, our trust game was modified slightly, and was similar to that used by Stanley, Sokol-Hessner, Banaji, and Phelps (2011) and Chang et al. (2010).

All of our participants played the trust game as Player A, and were provided with a cover story about Player B (the partner), which was used to increase the realism of the task (see also Stanley et al., 2011). The cover story closely followed the standard procedure for the trust game, and informed the participants that the photographs shown during the task were individuals who completed a previous experiment as Player B. Recall that we created composite faces using images from the KDEF database to present to participants as their partners in the trust game. In reality, we only measured the amount of money sent from the participant to each partner. The partners did not interact with participants by actually playing the trust game, nor did participants receive feedback about the behaviour of their partner on each trial.

The cover story informed participants that they would have 10 virtual dollars to share with each virtual partner (in whole dollar increments). The sum shared by the participant would be multiplied by 4 and given to the virtual partner, who had previously indicated whether they would keep the entire sum for themselves or split the funds evenly with Player A. This modified payoff structure meant that sharing money with a trustworthy partner (i.e., one who returned the money) would double the participant's money, while sending money to an untrustworthy partner (i.e., one who kept the money) would result in a loss of funds.

Participants were also informed there was an opportunity for a "bonus" cash payment at the end of the experiment, that would be equal to the amount a partner returned to them on one randomly selected trial (Stanley et al., 2011). Therefore, giving money to a trustworthy partner would result in a bonus payment, whereas no bonus would be given for sending money to an untrustworthy partner. We included the bonus payment in our cover story to encourage participants to examine the photographs of their partners and to actively evaluate their trustworthiness. All participants received the bonus payment at the conclusion of the experiment.

Procedure

Following consent, participants received written and verbal instructions, which included a knowledge test about the rules of the trust game and the bonus payment, before completing two practice trials. The instructions given to the participants did not include the terms "trust" or "trustworthiness". At the start of each trial, a face was presented for 3000 ms. A prompt reminding participants they had \$10 to share with the partner was then presented for 1500 ms. Participants then indicated how much money they would share with the partner. Invalid responses (i.e., amounts > \$10) were automatically rejected, and had to be re-entered by the participant. The FLANDERS handedness survey (Nicholls et al., 2013) was always completed at the end of the experiment to avoid making the lateralised nature of the task salient. The experiment took 30 minutes to complete.

Results

The amount of money shared with the virtual partners was analysed using a mixed-model analysis of variance (ANOVA), in which composite image (left-left, right-right) and sex of the virtual partner (female, male) were within-participant factors, and participant sex (female, male) was a between-participant factor. There was a significant main effect of sex of the virtual partner, $F(1,36) = 55.55, p < .001, \eta_p^2 = .607$, as participants shared more money with female ($M = 4.44$, $SD = 1.54$) than male ($M = 3.32$, $SD = 1.78$) partners. Critically, the main effect of composite image was non-significant, $F(1,36) = 1.95, p = .172, \eta_p^2 = .051$. Participants shared a similar amount of money with left-left (*M* = 3.85, *SD* = 1.58) and right-right (*M* = 3.91, *SD* = 1.62) composite faces. The main effect of sex of the virtual partner is evident in Figure 3.1, which also shows the non-significant effect of composite image. All remaining main effects and interactions failed to reach significance (*F*'s < 1.52, *p*'s > .227).

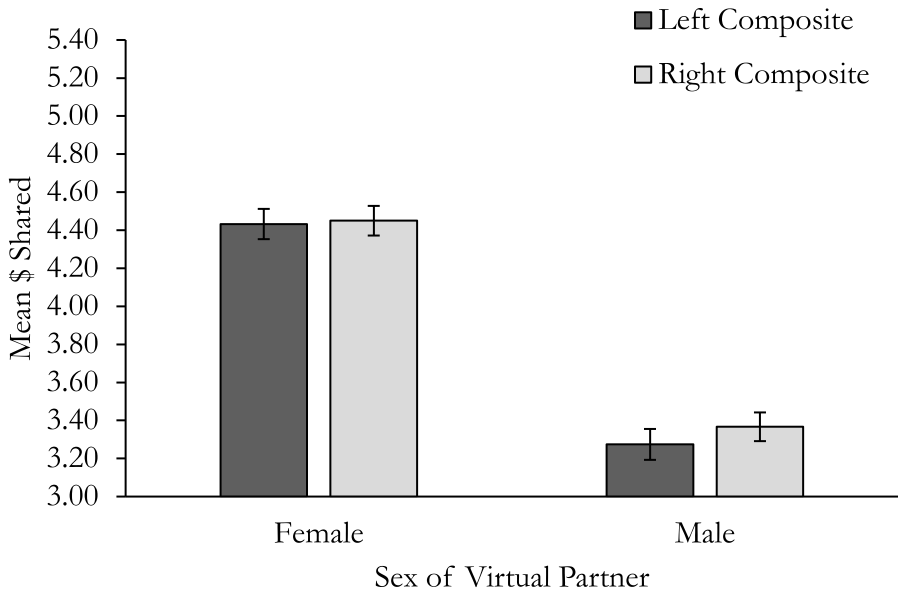

Figure 3.1. The amount of money shared with each partner separated by sex of the virtual partner and composite image. Within-participant error bars represent the standard error of the mean (Cousineau, 2005).

Discussion

When tasked with sharing money with virtual partners, participants shared a similar amount of money with left-left and right-right composite faces. Our results indicate that trustworthiness is not lateralised in the face, as we failed to observe any differences in the perceived trustworthiness of composite faces. This finding is consistent with Zaidel et al. (2003), who did not find a difference in perceived trustworthiness for composite faces. However, our findings are at odds with Okubo et al. (2013) who found that left-left composites of untrustworthy individuals were more trustworthy than their right-right composites. Okubo et al.

(2016) also showed that untrustworthy males were more likely to present the left cheek to their partners. Importantly, this result occurred amongst males, who typically exhibit weaker left posing biases than females (Lindell, 2013b), suggesting that untrustworthy individuals intuitively offer the left cheek when trying to appear trustworthy. It could be the case that asymmetries in relation to trustworthiness become apparent when people alter their pose to show a cheek prominently.

An unexpected, but robust, effect was observed in relation to the sex of the virtual partner. A female trustworthiness premium was observed by which participants shared approximately 25% more money with females than males, indicating that female partners were perceived as more trustworthy. Our finding appears to be novel in the trust game, as typically one gender of stimuli has been used in this context to simplify the design (i.e., Rezlescu et al., 2012). Our finding is consistent with prior research examining the influence of stable masculine facial appearance on perceived trustworthiness. The facial width to height ratio is a sexually dimorphic trait (though see Lefevre et al., 2012), by which the width of the lower face is influenced by testosterone exposure (Carré & McCormick, 2008; Stirrat & Perrett, 2010; Weston, Friday, & Liò, 2007). Facial width influences the perception of masculinity and dominance in the face; two traits negatively correlated with trustworthy appearance (Oosterhof & Todorov, 2008; Todorov et al., 2008). Typically, masculine faces are perceived as dominant and less trustworthy than feminine faces (Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Consequently, male faces often receive lower ratings of trustworthiness than female faces (Wincenciak, Dzhelyova, Perrett, & Barraclough, 2013), a finding replicated in the present study. The reduced trust in male partners observed in the present study may be the result of masculine facial appearance signalling dominance, making males appear less trustworthy than their female counterparts (Oosterhof & Todorov, 2008; Todorov et al., 2008).

Experiment 2

Previous research has shown that people alter their pose when attempting to either convey or conceal emotion (Nicholls et al., 1999b), because the left side of the face is more emotionally expressive than the right (Borod et al., 1988; Nicholls et al., 2004). The left side of the face moves more during a smile (Borod et al., 1988), and smiling is related to trustworthy appearance (Krumhuber et al., 2007; Oosterhof & Todorov, 2008). Interestingly, untrustworthy individuals are more likely to pose showing the left cheek, which is perceived as more trustworthy than the right (Okubo et al., 2016).

We investigated whether a difference in perceived trustworthiness would be evident for images of individuals predominantly showing either the left or the right cheek. In contrast to Experiment 1, which used composite images, faces showing one cheek prominently appear more naturalistic. Perhaps, as Burt and Perrett (1997) argue, the unnatural symmetry created in composite stimuli influenced the perceived trustworthiness of the faces shown in Experiment 1. We predicted that more money would be shared when partners posed the left cheek compared to the right cheek; although, no difference between the posing conditions would replicate our results from Experiment 1, and strongly suggest that trustworthiness is not lateralised in the face. Finally, it was also predicted that female partners would be sent more money than males, replicating the strong effect of the female trustworthiness premium observed in Experiment 1.

Method

Participants

Forty participants (M_{age} = 24.38, SD = 8.81, 20 females) were recruited from Flinders University. Participants received $20.00 AUD: $15.00 AUD for their participation and $5.00 AUD as a bonus payment (as in Experiment 1). All participants were right-handed as defined by the FLANDERS handedness survey (Nicholls et al., 2013). Three participants were excluded for using a predetermined strategy to share the maximum amount of money with the majority of partners (> 98% of trials) (Rezlescu et al., 2012). The final sample included 37 participants (M_{age} = 23.57, *SD* = 8.28, 19 females).

Apparatus, Stimuli & Procedure

The apparatus, trust game, and procedure were identical to Experiment 1, and the images selected for the virtual partners in the trust game were again of models expressing happiness from the KDEF database (Lundqvist et al., 1998). To test the effect of prominent cheek on trustworthy appearance, the images selected for the present experiment had been photographed simultaneously on 45° angles to the left and right of the centre of the model's face (Lundqvist et al., 1998). These images differed from the centrally photographed images used in Experiment 1. Mirror-reversals of each image were included in the stimuli set to ensure that any trustworthiness asymmetry observed between the left and right sides of the face was indeed caused by the emotional expressivity of the face itself, rather than aesthetic preferences or perceptual asymmetries on behalf of the viewer (Burt & Perrett, 1997). However, we also included image orientation as a factor in our analysis to explore whether perceptual asymmetries influenced perceived trustworthiness.

Participants saw each model 4 times, once in each possible orientation: left cheek original, left cheek mirror-reversal, right cheek original, and right cheek mirror-reversal. The stimuli were randomised into 4 blocks of 70 trials. Within each block, one image of each model was presented and stimuli were presented in a random order. The experiment had 280 trials and took 45 minutes to complete.

Results

The amount of money shared with the virtual partners was analysed using a mixed-model ANOVA, in which prominent cheek (left, right), orientation (original, mirror-reversed) and sex of the virtual partner (female, male) were within-participant factors, and participant sex (female, male) was a between-participant factor. The main effect of sex of the virtual partner was significant, $F(1,35) = 66.47$, $p < .001$, $\eta_p^2 = .655$. As in Experiment 1, participants shared more money with female (*M* = 5.07, *SD* = 1.33) than with male (*M* = 4.06, *SD* = 1.38) partners (see Figure 3.2). Critically, the main effect of prominent cheek was non-significant, $F(1,35) = 1.35$, $p = .253$, $\eta_p^2 = .037$. Participants shared a similar amount of money with partners when either the left (*M* = 4.55, *SD* = 1.30) or the right (*M* = 4.58, *SD* = 1.31) cheek was presented. As indicated in Figure 3.2, there was no evidence that left cheek poses attract larger investments than right cheek poses. The main effects of orientation and participant sex failed to reach significance, as did all lower order interactions (*F*'s < 2.47, *p*'s > .125).

The four way interaction between prominent cheek, orientation, sex of the virtual partner and participant sex was significant, $F(1,35) = 6.66$, $p = .014$, $\eta_p^2 = .160$. Analyses conducted to interpret the interaction showed that when female participants interacted with female virtual partners, there was a significant interaction between prominent cheek and orientation, $F(1,18) = 9.37$, $p = .007$, $\eta_p^2 = .342$. A significant perceptual asymmetry on behalf of the viewer would be indicated by a significant asymmetry in the original presentation orientation that was reversed in the mirror-reversal orientation (Dunstan & Lindell, 2012). Yet this was not the case; following Bonferroni corrections for multiple comparisons ($p = .025$), both post hoc paired-samples *t*-tests were non-significant. As such, we believe the observed four factor interaction was spurious.

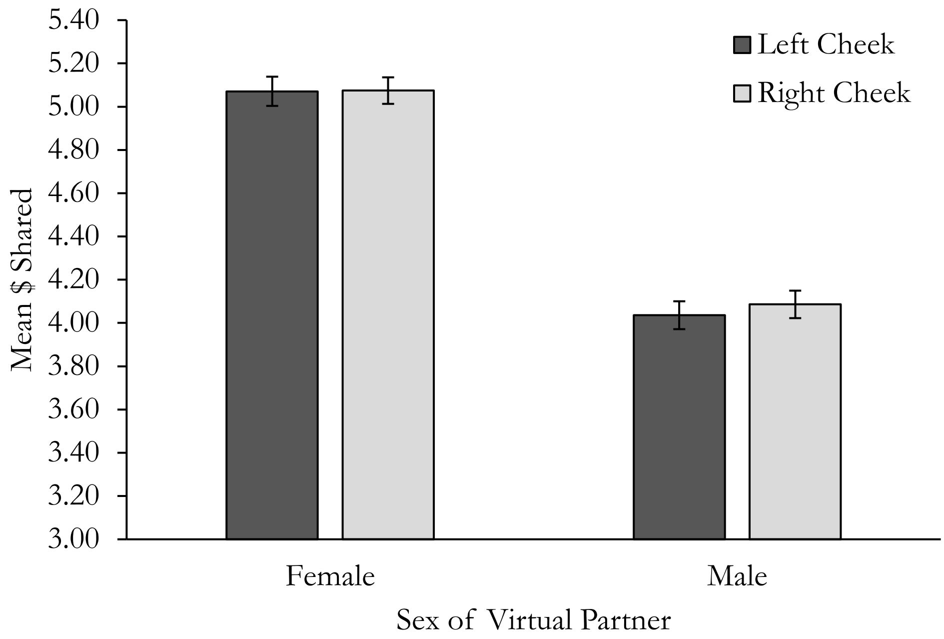

Figure 3.2. The amount of money invested with each partner separated by sex of the virtual partner and prominent cheek. Within-participant error bars represent the standard error of the mean (Cousineau, 2005).

Discussion

The use of real faces in lieu of composite faces did not change our results; there was no evidence that the left cheek was perceived as more trustworthy than the right. Collectively, our findings from Experiments 1 and 2 indicate that trustworthiness is not lateralised in the face. Interestingly, however, more money was once again shared with female compared to male partners. On average, females were sent approximately 20% more money than males. Across both experiments, our findings clearly show that females were seen as more trustworthy and receive a female trustworthiness premium.

General Discussion

The present study, which was the first to directly investigate the lateralisation of trustworthiness using a trust game, found no evidence that one side of the face is perceived as more trustworthy. Contrary to our hypothesis, the left cheek was not perceived as more trustworthy than the right, for either composite faces or for faces that featured one cheek more prominently than the other. Our results are consistent with Zaidel et al. (2003), who also reported no differences in perceived trustworthiness using composite stimuli. As we replicated these findings using real faces in Experiment 2, the lack of a significant difference in Experiment 1 was not the result of using artificial composite faces. Rather, our results support the suggestion that trustworthiness is not lateralised in the face (Zaidel et al., 2003). However, our results are at odds with Okubo et al. (2013), who found that amongst untrustworthy males, left-left composites are rated as more trustworthy than right-right composites. Furthermore, Okubo et al. (2016) found that untrustworthy males were more likely to present their left cheek in a photograph for a trust game, which was subsequently judged to be more trustworthy than showing the right cheek. The findings of Okubo and colleagues (2013; 2016) suggest that trustworthiness is lateralised in the face of objectively untrustworthy males.

We believe that the inconsistency between our results and those of Okubo et al. (2013) can be explained by prior research into the relationship between physical characteristics of the face and objective measures of trustworthiness. Facial width is a sexually dimorphic trait (Weston et al., 2007) that is related to perceived dominance and therefore untrustworthy appearance (Oosterhof & Todorov, 2008), whilst being correlated with both untrustworthy and aggressive behaviour in males (i.e., objective trustworthiness) (Carré & McCormick, 2008; Carré, McCormick, & Mondloch, 2009; Stirrat & Perrett, 2010). Males with wider faces are more likely to deceive their partners in a trust game (Stirrat & Perrett, 2010). Consistent with Stirrat and Perrett (2010), Okubo et al. (2013) reported that composite faces of untrustworthy males were wider than those of trustworthy males. Taken together, these findings suggest that untrustworthy individuals may have wider faces than trustworthy individuals. Wider faces might be prone to larger emotional asymmetries than typical faces, which could, in turn, influence the strength of emotional asymmetries.

The trustworthiness asymmetry reported by Okubo and colleagues (2013; 2016) is only apparent for *objectively* untrustworthy individuals, suggesting it is imperative for future research to use stimuli in which the objective trustworthiness of the photographed individuals is known. Presumably, any trustworthiness asymmetry that exists for untrustworthy individuals would be masked in a sample of faces that also includes trustworthy individuals (as the KDEF presumably does). As the faces used in the present study came from the KDEF database (Lundqvist et al., 1998), the objective trustworthiness of the models in the photographs is unknown (i.e., it is impossible to determine whether or not these individuals are actually trustworthy in the real world). As such, it is beyond the scope of the present study to investigate whether a trustworthiness asymmetry is only apparent for untrustworthy individuals.

Future research would be well served to identify objectively trustworthy and untrustworthy individuals, before re-examining the possibility that trustworthiness is lateralised in the face, using the design of the present study. Presently, this suggestion is limited because the trustworthiness of the individuals in face databases is very rarely known. In future, it will be important to investigate whether subjective ratings of trustworthiness can be used as a proxy for objective trustworthiness. In order to do so, it will be necessary to obtain information about the actual behaviour of the individuals who were being photographed, and to then obtain subjective ratings of trustworthiness. An interesting possibility could be that emotional asymmetries are exaggerated in individuals that are subjectively rated as very untrustworthy, as well as those who are objectively untrustworthy (Okubo et al., 2013).

Judgments of perceived trustworthiness are influenced by a myriad of features within the face, including those that are temporary and those that are stable. Though Zaidel et al. (1995) found that left-left composite faces had larger smiles, the same left-left composites were no more trustworthy than right-right composites (Zaidel et al., 2003), suggesting that the asymmetries in smiling displays are not strong enough to alter the perceived trustworthiness of composite faces. Furthermore, the results of Zaidel et al. (2003) also suggest that trustworthy appearance is not contingent on smiling. This suggestion is supported by research showing that faces without emotional expressions are also rapidly judged for trustworthiness (Engell et al., 2007). The dissociation between smiling and trustworthiness suggests that stable facial features may be more important in facilitating rapid judgments of trustworthiness than temporary expressions like smiles. This suggestion is consistent with the hypothesised role of trustworthiness judgments in approach-avoidance decision-making (Todorov, 2008). Importantly, a smile is one of the easiest facial expressions to fake (Ekman & Friesen, 1982; Okubo et al., 2012). An untrustworthy individual posing a fake smile could easily thwart an approach-avoidance mechanism if this decision relies too heavily on cues that can be faked. Rather, it would appear that approach-avoidance decision-making and consequently, judgments of trustworthiness, rely on stable traits that are not easily faked, such as dominance and facial width.

Cosmides and Tooby (1992) hypothesised that humans have evolved a specific mechanism that assists in identifying untrustworthy individuals, and indeed, trustworthiness judgments appear to have a biological basis. The amygdala shows an automatic response to the trustworthiness of faces (Engell et al., 2007), even when the trustworthiness of the face is task irrelevant (Winston et al., 2002). The activation of the amygdala is non-linear, such that untrustworthy faces elicit stronger responses than trustworthy faces (Said et al., 2009a; Todorov, 2008). The activation pattern of the amygdala suggests that the function of trustworthiness judgments is not to distinguish between individuals with highly trustworthy appearance; rather it is to avoid untrustworthy individuals. Perhaps the cost of mistakenly interacting with an untrustworthy individual is greater than that of missing the opportunity to co-operate with a highly trustworthy individual. Therefore, while the left cheek can pose a larger smile (Borod et al., 1988; Zaidel et al., 1995), the increase in positive expression due to a smiling asymmetry may not be large enough to influence the already established trustworthiness judgment (Zaidel et al., 2003).

In both of our experiments, female partners were given more money than males, suggesting that females were perceived as more trustworthy. This female trustworthiness premium was worth approximately 25% in Experiment 1 and 20% in Experiment 2. As noted above, this finding appears novel in the context of a trust game, because stimuli are typically selected on the basis of trait attributions, such as being trustworthy (Rezlescu et al., 2012) or attractive (Wilson & Eckel, 2006), rather than sex differences. Our finding is consistent with gender differences that have been shown in two player trust games, in which the participants did not know the gender of their partner. Females return a greater proportion of money sent to them by their partner (Buchan, Croson, & Solnick, 2008); whereas males are more likely to lie to their partners to increase their own payoff (Dreber & Johannesson, 2008). These findings suggest that the female trustworthiness premium may occur simply because females are more trustworthy than males in this setting. However, when only a name is used to reveal the partner's gender, there is no advantage for partners with female names relative to those with male names (Buchan et al., 2008). If the female trustworthiness premium were the result of learning that females are more trustworthy than males, a partner with a female name should be trusted more than a male. As this does not appear to be the case (Buchan et al., 2008), we suggest that feminine facial appearance is required to observe the female trustworthiness premium, shown in the present study.

As discussed above, trustworthy appearance is related to the perception of trait dominance, which in turn varies with masculine facial appearance (Oosterhof & Todorov, 2008; Todorov et al., 2013). Consequently, female faces are rated as more trustworthy than male faces (Wincenciak et al., 2013). Trustworthy appearance strongly influences decision making, as Rezlescu et al. (2012) found that facial trustworthiness continues to influence decision-making in the trust game, even when participants are given information about the past behaviour of the individual. We suggest that the female trustworthiness premium is consistent with the preference for trustworthy appearance observed by Rezlescu et al. (2012). Rather than being the result of females actually being more trustworthy than males, the female trustworthiness premium likely reflects the sexually dimorphic relationship through which facial width influences perceptions of dominance and consequently untrustworthy appearance. Although we cannot be certain whether the female trustworthiness premium reflects experiences in which females are actually more trustworthy or that feminine faces are perceived as more trustworthy, our results clearly show that gender of a seen partner can significantly influence the behaviour of the participants in the trust game.

Across two iterations of the trust game we found no evidence for lateralised trustworthiness, using both composite faces, and images in which the model posed the left or right cheek. Instead, we suggest that stable characteristics, such as masculinity and dominance, appear to influence perceived trustworthiness more than asymmetries in emotional expressions such as smiling. Though the left side of the face is more expressive, perhaps the function of trustworthiness judgments is not to distinguish between slight differences in trustworthy appearance, which may be induced by emotional asymmetry, but rather to avoid untrustworthy individuals entirely. We also found strong evidence that female faces are perceived as more trustworthy than males; females were sent between 20 and 25 per cent more money than males. This female trustworthiness premium may be another instance of rapid impressions of trustworthy appearance, namely feminine facial features, influencing subsequent decision-making.

Chapter 4: Do Visuospatial Asymmetries Modulate The Cheerleader Effect?

Citation

The entirety of this chapter has been published: Carragher, D. J., Lawrence, B. J., Thomas, N. A., & Nicholls, M. E. R. (2018). Visuospatial asymmetries do not modulate the cheerleader effect. *Scientific Reports*, *8*(1), 2548, doi: 10.1038/s41598-018-20784-5.

The article has been reproduced here in full. Authors BJL, NAT and MERN contributed to the conceptualisation of the current project and edited the manuscript.

Rationale

Despite the breadth of research conducted in the field of social perception, very little is known about the way trait judgments are made about an individual when they are seen in a social context. The few studies to investigate the influence of social context on social perception report that the attractiveness of an individual is modulated by social context. Most recently, Walker and Vul (2014) reported that the same individual was perceived to be more attractive when seen in a group compared to alone, a phenomenon they called "the cheerleader effect". Despite the popularity of the original article, the cheerleader effect had not been replicated, which I found curious given both the novelty of the original findings, and the range of possible implications these findings have for the wider field of social perception. In the following chapter, I investigated whether the cheerleader effect could be replicated (Walker & Vul, 2014). I also investigated whether an attentional bias toward the left side of space would influence the size of the cheerleader effect (Bowers & Heilman, 1980; Guo, Meints, Hall, Hall, & Mills, 2009).

Abstract

The cheerleader effect occurs when the same individual appears to be more attractive when seen in a group, compared to alone. As observers over-attend to visual information presented in the left visual field, we investigated whether the spatial arrangement of the faces in a group would influence the magnitude of the cheerleader effect. In Experiment 1, *target faces* were presented twice in the centre of the display: once alone, and once in a group. Group images featured two *distractor faces,* which were presented in either the left or the right visual field, or on either side of the target. The location of the *distractor faces* did not modulate the size of the cheerleader effect, which was observed in each group configuration. In Experiment 2, we manipulated the location of the *target faces*, which were presented at the far left, far right, or centre of the group. Faces were again significantly more attractive in each group configuration, and the spatial location of the *target face* did not influence the size of the cheerleader effect. Together, our results show that the cheerleader effect is a robust phenomenon, which is not influenced by the spatial arrangement of the faces in the group.

Introduction

Attractiveness is an important social cue that is rapidly evaluated from the face during first impressions (Oosterhof & Todorov, 2008; Willis & Todorov, 2006). Attractive individuals are attributed many positive stereotypes (Dion et al., 1972), including competence (Langlois et al., 2000) and intelligence (Eagly et al., 1991). Furthermore, attractive individuals receive more lenient criminal sentences (Stewart, 1980), and an increased vote share in elections (Stockemer & Praino, 2015), when compared to unattractive individuals. Facial attractiveness is signalled by the characteristics of the face being examined, including averageness, symmetry, and a sexually dimorphic appearance (Little et al., 2011b; Perrett et al., 1998; Rhodes, 2006). Because attractiveness is related to physical cues in the face, the majority of research has presented facial stimuli in isolation (i.e., a single face is presented at a time) (Phillips et al., 2014). Yet, we often meet strangers for the first time in social settings (e.g., in a boardroom or a bar). Recent findings have suggested that the perceived attractiveness of a face is influenced by social context (Furl, 2016; Geiselman et al., 1984; van Osch et al., 2015; Walker & Vul, 2014).

Previous research has shown that the presence of other faces in a group influences the attractiveness evaluations made for individual faces (Walker & Vul, 2014). For example, the attractiveness of an individual is raised in the presence of an attractive group, but lowered in an unattractive group (Geiselman et al., 1984). Furthermore, an unattractive, but task irrelevant, face can strongly influence the rate of preference choices made between two attractive faces (Furl, 2016). Most curiously, the same face is also perceived to be more attractive when it is seen in a group, compared to when seen alone; a phenomenon described as "the cheerleader effect" (Rashid & Fryman, 2008; Walker & Vul, 2014). The cheerleader effect occurs for both male and female faces, shown in groups of same gender faces. Furthermore, the cheerleader effect occurs for groups of various sizes, from 4-16 group members (Walker & Vul, 2014). The cheerleader effect strongly suggests that it is not only the attractiveness of the individual face that is evaluated, but that the surrounding faces are also encoded by the observer, which interfere with attractiveness evaluations (Phillips et al., 2014). Together, these findings show that attractiveness judgments change when an individual appears in a group, and that the social perception of an individual within a group is a unique process, whereby irrelevant faces influence our judgments of specific individuals (Phillips et al., 2014).

When meeting a group for the first time, each group member is evaluated (Phillips et al., 2014). However, if each group member was evaluated individually, group perception would be both time consuming and cognitively demanding to perform. Rather, through the process of ensemble coding, the visual system rapidly summarises the group display, which allows observers to identify the mean characteristics of the group (Alvarez, 2011). Through ensemble coding, observers are able to accurately report the average size of a group of circles (Ariely, 2001), or the average emotion displayed by a group of faces (Haberman & Whitney, 2007, 2009). Although observers can accurately recall the average size of a group of circles (Ariely, 2001), when asked to recall the size of an individual circle from the group, observers recall the circle as being similar in size to the group average (Brady & Alvarez, 2011). For example, a small individual circle presented among a group of large circles is recalled as being larger than it truly was. Brady and Alvarez (2011) suggest that ensemble coding occurs hierarchically, such that the average characteristics of the group influence the recall of individual items from the same group[9].

Walker and Vul (2014) proposed that the cheerleader effect occurs due to the hierarchical nature of ensemble coding. Initially, faces presented in a group image are automatically summarised into an ensemble average, through ensemble coding (Chong & Treisman, 2005b). The ensemble average has the average characteristics of the faces in the group (Haberman & Whitney, 2009), including the face being evaluated, and the irrelevant faces. Crucially, averageness is a trait that is perceived to be highly attractive in faces (Langlois & Roggman, 1990; Rhodes, 2006; Rhodes et al., 1999; Said & Todorov, 2011). Average faces that

[9] With the benefit of hindsight, I now realise that this explanation of hierarchical encoding does not appropriately explain that it is the structure of visual working memory, not ensemble coding, that is proposed to be hierarchical. However, as this chapter has already been published, I have elected to present this work as it was written at the time.

are created by digitally averaging many faces together are perceived to be more attractive than the individual faces included in the averaging process (Langlois & Roggman, 1990; Rhodes et al., 2001). Walker and Vul (2014) suggest that the ensemble average for a group of faces is also perceived to be highly attractive, because it has the average facial characteristics of the individual faces in the group. Walker and Vul (2014) suggest that the hierarchical structure of ensemble coding gives rise to the cheerleader effect, because an observer will recall the attractiveness of an individual face from the group as being similar to that of the ensemble average. Because the ensemble average is perceived to be highly attractive, observers will systematically recall any individual face seen in the group as being more attractive than when previously seen alone (Walker & Vul, 2014).

The cheerleader effect demonstrates that irrelevant faces in the group (i.e., those not being evaluated), influence the perceived attractiveness of an individual. When considering how a group of individuals is commonly seen, it is clear that most groups are arranged horizontally so that the group members are standing side by side. The spatial arrangement of the faces within the group may modulate the strength of the cheerleader effect, because most people over-attend to visuospatial information that is presented within the left visual field (LVF); a phenomenon known as pseudoneglect (Bowers & Heilman, 1980; Jewell & McCourt, 2000). This LVF bias likely arises because the right hemisphere, which processes the visual information in the LVF, is dominant for visuospatial processing (Kinsbourne, 1970). Pseudoneglect is demonstrated in line bisection tasks, whereby observers erroneously mark the centre of a horizontal line to the left of the true centre (Jewell & McCourt, 2000; Nicholls, Bradshaw, & Mattingley, 1999a). Interestingly, pseudoneglect also influences representational memory (Brooks, Della Sala, & Darling, 2014), whereby observers show greater accuracy when recalling landmarks that are seen in the LVF compared to the right visual field (RVF) (McGeorge, Beschin, Colnaghi, Rusconi, & Della Sala, 2007). Similarly, observers are more accurate in recalling changes in complex visual patterns when they occur in the LVF as opposed to the RVF (Aniulis, Churches, Thomas, &

Nicholls, 2016). The cheerleader effect might be modulated by the spatial arrangement of the faces in the group, because the attention of the observer is not equally distributed across the visual field, and consequently, the individual faces in the group.

Attentional asymmetries have also been shown to influence the processing of human faces (Kanwisher et al., 1997; Yovel, Tambini, & Brandman, 2008). When viewing a human face, observers gaze toward the right side of the face, which falls within the over-attended LVF (Butler et al., 2005; Guo et al., 2009; Guo, Smith, Powell, & Nicholls, 2012; Thomas, Wignall, Loetscher, & Nicholls, 2014). This preference to examine the side of the face that falls within the LVF may further reflect the lateralised functions of the right hemisphere, which is not only dominant for visuospatial processing (Kinsbourne, 1970), but also face processing (Kanwisher et al., 1997; Yovel et al., 2008). Human infants, adults and rhesus monkeys, have all been shown to fixate on the left side of the human face, suggesting that the left gaze bias for human faces might be innate (Guo et al., 2009). Furthermore, the visual scan paths displayed by the majority of individuals when examining a face also demonstrate an automatic LVF bias, which is not observed when the same individuals gaze at landscapes or symmetrical objects (Leonards & Scott-Samuel, 2005). Finally, when given the opportunity to examine faces for an extended period of time, observers continue to spend significantly more time fixating the side of the face that falls within the LVF (Guo et al., 2009; Guo et al., 2012). Therefore, when gazing at a group of faces, observers likely make more fixations toward the faces in the LVF, even over an extended time period. This left gaze bias for human faces is also reflected in the perceptual asymmetries shown by observers when making trait evaluations from faces (Burt & Perrett, 1997; Lindell, 2013a).

Observers not only spend longer exploring the right side of the face (i.e., the LVF) (Guo et al., 2012), but base their trait evaluations of individuals upon the visual information present on the right side of the face (Burt & Perrett, 1997; Butler et al., 2005; Dunstan & Lindell, 2012; Indersmitten & Gur, 2003). Observers display a strong perceptual bias, which influences the perceived attractiveness of faces, such that the right side of the face is perceived to be more attractive than the left (Burt & Perrett, 1997; Dunstan & Lindell, 2012; Franklin & Adams, 2010; Zaidel et al., 1995), and faces presented entirely within the LVF are perceived as more attractive than faces presented in the RVF (Franklin & Adams, 2010). Burt and Perrett (1997) created chimeric faces, where one half of the face was highly attractive, while the other was unattractive. Participants were then presented with two identical faces, one which showed the attractive hemiface in the LVF, while the other was mirror reversed to show the attractive hemiface in the RVF. When asked which face was more attractive, observers displayed a strong bias to select the face with the attractive side presented in the LVF, despite the two faces being identical (Burt & Perrett, 1997). Furthermore, Dunstan and Lindell (2012) found that female faces were perceived to be more attractive when they showed the right cheek more prominently. However, when the same faces were mirror reversed, observers indicated that the left cheek was more attractive (Dunstan & Lindell, 2012). Crucially, the right side of the face is naturally viewed in the LVF, as is the left cheek when it is mirror reversed. Together these findings indicate a perceptual bias, such that faces are perceived to be more attractive when seen in the LVF.

Our aim was to investigate whether the LVF bias for face perception influences the magnitude of the cheerleader effect. We manipulated the spatial arrangement of the faces in the group image, such that the target face (i.e., the face being evaluated) would always be presented in the centre of the display, while the two distractor faces could be presented to the LVF, RVF, or on either side of the target. As attention is biased toward the LVF (Bowers & Heilman, 1980; Butler et al., 2005; Jewell & McCourt, 2000), we expected that distractor faces within the LVF would be more salient to the observer than those in the RVF (Butler et al., 2005; Guo et al., 2009). Increased visual exploration of the LVF could facilitate ensemble coding, as observers show greater accuracy in recalling complex visual scenes viewed in the LVF compared to the RVF (Aniulis et al., 2016). Furthermore, the left gaze bias might also increase the perceived attractiveness of the distractor faces when they are seen in the LVF compared to the RVF (Burt

& Perrett, 1997; Franklin & Adams, 2010). Consequently, if the distractors in the LVF are perceived to be more attractive, the attractiveness of the ensemble average created from the group should also be increased. Under these conditions, the size of the cheerleader effect would increase, because the discrepancy between the attractiveness of the ensemble average and the individual face being evaluated should be greater. We predicted that the cheerleader effect would be larger when the distractor faces appeared in the LVF compared to the RVF.

Experiment 1

Method

Participants

Sixty-four participants from Flinders University (51 females, M_{age} = 25.58, *SD* = 8.42) received course credit for their participation. The Flinders Handedness Survey (FLANDERS) was used to assess participant handedness (Nicholls et al., 2013). Scores on the FLANDERS can range from -10 (strongly left handed) to +10 (strongly right handed). Data from left- and mixed-handed participants (scores ≤ +5; *n* = 9) were excluded from analysis. Participants with a cheerleader effect score that was further than 3*SD* from the condition mean (*n* = 3), who were visually impaired (*n* = 1), or who did not complete the task as instructed (*n* = 1), were also excluded from analyses. The final sample consisted of 50 strongly right-handed (*M* = 9.72, *SD* = 0.81) participants (41 females, M_{age} = 26.44, *SD* = 9.26). The procedures in the present research were approved by, and carried out in accordance with the guidelines of, the Social and Behavioural Research Ethics Committee of Flinders University.

Stimuli

Images of female faces were collected online, by querying an image search engine using the search term 'Bridesmaids' (van Osch et al., 2015; Walker & Vul, 2014). To control for the possibility that individuals might pose differently in a group, all face stimuli originally came from photographs of groups. The faces of individual group members were closely cropped from the image to create individual portraits. Facial stimuli were selected that directly faced the camera, and had both eyes directed toward the camera. The majority of facial stimuli were estimated to be between 20-40 years old, showing joyous or happy expressions, and appeared to be of Caucasian ethnicity. Three individual portraits were shown horizontally side by side to create each group stimulus (see Figure 4.1). Each individual portrait was 68 mm x 80 mm (7.78°, 9.15°) in size, and group images were 204 mm x 80 mm (23.06°, 9.15°). Stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), interfaced with a 22" monitor (1680 x 1050) running at 60Hz, which was positioned approximately 500 mm from the participant.

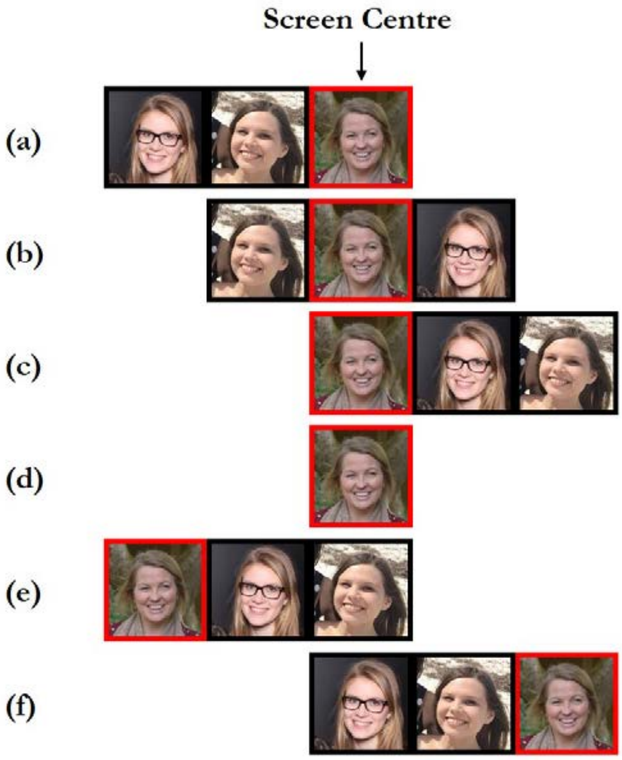

Figure 4.1. Example stimulus configurations in Experiment 1 a) LVF distractors b) BL distractors c) RVF distractors d) alone target e) left dummy trial f) right dummy trial. The target face [red frame] was presented in the centre of the display for critical trials (a, b, c, d).

Target images were presented twice: once in a group with two distractors, and once alone as a portrait. Each distractor face appeared in only one group image. Target faces always appeared in the veridical centre of the display. On group trials, three distractor configurations were used: both distractors to the left of the target (LVF distractors), one on either side (bilateral distractors; BL), or both distractors on the right (RVF distractors). The distractor configuration for each target face was counterbalanced between-participants, such that each target face was rated with LVF, BL, and RVF distractors across participants. Importantly, each target face was shown with the same two distractor faces across participants, ensuring the attractiveness of the group remained constant (Geiselman et al., 1984). The only difference between distractor conditions was the configuration of the distractors themselves.

Eighty-four faces were randomly selected from the stimulus set to appear as targets. All target faces were presented once as an individual portrait (n = 84) and once within a group (n = 84). The group image trials consisted of 28 trials for each distractor configuration (LVF, BL, RVF). Dummy group trials (n = 22), in which the "target" was the far left or the far right face in the group and the distractors filled the two adjacent spaces (see Figure 4.1e, 4.1f), were also included in the design to prevent participants from fixating on the central face, which was the target location on all critical trials. The dummy target images (n = 22) were also shown in the alone condition to conceal their purpose. All dummy trials were discarded from analyses. In total, the experiment consisted of 212 trials, which were intermixed and randomised.

Procedure

Small groups of participants (n = 6-8) completed the experimental task individually. Informed consent was obtained from all participants prior to participation. Participants were asked to rate the attractiveness of the target face, which would be identified from the group image by a red frame appearing around the image. Six practice trials were completed to familiarise participants with all possible trial conditions (including dummy trials).

In group image trials, the image was initially presented for 2000 ms, during which time each face in the group was surrounded by a black frame. The target face was not identified from the group image during this initial free viewing phase, and participants were encouraged to examine each face in the group. The target face was then identified from the group image by a red frame, which surrounded the target face. The target was cued for 1000 ms, before all faces then disappeared from the display, and participants gave an attractiveness rating. During alone presentation trials, the target face was initially presented for 1000 ms with a black frame, which was replaced by a red frame for an additional 1000 ms. This presentation timing replicates that used by Walker and Vul (Experiment 4; 2014). The FLANDERS questionnaire was completed at the end of the experiment, to avoid priming participants about the lateralised nature of the task. The experiment took approximately 30 minutes to complete.

Analysis

Attractiveness judgments were made, via mouse click, along a visual analogue scale (width = 192mm, 21.74°) that ranged from "Very Unattractive" (0%) to "Very Attractive" (100%). The spatial location of scale anchors was counterbalanced between-participants, such that half saw 'Very Unattractive' on the left and 'Very Attractive' to the right of the scale, while the other half of participants saw the anchors in the opposite orientation. Scale anchors were counterbalanced between participants to avoid stimulus-response compatibility effects, whereby participants might produce extreme responses on the side of the scale where the distractors appeared (i.e., lower attractiveness ratings for LVF distractors when "Very Unattractive" also appeared in the LVF) (Nicholls, Orr, Okubo, & Loftus, 2006). The dependent variable was the x-coordinate of the mouse click along the visual analogue scale, which was converted into a percentage of attractiveness prior to analysis.

The cheerleader effect refers to the change in attractiveness of the same face when seen in a group compared to alone (Walker & Vul, 2014). As such, a cheerleader effect measure was calculated by subtracting the rating of attractiveness of targets when seen alone, from the attractiveness ratings made when the target faces were presented in each of the three group conditions. The three resulting change scores (one for each group condition), indicated as a percentage, the change in attractiveness experienced when a target face was seen in a group, with positive values indicating an increase in attractiveness.

Data Availability

The datasets generated and analysed in the current study are available in the Open Science Framework repository, [https://osf.io/rbg8q/].

Results

To first establish whether the cheerleader effect occurred in each group condition, we used three one sample *t*-tests to examine whether the change in attractiveness was statistically significant. Faces were perceived to be significantly more attractive in all group conditions: LVF, $t(49) = 4.52$, *95%CI*[1.00, 2.59], $p < .001$, $d = 0.64$; BL, $t(49) = 3.24$, *95%CI*[0.49, 2.09], $p = .002$, $d = 0.46$; RVF, $t(49) = 4.45$, *95%CI*[0.87, 2.31], $p < .001$, $d = 0.63$. The cheerleader effect was observed in each group condition, regardless of the spatial arrangement of the distractor faces (see Figure 4.2).

A one-way within-participants analysis of variance (ANOVA) was used to investigate whether the strength of the cheerleader effect differed depending on the configuration of the distractor faces (LVF, BL, RVF). The effect of distractor configuration was non-significant, $F(2, 98) = .610$, $p = .545$, $\eta^2 = .012$. Finally, we used a Bayesian ANOVA (Rouder, Morey, Speckman, & Province, 2012) to investigate whether the observed data provided evidence in support of the null hypothesis. Our data were 8.96 times more likely to have occurred in the absence of an effect of distractor configuration, and therefore provide moderate evidence for the null hypothesis ($BF_{10} = 0.11$) (Jeffries, 1961). Together, our results strongly suggest that the spatial arrangement of the distractor faces does not modulate the strength of the cheerleader effect.

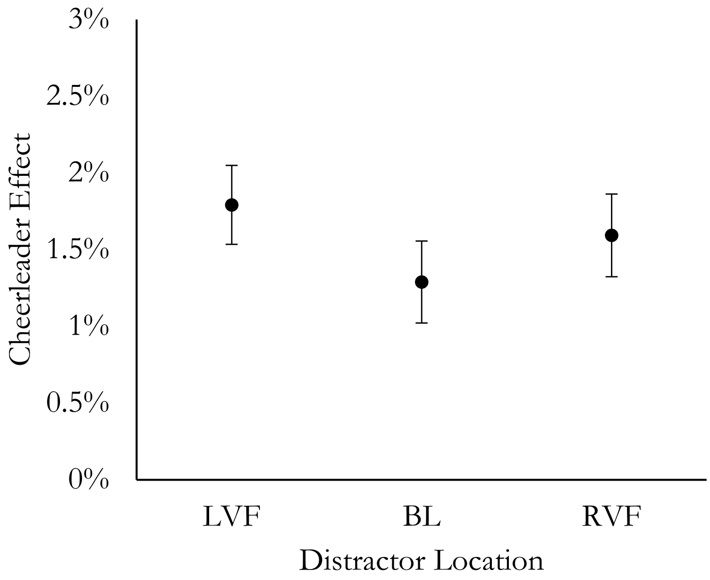

Figure 4.2. The cheerleader effect in each distractor configuration. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

Faces were perceived to be significantly more attractive when they appeared in a group, compared to when those same faces were seen alone. Our results offer the first replication of the cheerleader effect reported by Walker and Vul (2014). Furthermore, we observed the cheerleader effect in each group condition, suggesting that the effect is robust. Despite our prediction that LVF distractors would be more salient, the location of the distractor faces in the group images did not modulate the size of the cheerleader effect. A Bayesian analysis also indicated that the observed data were consistent with the null hypothesis. Indeed, the size of the cheerleader effect was similar regardless of the spatial position of the distractor faces around the target face.

A strength of our experimental design was that the target was always presented centrally. Any change to the size of the cheerleader effect could only be attributed to the spatial configuration of the distractor faces around the target face. However, this design could also be considered a limitation, as the target face was always presented centrally, rather than appearing within the LVF or RVF of the observer. The target face is the most important face in the group image, because it is the only face that the observer is required to evaluate. Perhaps it is the spatial location of the salient target face that modulates the size of the cheerleader effect, rather than the position of the irrelevant distractor faces. We conducted an exploratory follow up experiment, wherein the location of the target face was manipulated, in order to identify whether the cheerleader effect is sensitive to any manipulation of the spatial arrangement of the group.

Experiment 2

Experiment 1 showed that the spatial arrangement of the *distractor faces* in the group did not modulate the size of the cheerleader effect. To investigate whether any change to the spatial composition of the group image might influence the cheerleader effect, we manipulated the spatial location of the *target face* in the group. Target faces were presented *furthest* from the centre of the display, either at the *far left* or *far right* of the group. Visual field differences might be revealed when the target face is shifted, because unlike the distractor faces, the observer is required to make an attractiveness evaluation of the target face. As the present study was exploratory in nature, we entertained the plausibility of multiple hypotheses (H_0, H_1, H_2).

As illustrated by previous research, observers are likely to spend more time visually exploring the target face when it appears in the LVF (Butler et al., 2005; Guo et al., 2012; Thomas et al., 2014). Furthermore, the target might be perceived as more attractive when presented entirely within the LVF, compared to the RVF (Burt & Perrett, 1997; Dunstan & Lindell, 2012; Franklin & Adams, 2010). Consequently, when the target face is presented in the LVF, the group should be summarised to create an ensemble average that is more attractive than when the same target is presented in the RVF. As such, the increased attractiveness of the ensemble average in the LVF condition could result in a larger cheerleader effect, compared the RVF condition (H_1).

Although it is possible that placing the target face within the LVF will increase the cheerleader effect (H_1), it is also possible that positioning the target face within the LVF will produce a smaller cheerleader effect (H_2), because observers are more likely to rely on the ensemble average under conditions of uncertainty (Brady & Alvarez, 2011). When located in the LVF, observers are likely to spend more time examining the target face, compared to when it appears in the RVF (Guo et al., 2012). If the observer has spent more time examining the target in the LVF, they may be less likely to rely on the ensemble average when recalling the attractiveness of the target face. In contrast, if fewer fixations are made to the target in the RVF, the observer may be uncertain about the attractiveness of the target face, and instead rely more on the attractiveness of the ensemble average. Therefore, it is also possible that the cheerleader effect will be smaller when the target face appears in the LVF and larger when the target appears in the RVF (H_2).

Finally, it is also possible that the size of the cheerleader effect will not be influenced by the spatial position of the target face (H_0). This pattern of results would be consistent with the evidence in favour of the null hypothesis reported in Experiment 1. Further evidence in favour of the null hypothesis would suggest that the spatial arrangement of the group does not influence the size of the cheerleader effect.

Method

Participants

Sixty-six participants from Flinders University (44 females, M_{age} = 24.52, *SD* = 6.27) received course credit for their participation. As in Experiment 1, data from left- and mixed-handed participants were excluded (*n* = 6). All participants completed the task as instructed, and no participant data fell outside the 3*SD* exclusion criterion used in Experiment 1. The final sample consisted of 60 strongly right-handed (*M* = 9.77, *SD* = 0.75) participants (40 females, M_{age} = 24.67, *SD* = 6.51).

Stimuli, Apparatus & Procedure

The apparatus, design, and procedure were identical to Experiment 1. The target faces and their accompanying distractor faces were also those presented in Experiment 1. In Experiment 2, we manipulated the spatial location of the *target face* within the group. Target faces could be presented in the LVF, Centre, or RVF (see Figure 4.3). The location of each target image was counterbalanced between-participants, such that each target was rated in each of the three possible locations. Each target face was presented with the same two distractor faces in each group condition; only the position of the target face in the group differed between the group conditions. Dummy trials, which were not analysed, were included to prevent participants from fixating on the far ends of group images, where the target faces appeared during most critical trials. Dummy targets were presented in the centre of the display, with both distractors either to the left or right. In total, the experiment consisted of 212 trials, which were intermixed and randomised.

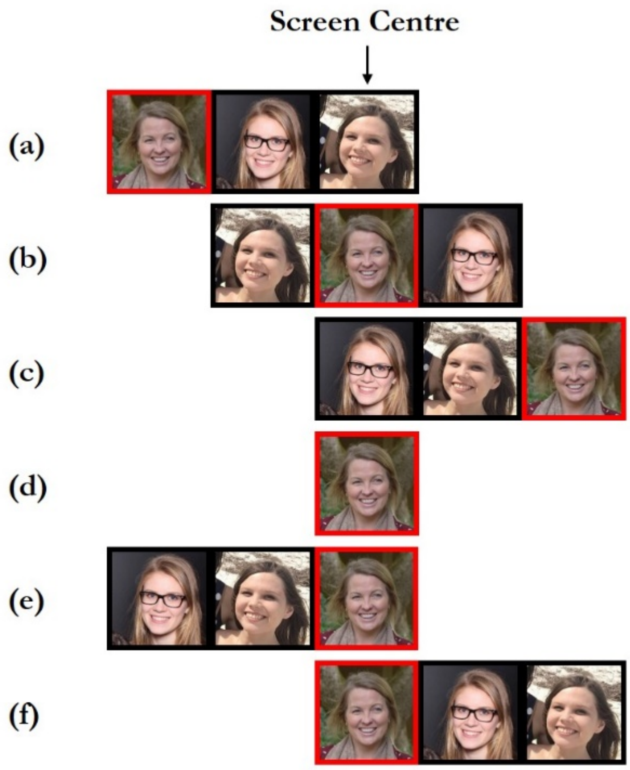

Figure 4.3. Example stimulus configurations for Experiment 2 a) LVF target b) Centre target c) RVF target d) alone target e) left dummy trial f) right dummy trial.

Results

Three one sample *t*-tests were first used to examine whether the cheerleader effect was observed in each group condition. Faces were significantly more attractive in all group conditions: LVF, $t(59) = 3.82$, *95%CI*[0.73, 2.34], $p < .001$, $d = 0.49$; Centre, $t(59) = 3.79$, *95%CI*[0.61, 1.96], $p < .001$, $d = 0.49$; RVF, $t(59) = 5.16$, *95%CI*[1.32, 3.00], $p < .001$, $d = 0.67$. Once again, the cheerleader effect was observed in each group condition, regardless of the location of the target face in the group (see Figure 4.4).

A one-way within-participants ANOVA was used to investigate whether the strength of the cheerleader effect was influenced by target location (LVF, Centre, RVF). As the assumption of sphericity was violated, Greenhouse-Geisser adjusted degrees of freedom are reported. The effect of target location was non-significant, $F(1.75, 103.29) = 1.49$, $p = .230$, $\eta^2 = .025$. Finally, we used a Bayesian ANOVA (Rouder et al., 2012) to investigate whether the data provided evidence in support of the null hypothesis. Our data were 4.41 times more likely to have occurred in the absence of an effect of target location, and therefore provide moderate evidence for the null hypothesis ($BF_{10} = 0.23$) (Jeffries, 1961). Taken together, our results strongly suggest that the location of the target face in the group image did not modulate the strength of the cheerleader effect.

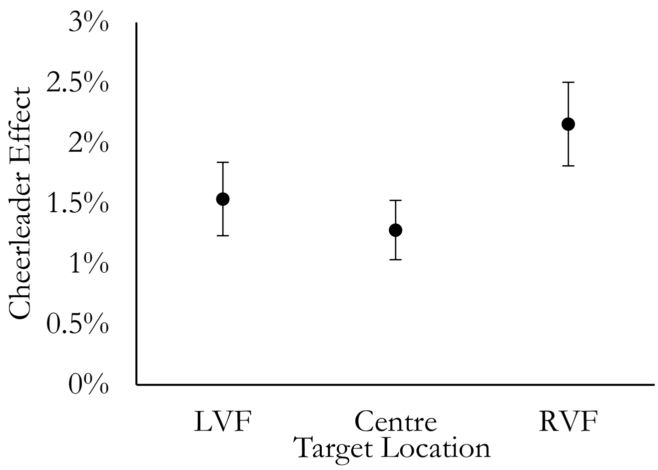

Figure 4.4. The cheerleader effect for each target location. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

The results of Experiment 2 showed a consistent cheerleader effect, regardless of the location of the target within the group. Our findings support the null hypothesis (H_0), as the position of the target face in the group did not influence the strength of the cheerleader effect. Taken together, the results of Experiment 1 and Experiment 2 strongly suggest that the spatial configuration of the group does not influence the strength of the cheerleader effect.

General Discussion

Across two experiments, we found strong evidence in support of the cheerleader effect (Walker & Vul, 2014). In contrast to our predictions, the cheerleader effect was not influenced by perceptual or visual field biases, and occurred regardless of the spatial configuration of the group. The size of the cheerleader effect appears to be consistent, such that attractiveness is increased within the range of 1.5-2%. Although the cheerleader effect is known colloquially (Rashid & Fryman, 2008), scientific investigation has been limited (Walker & Vul, 2014). Our findings show that the cheerleader effect is a robust phenomenon that can be observed using a relatively unconstrained set of images, collected from the internet. The cheerleader effect occurred in all group conditions, replicating Walker and Vul (2014), and extending upon their findings to show that the cheerleader effect is not modulated by the spatial configuration of the group image.

While the present study was not designed to directly test the proposed hierarchical encoding mechanism of the cheerleader effect, our findings are nonetheless consistent with the framework provided by Walker and Vul (2014). It is possible that the spatial configuration of the group did not influence the cheerleader effect, because ensemble coding serves to reduce perceptual redundancy by rapidly encoding and summarising complex group displays (Alvarez, 2011). Haberman and Whitney (2009) demonstrated that the average emotion shown in a group of 16 emotional faces could be accurately identified after being presented for only 500 ms. Therefore, ensemble coding can create accurate summary representations of much larger groups, which have been presented for less time, than the groups used in the present experiments. Furthermore, participants are sensitive to small changes in facial emotional expressions in large groups, without being able to identify the source of the change (Haberman & Whitney, 2011), suggesting that encoding occurs without awareness for individual item locations within the set. The findings of Haberman and Whitney (2011) are consistent with previous research, which has shown that distractors are encoded and tracked, even though they are task-irrelevant and fall outside of the attended area of the visual display (Alvarez & Oliva, 2008). Thus, the results of the present study are consistent with an ensemble coding mechanism that can rapidly summarise large set sizes, and as such, is not influenced by visual asymmetries within a small set of images.

The hierarchal encoding framework offered by Walker and Vul (2014) is currently the only framework proposed to explain the cheerleader effect. Yet, the cheerleader effect shares many similarities with the *Group Attractiveness Effect*, whereby the attractiveness of a whole group of individuals is overestimated relative to the average attractiveness ratings of each individual group member (van Osch et al., 2015). The Group Attractiveness Effect is driven by selective attention towards the most attractive faces within the group, which results in an increased estimate of the attractiveness of the whole group (van Osch et al., 2015). The selective attention framework suggested by van Osch et al. (2015) could also underlie the cheerleader effect, whereby the attractiveness of the target face is overestimated in a group, because attention is primarily directed towards the most attractive faces in the group (Maner et al., 2003). As our data are consistent with both accounts of hierarchical encoding (Walker & Vul, 2014) and selective attention (van Osch et al., 2015), it is clear that future research is necessary to directly contrast the unique predictions of each model to determine the mechanism underlying the cheerleader effect.

Although imaging (Kanwisher et al., 1997; Proverbio, Brignone, Matarazzo, Del Zotto, & Zani, 2006; Rossion et al., 2000; Rossion, Joyce, Cottrell, & Tarr, 2003) and behavioural (Borod et al., 1998; Bourne, 2005, 2010; Burt & Perrett, 1997) studies indicate that the right hemisphere shows greater activation during face processing, bilateral and left hemisphere activation (Kanwisher et al., 1997; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005) have also been reported. Proverbio et al. (2006) found that females showed bilateral activation during face processing, whereas males showed asymmetric activation of the right hemisphere, suggesting that contradictory previous findings may be the result of sex differences in cerebral lateralisation. Similarly, Bourne (2005) found that males exhibited a stronger behavioural bias on a chimeric face task than females, further suggesting that males have stronger lateralisation of face processing than females, which manifests in stronger visual field asymmetries. As such, it is possible that the spatial configuration of the group images in the present study did not modulate the cheerleader effect because the majority of our participants were female. Although there were too few males in our sample to perform a reliable sex analysis, future research should consider whether the spatial configuration of the group image does modulate the cheerleader effect among males.

Our results clearly show that attractiveness judgments made for an individual face within a group are not the same as those made for the same face presented alone (Geiselman et al., 1984; van Osch et al., 2015; Walker & Vul, 2014). The presence of other faces in a group interferes with the perceived attractiveness of an individual, even when the individual target is clearly identified from the group. As the majority of previous research has examined trait perception of individual faces, future research of group social perception is vital (Phillips et al., 2014). For example, is the cheerleader effect a phenomenon that is specific to attractiveness judgments, or is the perception of other traits also influenced in a group scene? Other trait judgments, such as trustworthiness and competence, are strongly correlated with attractiveness (Oosterhof & Todorov, 2008), and consequently may also be increased when an individual is seen within a group. However, the cheerleader effect might be unique to attractiveness judgments, because other traits are not as strongly related to facial averageness, and should not be systematically increased due to the average properties of the ensemble average of the group.

Investigating whether the cheerleader effect extends to other traits is an exciting avenue for future research.

The cheerleader effect is a robust phenomenon, wherein faces appear more attractive in a group than alone. Our findings replicate those of Walker and Vul (2014), and show that the spatial configuration of the group does not influence the magnitude of the cheerleader effect. Ensemble coding of small groups might not be subject to the visual or perceptual biases that affect the perception of a single face. Our interpretation is consistent with previous research showing that group displays are encoded and summarised rapidly, and with little awareness of the individual group members. Our findings suggest that if you are looking to increase your own attractiveness, you could do so by appearing in a group, though you needn't worry where you appear.

Chapter 5: Does The Cheerleader Effect occur for Trustworthiness?

Citation

At the time of thesis submission, the entirety of this chapter is under-review: Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. (Submitted: 8/7/18). The dissociable influence of group presence on evaluations of attractiveness and trustworthiness. *The Quarterly Journal of Experimental Psychology*. Authors NAT and MERN contributed to the conceptualisation of the current project and edited the manuscript.

Rationale

To date, the cheerleader effect has only been investigated for judgments of attractiveness. In the following chapter, I investigated whether the cheerleader effect also occurred for judgments other than attractiveness. Attractiveness and trustworthiness judgments are signalled by similar traits in the face, and are highly correlated with each other (Oosterhof & Todorov, 2008). Therefore, I investigated whether faces would also become more trustworthy in a group, compared to when seen alone.

In addition to investigating whether the cheerleader effect occurs for judgments of trustworthiness, I also turned my attention toward testing the underlying cause of the cheerleader effect. To date, the hierarchical encoding mechanism proposed by Walker and Vul (2014) has not been tested. To begin, I examined whether the attractiveness of the target face, or the attractiveness of the distractor faces in the group, influenced the size of the cheerleader effect. In the final experiment of this chapter, I investigated whether the size of the cheerleader effect that each face experienced was related to the attractiveness of the average face of the group it appeared within.

Abstract

The *cheerleader effect* occurs when the same face appears more attractive in a group than alone. As attractiveness is positively correlated with trustworthiness, we investigated whether the cheerleader effect also extends to judgments of trustworthiness. In three experiments, observers made attractiveness or trustworthiness judgments about faces presented in a group, and alone. We observed a robust cheerleader effect, whereby naturalistic female faces were significantly more attractive in a group than alone. The size of the attractiveness increase that each face experienced was unrelated to the attractiveness of the evaluated individual, or to the attractiveness of the group. A strikingly different pattern was observed for trustworthiness judgments. A negative correlation between trustworthiness increase in a group and the trustworthiness of the evaluated individual was consistently observed. Untrustworthy faces received a much larger increase in trustworthiness when seen in a group than faces that were already trustworthy. Our results demonstrate that group presence has unique influences on our perceptions of attractiveness and trustworthiness. A consistent increase in attractiveness is observed when faces appear in a group, whereas the trustworthiness of an individual can be increased or decreased, depending on the trustworthiness of the individual being evaluated.

Introduction

Many traits are rapidly evaluated from the face during first impressions (Bar et al., 2006; Willis & Todorov, 2006). These evaluations are pervasive (Chang et al., 2010), influencing the decisions we make in economic games (Rezlescu et al., 2012; Stirrat & Perrett, 2010), voting behaviour in elections (Antonakis & Dalgas, 2009; Todorov et al., 2005), and extreme sentencing outcomes for criminals (Wilson & Rule, 2015, 2016). Many models describe the facial characteristics that influence the trait evaluations we make about others (Oosterhof & Todorov, 2008; Said & Todorov, 2011; Todorov, 2008). Even though many of our daily interactions take place in social settings, the majority of these social perception models only consider how we perceive faces in isolation (Barrett et al., 2011; Phillips et al., 2014).

A growing number of studies illustrate that the perceived attractiveness of an individual face is not constant, but rather, will fluctuate in response to the presence of other faces (Carragher, Lawrence, Thomas, & Nicholls, 2018; Furl, 2016; Geiselman et al., 1984; van Osch et al., 2015; Walker & Vul, 2014). The television show '*How I Met Your Mother*' popularised the idea of "the cheerleader effect", a phenomenon said to occur when an individual appears more attractive in a group compared to when they are seen alone (Rashid & Fryman, 2008). Walker and Vul (2014) offered the first empirical account of the cheerleader effect and found that observers do, indeed, rate the same individual faces to be more attractive when shown in a group compared to alone. The cheerleader effect is a robust visual phenomenon, as neither the number of individuals in the group, nor the gender of the faces presented within the group, modulate the strength of the effect (Walker & Vul, 2014). Furthermore, the same increase in attractiveness occurs regardless of the visual field location of the faces within the group (Carragher et al., 2018). The cheerleader effect typically increases the attractiveness of an individual by 1.5 to 2% (Carragher et al., 2018).

To date, investigation of the cheerleader effect has been limited to judgments of attractiveness (Carragher et al., 2018; Walker & Vul, 2014). However, attractiveness is just one of many traits initially evaluated from the face (Willis & Todorov, 2006; Zebrowitz & Montepare, 2008). Observers also automatically evaluate the trustworthiness of unfamiliar faces (Engell et al., 2007), and will modify their behaviour in order to cooperate with individuals who have a trustworthy appearance (Rezlescu et al., 2012; Stirrat & Perrett, 2010). Trustworthiness judgments are strongly correlated with the other traits that are evaluated from the face during first impressions (Oosterhof & Todorov, 2008), and particularly, with judgments of attractiveness (Todorov, 2008). Furthermore, trustworthiness and attractiveness have been shown to load on the same factor in face perception, suggesting that both traits are signalled by similar facial characteristics (Oosterhof & Todorov, 2008). Our first aim was to investigate whether the cheerleader effect also extends to perceptions of trustworthiness. Because the two judgments are strongly correlated, we predicted that faces would be perceived to be both more attractive, and more trustworthy, when seen in a group, compared to when seen alone (Carragher et al., 2018; Walker & Vul, 2014).

The second aim of our study was to investigate the mechanism underlying the cheerleader effect. Walker and Vul (2014) proposed that the cheerleader effect is caused by the interplay between automatic visual processes and the properties of average faces. When presented with a complex visual scene, the visual system rapidly creates an ensemble average (Alvarez, 2011; Haberman et al., 2015a; Whitney & Yamanashi Leib, 2018). The ensemble average is a single, summary representation of the individual components of the group (Alvarez, 2011; Whitney & Yamanashi Leib, 2018). The ensemble average allows observers to accurately identify the average characteristics of the group, such as the average size of a group of circles (Ariely, 2001; Chong & Treisman, 2003, 2005b; Oriet & Brand, 2013). Ensemble coding also occurs for groups of faces, as observers can accurately perceive the average identity in a group of faces (Neumann et al., 2017; Neumann et al., 2013), or the average emotion shown by a group (Griffiths et al., 2018; Haberman & Whitney, 2007, 2009).

Although the ensemble average can be used to accurately identify the average characteristics of the group, an observer will incorrectly recall an individual item from the group as being more similar to the ensemble average than it was (Brady & Alvarez, 2011). For instance, observers can accurately identify that a group of faces has a happy average expression (Haberman & Whitney, 2007, 2009), but recall a face with a relatively neutral expression from that same group as being happier than it truly was (Griffiths et al., 2018). The biased recall of emotion toward the ensemble average demonstrates that memory for group scenes has a hierarchical structure, whereby the average characteristics of the group interfere with the encoded characteristics about the individual items within the group (Brady & Alvarez, 2011; Griffiths et al., 2018).

The ensemble average accurately reflects the average facial characteristics of the individuals within the group (Haberman & Whitney, 2009, 2011). As such, observers often mistake a novel face, which is the true digital average face of a previously seen group, as a face that itself has previously been seen (Neumann et al., 2013). The findings of Neumann and colleagues (2013) suggest that the mentally summarised ensemble average is conceptually similar to digitally averaged faces. Digitally averaged faces have been used extensively to study attractiveness. Crucially, rather than being of "average" attractiveness, faces with average characteristics are perceived to be highly attractive (Langlois & Roggman, 1990; Rhodes et al., 2001; Said & Todorov, 2011; Valentine et al., 2004). Indeed, Langlois and Roggman (1990) showed that digitally averaged faces are perceived to be significantly more attractive than each of the individual faces initially included in the average. The attractiveness of average facial characteristics is tied to their role as a signal of the genetic diversity of a potential mate (Lie et al., 2008; Little et al., 2011b; Rhodes, 2006; Thornhill & Gangestad, 1999). Walker and Vul (2014) proposed that because the mentally summarised ensemble average possesses the average facial characteristics of the group, it too is perceived to be highly attractive (Langlois & Roggman, 1990; Rhodes et al., 1999). Consequently, Walker and Vul (2014) hypothesised that the cheerleader effect occurs because an individual face seen in a group is recalled as being similar to the highly attractive ensemble average, which results in the consistent increase in attractiveness that is typical of the cheerleader effect.

While the cheerleader effect is consistent with the hierarchical encoding mechanism suggested by Walker and Vul (2014), the role of such a mechanism in the cheerleader effect has not been tested. Crucially, if the cheerleader effect is the product of hierarchical encoding, an increase in attractiveness can only occur when the ensemble average is more attractive than the individual being recalled from the group. Although average faces are undoubtedly attractive (Rhodes, 2006), there are conflicting accounts about just how attractive average faces are (Alley & Cunningham, 1991; DeBruine et al., 2007; Perrett et al., 1994). Therefore, we investigated whether the attractiveness of the individual being recalled from the group was related to the size of the cheerleader effect that each face experienced.

Digitally averaged faces are often rated as more attractive than each of the individual faces they are composed from (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001). If the mentally summarised ensemble average is analogous to these digitally averaged faces, then the cheerleader effect might occur regardless of the attractiveness of an individual, because the ensemble average will always be more attractive than the individual being recalled from the group. However, manipulating specific shape characteristics of digitally averaged faces has been shown to make these faces more attractive, but less average (DeBruine et al., 2007; Perrett et al., 1994). If highly attractive individuals are more attractive than the ensemble average, then hierarchical encoding would predict a negative correlation between the attractiveness of the individual being evaluated and the size of the cheerleader effect. An increase in attractiveness (i.e., the cheerleader effect) would only occur for faces that were less attractive than the ensemble average, whereas individuals more attractive than the ensemble average (i.e., a highly attractive individual) would be recalled to be less attractive in a group (Brady & Alvarez, 2011; Griffiths et al., 2018).

Experiment 1

Method

Participants

All research was approved by the Social and Behavioural Research Ethics Committee of Flinders University (project #6927). All participants provided written informed consent prior to participation, were debriefed upon completion, and were reimbursed $10 AUD or course credit for their participation. All experiments took between 20 and 30 minutes to complete. We report all experimental measures, manipulations, and participant exclusions within the manuscript.

To establish whether the cheerleader effect is statistically significant, the increase in attractiveness is compared to 0 in a one-sample *t*-test. We have previously shown that the cheerleader effect occurs with a medium effect size ($d = 0.60$; Carragher et al., 2018). A power analysis (G*Power; Faul, Erdfelder, Lang & Buchner, 2007) indicated that a sample of 32 participants was required to detect an effect of $d = 0.60$, in a two-tailed, one sample *t*-test with an alpha of $\alpha = .05$, and a power level of 90%.

Attractiveness. Thirty-two participants from Flinders University were recruited to make attractiveness judgments (25 females, $M_{age} = 23.13$, $SD = 6.29$). Prior to analysis (Carragher et al., 2018), exclusion criteria were established to remove participants with a cheerleader effect score further than 3*SD* from the condition mean ($n = 0$), and those who did not complete the task as instructed ($n = 1$), resulting in a final sample of 31 participants.

Trustworthiness. Thirty-two participants from Flinders University were recruited to make trustworthiness judgments (24 females, $M_{age} = 23.09$, $SD = 8.70$). All data were within 3*SD* of the condition mean, and all participants completed the task as instructed. As such, no data were excluded prior to analysis.

Stimuli

We collected images of female faces online from photographs of groups (Carragher et al., 2018; Walker & Vul, 2014). Individual faces were cropped from each group photograph, to create a stimulus set of individual portrait images (75 mm x 87.5 mm; 8.34° x 10.00°). Target faces were shown twice in the experiment; once alone, and once in a group (see Figure 5.1). Group images were created by presenting four individual portraits in a 2 x 2 array (150 mm x 175 mm; 17.30° x 19.51°). Stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), interfaced with a 22" monitor (1680 x 1050) running at 60 Hz, and positioned approximately 500 mm from the participant.

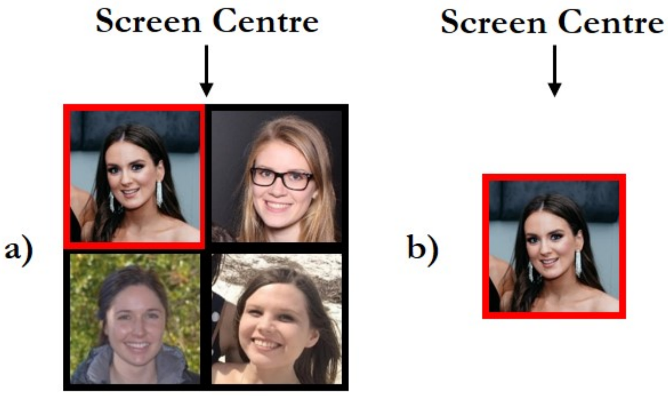

Figure 5.1. Example stimulus configurations a) group trial b) alone trial. The *target* face was identified by a red frame [top left]. The *distractor* faces were shown with black frames.

Although separate samples of participants made attractiveness and trustworthiness judgments, the same face stimuli were shown in both conditions. Eighty faces were randomly selected from the stimulus set to be the *target* faces that participants would rate. The remaining faces in the stimulus set were randomised into sets of three and were presented as *distractor* faces in the group images. Each target face was presented once in a group (n = 80), and once alone as a portrait (n = 80). Target faces were not re-purposed as distractor faces, and each distractor face only appeared once in the experiment. The location of the target face within the group array was randomised and balanced. Group and alone trials were presented in separate blocks, and the order of block presentation was counterbalanced between participants. Stimulus presentation was randomised within each block. The experiment consisted of 160 trials.

Procedure

After providing informed consent, participants were asked to rate either the attractiveness or the trustworthiness of each target face. The target face was identified from the group when a red frame appeared around the border of a single face. Participants completed four practice trials to familiarise themselves with the task. Group images were initially presented without the target identified for 2000 ms, during which time each face was surrounded by a black frame, and participants were asked to look at each face in the group. A red frame then appeared around the target face for an additional 1000 ms, after which time all faces disappeared from the screen. In alone trials, the target face was presented for 2000 ms before disappearing from the screen. The trial procedure and timing replicates that used by Walker and Vul (Experiment 4; 2014). Once the image had disappeared from the screen, participants made their judgment via mouse click along a visual analogue scale (width = 192 mm; 21.74°), which ranged from "*Very Unattractive/Untrustworthy*" (0%) to "*Very Attractive/Trustworthy*" (100%; Carragher et al., 2018; Walker & Vul, 2014).

Analysis

We first assessed the reliability of the trait judgments made between observers, both when faces were presented alone and in a group. Reliability analysis using Cronbach's alpha showed very high inter-rater reliability for judgments of attractiveness (alone, $\alpha = .94$; group = $\alpha = .95$), and trustworthiness (alone, $\alpha = .86$; group, $\alpha = .85$).

A *cheerleader effect measure* was calculated by subtracting the judgments given to the target faces seen alone, from the judgments given to the targets when presented in a group (Carragher et al., 2018). Expressed as a percentage, a positive change score represents an increase in the trait judgment for faces seen in a group, whereas negative values indicate a decrease.

Data Availability

All data reported in the current study are available in the Open Science Framework repository, [https://osf.io/c98hb/?view_only=82effbf8a91d4250a7819848e071df72].

Results

The Cheerleader Effect

One sample *t*-tests were used to investigate whether the cheerleader effect was statistically significant in each judgment condition. Faces were significantly more attractive in a group, $t(30) = 3.72$, *95%CI*[1.71, 5.87], $p = .001$, $d = 0.67$, replicating the cheerleader effect (see Figure 5.2). Unexpectedly, although the effect approached significance, faces were not significantly more trustworthy in a group, $t(31) = 1.97$, *95%CI*[-0.06, 3.19], $p = .058$, $d = 0.35$.

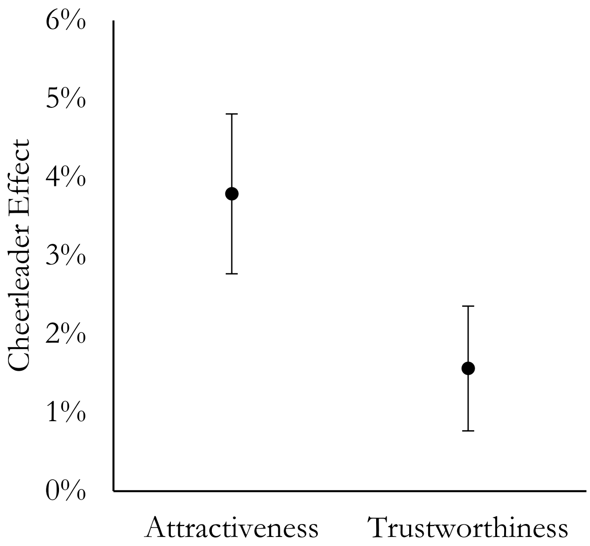

Figure 5.2. The cheerleader effect for each judgment condition. Error bars show standard error of the mean (SEM).

Individual Target Analysis

The same target faces were presented to observers in both judgment conditions. We examined whether judgments of attractiveness and trustworthiness were correlated for each target face, using the ratings for each image when presented alone. Attractiveness and trustworthiness judgments were strongly correlated, $r(80) = .630$, *95%CI*[.476, .746], $p < .001$, (see Figure 5.3).

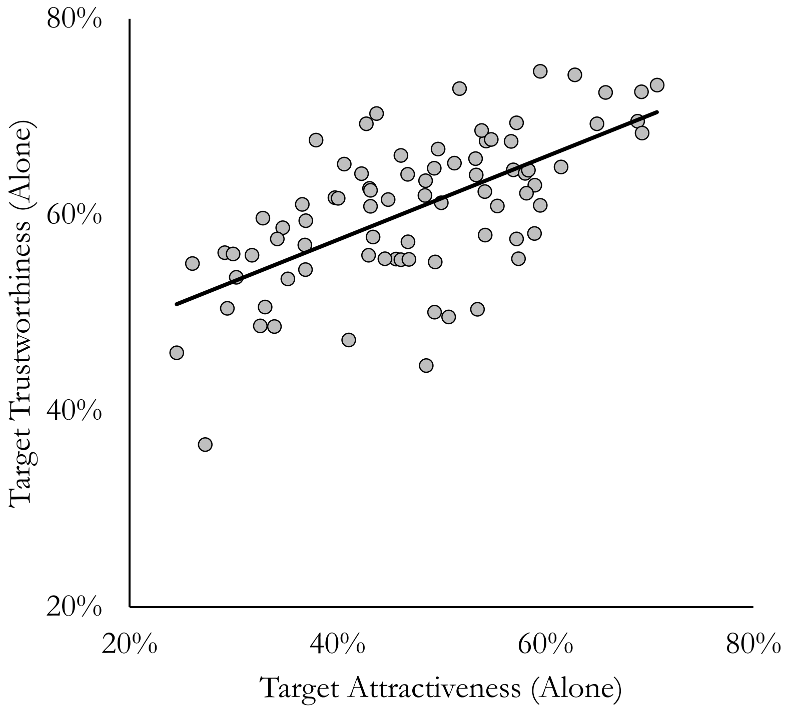

Figure 5.3. The relationship between the perceived attractiveness and trustworthiness for each individual target face.

Next, we examined whether the size of the cheerleader effect was influenced by the attractiveness of the face being evaluated. We used the average attractiveness rating for each target face when seen alone as the measure of the attractiveness of each target image. The size of the cheerleader effect each target face experienced was unrelated to the attractiveness of the face being evaluated, $r(80) = .085$, *95%CI*[-.137, .299], $p = .453$, (see Figure 5.4).

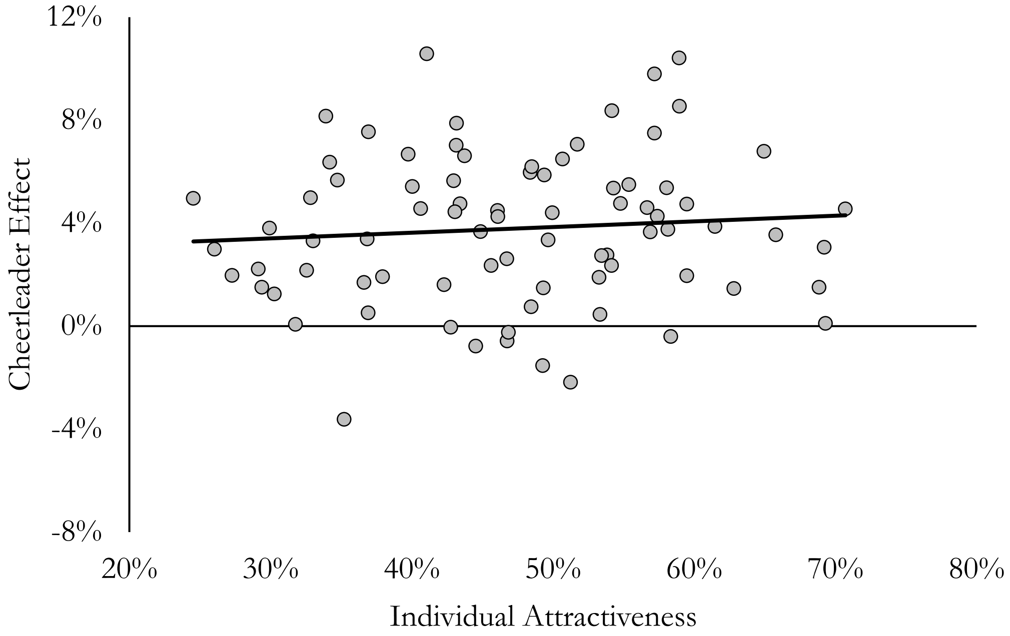

Figure 5.4. The relationship between individual attractiveness ratings for each target face and the size of the cheerleader effect each face experienced when seen in a group.

We also investigated whether the change in perceived trustworthiness for a face seen in a group was influenced by the trustworthiness of the individual being recalled. Interestingly, we observed a negative correlation between the trustworthiness of an individual and the change in their perceived trustworthiness in a group, $r(80) = -.453$, *95%CI*[-.260, -.612], $p < .001$ (see Figure 5.5). Very untrustworthy faces received a large increase in trustworthiness, whereas many trustworthy faces were perceived to be less trustworthy in a group.

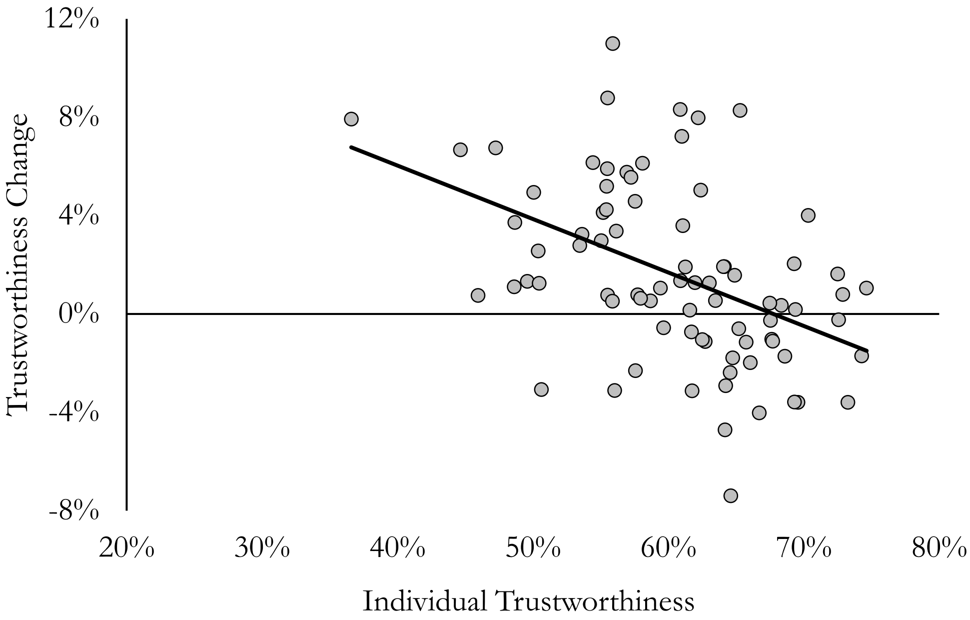

Figure 5.5. The trustworthiness of each target face seen alone and the change in perceived trustworthiness when seen in a group.

Discussion

Target faces were perceived to be more attractive in a group than when seen alone, replicating the cheerleader effect (Carragher et al., 2018; Walker & Vul, 2014). Surprisingly, this robust increase in attractiveness did not extend to perceptions of trustworthiness. The dissociation between an individual appearing more attractive, but not more trustworthy, in a group was unexpected, particularly because attractiveness and trustworthiness were strongly correlated. Visual inspection of Figures 5.4 and 5.5 makes clear that attractiveness and trustworthiness judgments are both influenced by group presence; however, the changes to each judgment are unique. As the same target faces were shown in both judgment conditions, the different correlation patterns cannot be attributed to stimulus differences, but rather, result from the judgments themselves.

Our results clearly show that the attractiveness of the face being evaluated did not influence the size of the cheerleader effect. Furthermore, 90% of the target faces became more attractive in a group, demonstrating that the cheerleader effect was not driven by a subset of images. The near systematic increase in the perceived attractiveness of individuals presented in a group is potentially consistent with the hierarchical encoding mechanism suggested by Walker and Vul (2014). Our findings suggest that if hierarchical encoding does cause the cheerleader effect, the attractiveness of the ensemble average might be analogous to digitally averaged face stimuli, which have been found to be more attractive than the individual faces they are composed from (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001).

In contrast to attractiveness, there was a strong correlation between the trustworthiness judgments made for faces when seen alone and those made for the same face presented in a group. Untrustworthy individuals received a large increase to their perceived trustworthiness in a group, while in many cases, trustworthy individuals appeared to be less trustworthy. The negative correlation between the trustworthiness of the individual and trustworthiness change in a group is consistent with a conventional model of hierarchical encoding, whereby the recalled characteristics of an individual item in a group are biased toward the ensemble average (Brady & Alvarez, 2011). While averageness strongly influences attractiveness judgments (Little et al., 2011b; Rhodes, 2006; Valentine et al., 2004), less is known about the relationship between average face shape and perceived trustworthiness[10]. The ensemble average likely conveys the "average" trustworthiness of the individual faces within the group. Crucially, both the target and distractor faces were randomly selected from the stimulus set, and randomly sorted into groups. If we assume a random distribution of trustworthiness within each group, the target faces that were rated as very untrustworthy alone, were also likely the least trustworthy faces in their group, whereas very trustworthy faces were likely to be the most trustworthy in each group. Consequently, our results are consistent with the idea that the trustworthiness of each individual was biased toward the trustworthiness of the ensemble average. However, our reasoning relies

[10] Dotsch, Hassin, and Todorov (2016) found that learned familiarity with the typical face of a population can increase the perceived trustworthiness of typical faces.

on the assumption that each group consisted of an average level of trustworthiness. We conducted a second experiment to manipulate the characteristics of each group image.

Experiment 2

In Experiment 1, we considered whether the attributes of the *target face* were related to the size of the cheerleader effect. However, the *distractor faces* are also summarised in the ensemble average, influencing the characteristics of the ensemble average, and potentially the size of the cheerleader effect. As the target and distractor faces were randomly sorted into groups in Experiment 1, the attributes of each ensemble were unknown. The aim of Experiment 2 was to investigate whether the characteristics of the distractor faces influence the size of the cheerleader effect. As such, target and distractor faces were selected based upon their perceived attractiveness or trustworthiness. Participants in the attractiveness judgment condition saw unattractive and attractive target faces, presented in both unattractive and attractive groups. Participants in the trustworthiness judgment condition saw untrustworthy and trustworthy target faces, presented in both untrustworthy and trustworthy groups.

The results of Experiment 1 were consistent with a model of hierarchical encoding that could occur if the ensemble average was more attractive than the individual faces in the group (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001). If the ensemble average is perceived to be more attractive than the individual faces in the group, then neither the attractiveness of the target, nor the attractiveness of the group, should influence the size of the cheerleader effect. In keeping with the findings from Experiment 1, we predicted the cheerleader effect would be of the same magnitude for both unattractive and attractive target faces, when presented in unattractive and attractive groups.

Our findings from Experiment 1 suggest that the perceived trustworthiness of an individual in a group is biased toward the average trustworthiness of the group. Therefore, manipulating the trustworthiness of the distractor faces within the group should, in turn, modulate the trustworthiness of the ensemble average. A group of untrustworthy faces should be summarised to create an untrustworthy ensemble average, whereas a group of trustworthy faces should be summarised to create a trustworthy ensemble average. We expected that the largest changes in trustworthiness would occur when the trustworthiness of the individual was furthest from the trustworthiness of the ensemble average. Therefore, we predicted that the largest increase in trustworthiness would occur for untrustworthy targets presented in trustworthy groups, while the largest decrease in trustworthiness would occur for trustworthy targets in untrustworthy groups.

Method

Participants

Attractiveness. Thirty-eight participants from Flinders University were recruited to make attractiveness judgments (27 females, M_{age} = 20.39, SD = 2.50). Participants who did not complete the task as instructed were excluded from analyses (n = 2). Participants with incomplete data due to technical failure were also excluded (n = 1). The final attractiveness sample consisted of 35 participants.

Trustworthiness. Thirty-eight participants from Flinders University were recruited to make trustworthiness judgments (31 females, M_{age} = 21.55, SD = 4.94). Participants with a cheerleader effect score that was further than 3SD from the condition mean (n = 1), or who did not complete the task as instructed (n = 1), were excluded from analyses. The final trustworthiness sample consisted of 36 participants.

Stimuli

The stimulus set, trial procedure, and apparatus were identical to Experiment 1. Unlike Experiment 1, the target faces and distractor groups were selected separately for the attractiveness and trustworthiness conditions. Different stimuli were selected between judgment conditions in order to create groups that varied according to attractiveness or trustworthiness.

Independent attractiveness and trustworthiness ratings were collected to categorise the stimuli for Experiment 2. Two samples of participants were recruited online (www.crowdflower.com), to make either attractiveness (n = 50, 18 females, M_{age} = 33.32, SD = 8.40) or trustworthiness (n = 62, 19 females, M_{age} = 32.82, SD = 9.04) judgments for each of the 320 individual images in the stimulus set[11]. Inter-rater reliability was excellent for judgments of attractiveness, α = .97, and very good for judgments of trustworthiness, α = .86. To ensure that judgments collected online were similar to those collected in the laboratory, we compared the ratings for the 80 target faces when presented alone in Experiment 1, with the judgments given to the same images by the online participants. Attractiveness, $r(80)$ = .807, *95%CI*[.714, .872], p < .001, and trustworthiness ratings from both samples were strongly correlated, $r(80)$ = .823, *95%CI*[.736, .883], p < .001. Therefore, we used the ratings from the online sample to categorise the stimuli.

Attractiveness Stimuli. The stimulus set was separated into faces that were less attractive than average (unattractive stimuli) and faces more attractive than average (attractive stimuli). From these sets, we randomly selected 40 unattractive targets and 40 attractive targets. An independent samples t-test confirmed that attractive targets (M = 64.10, SD = 6.19) were significantly more attractive than unattractive targets (M = 45.27, SD = 6.70), $t(78)$ = 13.06, *95%CI*[15.96, 21.70], p < .001, d = 2.92. The remaining stimuli in each attractiveness condition were randomised into sets of 3 distractor faces, creating 40 unattractive and 40 attractive groups. An independent samples t-test confirmed that average attractiveness of the attractive groups (M = 63.33, SD = 4.27) was significantly greater than that of the unattractive groups (M = 46.25, SD = 3.57), $t(78)$ = 19.40, *95%CI*[15.32, 18.83], p < .001, d = 4.34.

Trustworthiness Stimuli. Stimuli were categorised as either less trustworthy than average (untrustworthy) or more trustworthy than average (trustworthy). We randomly selected 40 untrustworthy and 40 trustworthy target faces. An independent samples t-test confirmed that trustworthy targets (M = 64.80, SD = 2.64) were significantly more trustworthy than

[11] Only data from online participants who completed the entire task, and passed all attention checks, was included in the final analysis, which was performed to categorise the stimuli.

untrustworthy targets (*M* = 55.03, *SD* = 3.25), *t*(78) = 14.76, *95%CI*[8.45, 11.09], *p* < .001, *d* = 3.30. The remaining stimuli in each trustworthiness condition were randomised into sets of 3 distractor faces, creating 40 untrustworthy and 40 trustworthy groups. An independent samples *t*-test confirmed that the average trustworthiness of the trustworthy groups (*M* = 65.00, *SD* = 1.99) was significantly greater than that of the untrustworthy groups (*M* = 55.70, *SD* = 2.26), *t*(78) = 19.52, *95%CI*[8.35, 10.25], *p* < .001, *d* = 4.36.

Design

In each judgment condition (attractiveness, trustworthiness), targets were shown three times; once in an unattractive (untrustworthy) group, once in an attractive (trustworthy) group, and once alone. Each distractor group was also presented twice; once with an unattractive (untrustworthy) target, and once with an attractive (trustworthy) target. In both judgment conditions, alone target trials were presented in a single block (*n* = 80), whereas both group conditions (unattractive/attractive, untrustworthy/trustworthy), were intermixed and presented across two blocks (*n* = 160). Each target face and distractor group appeared once per block of group trials. Stimulus presentation was randomised within each block, and block order was randomised between participants. The experiment consisted of 240 trials.

Analysis

A cheerleader effect measure was created for each of the four group conditions that were created in a factorial combination of target attractiveness (trustworthiness) and group attractiveness (trustworthiness). Initially, each of the four cheerleader effect measures was subjected to a one-sample *t*-test, which was used to establish whether a significant increase in attractiveness (trustworthiness) had occurred in each condition. Subsequently, a 2 x 2 within-participants analysis of variance (ANOVA) was used to investigate whether the cheerleader effect measures in each group condition differed significantly from one another.

Results

Attractiveness

One sample *t*-tests showed that unattractive faces were significantly more attractive in both unattractive, $t(34) = 2.40$, *95%CI*[0.31, 3.78], $p = .022$, $d = 0.41$, and attractive groups, $t(34) = 2.36$, *95%CI*[0.28, 3.70], $p = .024$, $d = 0.40$ (see Figure 5.6). Attractive faces were significantly more attractive in attractive groups, $t(34) = 2.47$, *95%CI*[0.34, 3.55], $p = .019$, $d = 0.42$; however, this effect only approached significance when attractive faces were presented in unattractive groups, $t(34) = 1.96$, *95%CI*[-0.06, 3.05], $p = .059$, $d = 0.33$.

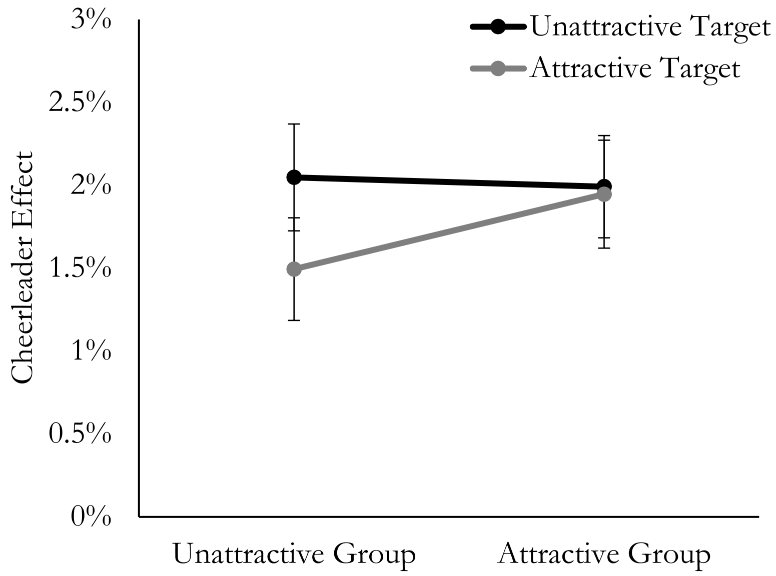

Figure 5.6. The cheerleader effect for unattractive [black] and attractive target faces [grey], presented in unattractive and attractive groups. Error bars show within-participants SEM (Cousineau, 2005).

A 2 x 2 within-participants ANOVA was used to investigate whether the strength of the cheerleader effect was influenced by target attractiveness (unattractive, attractive) and group attractiveness (unattractive, attractive). The main effects of target attractiveness, $F(1, 34) = .282$, $p = .599$, $\eta_p^2 = .008$, and group attractiveness were non-significant, $F(1, 34) = 1.02$, $p = .320$, $\eta_p^2 =$

.029. The interaction between target attractiveness and group attractiveness was also non-significant, $F(1, 34) = 1.43$, $p = .240$, $\eta_p^2 = .040$.

Trustworthiness

One sample *t*-tests identified two non-significant trends indicating that untrustworthy faces appeared more trustworthy in both untrustworthy, $t(35) = 2.01$, *95%CI*[-0.02, 3.37], $p = .053$, $d = 0.33$, and trustworthy groups, $t(35) = 1.75$, *95%CI*[-0.25, 3.39], $p = .089$, $d = 0.29$ (see Figure 5.7). In contrast, trustworthy faces did not appear more trustworthy in either untrustworthy, $t(35) = 1.42$, *95%CI*[-0.52, 2.94], $p = .163$, $d = 0.24$, or trustworthy groups, $t(35) = 1.13$, *95%CI*[-0.76, 2.66], $p = .268$, $d = 0.19$.

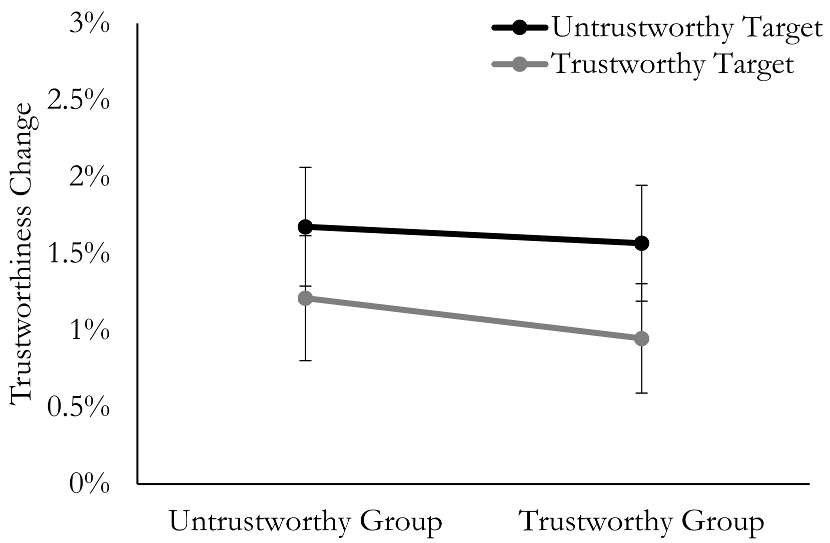

Figure 5.7. The change in perceived trustworthiness for untrustworthy [black] and trustworthy target faces [grey], when presented in unattractive and attractive groups. Error bars represent within-participants SEM (Cousineau, 2005).

A 2 x 2 within-participants ANOVA was used to examine whether the change in trustworthiness was influenced by target trustworthiness (untrustworthy, trustworthy) and group trustworthiness (untrustworthy, trustworthy). The main effects of target trustworthiness, $F(1, 35) = .986$, $p = .327$, $\eta_p^2 = .027$, and group trustworthiness were non-significant, $F(1, 35) = .213$, $p =$

.647, η_p^2 = .006. The interaction between target trustworthiness and group trustworthiness was similarly non-significant, $F(1, 35) = .047$, $p = .829$, $\eta_p^2 = .001$.

Discussion

The cheerleader effect was observed for unattractive faces, regardless of the attractiveness of the distractors, and for attractive faces presented in attractive groups. Curiously, the effect only approached significance when attractive faces were shown in unattractive groups. Notably, this is the first condition for which the cheerleader effect has not reached statistical significance (Carragher et al., 2018). This smaller cheerleader effect could be consistent with the notion that average faces are not *maximally* attractive (DeBruine et al., 2007; Perrett et al., 1994). Indeed, Perrett et al. (1994) found that the digital average of 15 highly attractive individuals was more attractive than the average of 60 randomly selected individuals. Perhaps if a group consists of two unattractive faces, and a single attractive target face, the ensemble average will be less attractive than if the group consisted of three attractive faces (i.e., Perrett et al., 1994). When presented in an unattractive group, the attractive target face might be closer in attractiveness to the less attractive ensemble average, resulting in a smaller cheerleader effect. However, caution is needed in adopting this explanation. If the attractiveness of the ensemble average changes to reflect the composition of the group, then a group composed of two attractive distractors and a single unattractive target should produce a relatively attractive ensemble average. In this scenario, the unattractive target should be further from the ensemble average in attractiveness and would receive a larger cheerleader effect. Yet, our results give no indication that the cheerleader effect for unattractive targets in attractive groups is increased.

In contrast to our predictions, the trustworthiness of the group did not influence the perceived trustworthiness of the target. Instead, untrustworthy faces appear to experience a larger increase in their perceived trustworthiness, regardless of the group in which they appear. This pattern of results is consistent with the negative correlation between individual trustworthiness and trustworthiness change found in Experiment 1 (see Figure 5.5). Our results suggest that the trustworthiness of the face being evaluated is the primary determinant of the influence that group presence will have on an individual's perceived trustworthiness. However, the increase in trustworthiness for untrustworthy faces in a group failed to reach significance.

The trend for untrustworthy faces to appear more trustworthy in a group might have failed to reach significance due to a limitation in the method used to select the target stimuli (in both judgment conditions). Although an equal number of target faces were more and less trustworthy than average, the faces were initially categorised using a binary split-half procedure, and then selected randomly from each trait category, instead of deliberately selecting those faces that differed most from one another (i.e., the least and most trustworthy faces). Therefore, it is possible that the randomly selected target faces might not have obviously differed in trustworthiness from the faces they were presented with. Perhaps a statistically significant change in trustworthiness might only occur for extremely untrustworthy or trustworthy faces. To investigate this possibility, we conducted a follow-up experiment wherein the stimuli were clearly differentiated on trustworthiness and attractiveness.

Experiment 3

Experiment 2 provided preliminary evidence that the effect of group presence is modulated by the characteristics of the individual being recalled from the group; however, these effects failed to reach statistical significance. One potential explanation for the non-significant trends is that the target faces randomly selected in Experiment 2 did not differ from one another enough to produce robust differences. To address this possibility, we performed a replication of Experiment 2, using computer-generated stimuli that had been manipulated to vary in attractiveness or trustworthiness (Todorov et al., 2013). Unlike the natural female faces used in Experiments 1 and 2, all stimuli in the present experiment were computer-generated, resembled bald male faces, and had been manipulated to appear extremely unattractive (untrustworthy) or attractive (trustworthy) (Todorov et al., 2013). To our knowledge, the present study is the first to investigate the cheerleader effect using computer-generated stimuli.

In addition to investigating the boundary conditions of the cheerleader effect, we also sought direct evidence of the hypothesised hierarchical encoding mechanism (Walker & Vul, 2014). To test the hierarchical encoding mechanism, we created an approximation of the mentally summarised ensemble average for each group image, by digitally averaging the individual faces within each group, together to produce a single average face. Neumann et al. (2013) found observers were highly familiar with novel faces, which were actually the digital average of previously seen face groups, suggesting that digitally averaged faces are an accurate representation of the mentally summarised ensemble average. In the present study, the average face of each group was presented alone as a target face, and received a trait rating from participants.

The cheerleader effect is remarkably robust (Carragher et al., 2018; Walker & Vul, 2014); yet, in Experiment 2, the effect was not statistically significant when attractive faces were presented in unattractive groups. The smaller cheerleader effect for attractive faces in unattractive groups might indicate that the ensemble average is less attractive when the group consists of two unattractive distractor faces, and a single attractive target face. We tested this hypothesis explicitly in Experiment 3, using the attractiveness ratings given to the average face of each group. We predicted that highly attractive target faces would become more attractive when averaged with highly attractive groups, but less attractive when averaged with highly unattractive groups. Conversely, we predicted that highly unattractive target faces would become more attractive when averaged with other unattractive faces, and with highly attractive faces.

As in Experiment 2, we calculated a cheerleader effect measure for attractive and unattractive target faces presented in attractive and unattractive groups. If the cheerleader effect is related to the attractiveness of the ensemble average, we expected that the size and direction of the cheerleader effect would mirror the attractiveness ratings given to the average faces of each group. Specifically, we predicted that the cheerleader effect would occur when attractive targets were presented in attractive groups, and for unattractive targets in unattractive groups. Furthermore, we predicted that the cheerleader effect would be largest when unattractive targets were presented in attractive groups, whereas a decrease in attractiveness would be found when attractive targets were presented in unattractive groups.

Finally, we used the attractiveness ratings given to each average face to create a new change measure, called "*ensemble effect*", to test the hierarchical encoding mechanism directly (Walker & Vul, 2014). Ensemble effect was used to examine whether the size of the cheerleader effect for each target image was related to the attractiveness of the average face of the group it appeared within. The ensemble effect measure was calculated for each target face by subtracting the attractiveness of the original target face when seen alone from the attractiveness rating given to the average face of its group. Ensemble effect, therefore, shows how much more attractive the group average was compared to the target face in the group. We hypothesised that hierarchical encoding would be demonstrated by a *positive* correlation between the magnitude of ensemble effect and the size of the cheerleader effect. This positive correlation would show that target faces much less attractive than the group average (i.e., those with positive ensemble effect) receive the largest cheerleader effect, whereas any face more attractive than the group average (i.e., those with a negative ensemble effect) are perceived to be less attractive in a group.

In relation to trustworthiness, the results of Experiments 2 suggested that untrustworthy faces receive a larger gain in trustworthiness when viewed in a group than do trustworthy faces. This pattern of results is consistent with the negative correlation between trustworthiness when seen alone, and the change in trustworthiness when in a group that we identified in Experiment 1. However, the increase in the trustworthiness of untrustworthy faces fell just short of significance in Experiment 2. We expected that using extremely untrustworthy and extremely trustworthy faces would reveal the extent of trustworthiness change in a group. In keeping with our previous findings, we predicted that extremely untrustworthy faces would be significantly more trustworthy in a group. Furthermore, we expected that if trustworthy faces were to ever appear less trustworthy in a group (as indicated in Figure 5.5), then the effect would become apparent using highly trustworthy faces.

Method

Participants

In contrast to Experiments 1 and 2, Experiment 3 used a within-participants design. Thirty-one participants from Flinders University were recruited to make both attractiveness and trustworthiness judgments (25 females, M_{age} = 21.68, *SD* = 5.55). All participants completed the task as instructed. Data from 1 participant were excluded from all analyses, as their cheerleader effect mean fell further than 3 *SD* from the trustworthiness condition mean. The final analyses consisted of 30 participants. Participants also completed an additional task, which was unrelated to the aims of the current study. Task order was counterbalanced between participants.

Design

Our stimuli were selected from the social perception databases created by Todorov and colleagues (2013). We selected 24 highly untrustworthy (-3 *SD*) and 24 highly trustworthy (+3 *SD*) faces to present in the trustworthiness judgment condition, and 24 highly unattractive (-3 *SD*) and 24 highly attractive (+3 *SD*) faces for the attractiveness judgment condition. For full methodological detail concerning the creation of the stimuli, see Todorov et al. (2013) and Oosterhof and Todorov (2008). To maximise the number of possible trials using the stimuli from Todorov et al. (2013), we reduced the size of each group to three faces; one target and two distractor faces (Carragher et al., 2018). Importantly, Walker and Vul (2014) found that the size of the cheerleader effect did not differ depending on the number of faces in the group. Three individual portraits were shown horizontally, side by side, to create each group image. Each individual portrait was 70 mm x 80 mm (8.00°, 9.15°) in size, and group images were 210 mm x 80 mm (23.72°, 9.15°).

Within each judgment condition, Todorov and colleagues (2013) created the high trait and low trait value faces from an initial sample of 24 neutral identities. We selected 8 identities, at random, to be the target faces for both the unattractive (untrustworthy) and attractive (trustworthy) conditions. It was necessary to select the same stimulus identities for each attractiveness (trustworthiness) condition to avoid presenting an attractive (trustworthy) target identity in the same group as its own unattractive (untrustworthy) counterpart. Within each attractiveness (trustworthiness) condition, the remaining 16 identities were randomised into pairs of distractor faces for the unattractive (untrustworthy) and attractive (trustworthy) groups. Target faces were randomly sorted to appear in unattractive (untrustworthy) and attractive (trustworthy) groups.

In the attractiveness condition, attractive and unattractive targets were shown three times: once in an unattractive group, once in an attractive group, and once alone. Similarly, in the trustworthiness condition, trustworthy and untrustworthy targets were shown three times: once in an untrustworthy group, once in a trustworthy group, and once alone. Each target was presented in the same location within the group for both attractiveness (trustworthiness) conditions. Eight target faces could not be presented with equal frequency in the three possible positions in the group. Within each of the factorial conditions of each judgment condition [target attractiveness(trustworthiness) x group attractiveness(trustworthiness)] target faces appeared three times in the left position of the group image, twice in the centre, and three times in the right position. Importantly, Carragher et al. (2018) found that the location of the target face in the group image does not influence the size of the cheerleader effect.

We used FantaMorph Deluxe 5.4.8 to create an average face for each group (Abrosoft, 2016). The three individual faces within the group were weighted equally in the averaging procedure. Each target face was included in two group averages, one created by averaging the face with its unattractive (untrustworthy) group, and one created by averaging the face with its attractive (trustworthy) group. Each group average face was presented once as a single portrait image.

As all participants rated both attractiveness and trustworthiness, judgment order was blocked and counterbalanced between participants. To minimise any confusion from making two judgments, a small prompt with the judgment condition ('attractiveness' or 'trustworthiness') was presented centrally, at the bottom of the screen, on each trial. Participants completed 160 trials in total. The attractiveness judgment condition consisted of 80 trials; 16 unattractive group trials, 16 attractive group trials, and 48 alone trials, of which 32 were presentations of group average faces. The trustworthiness judgment condition also consisted of 80 trials; 16 untrustworthy group trials, 16 trustworthy group trials, 48 alone trials, of which 32 were group average trials. Within each judgment condition, all trial conditions were intermixed and randomised. The trial procedure was identical to previous experiments.

Results

Attractiveness

The Cheerleader Effect. One samples *t*-tests showed that unattractive faces were significantly more attractive in unattractive groups, $t(29) = 3.69$, *95%CI*[1.22, 4.28], $p = .001$, $d = 0.67$, but not in attractive groups, $t(29) = 1.40$, *95%CI*[-0.41, 2.19], $p = .172$, $d = 0.26$ (see Figure 5.8). Conversely, attractive faces were significantly more attractive in attractive groups, $t(29) = 2.10$, *95%CI*[0.06, 4.74], $p = .045$, $d = 0.38$, but not in unattractive groups, $t(29) = 1.59$, *95%CI*[-0.50, 4.01], $p = .122$, $d = 0.29$.

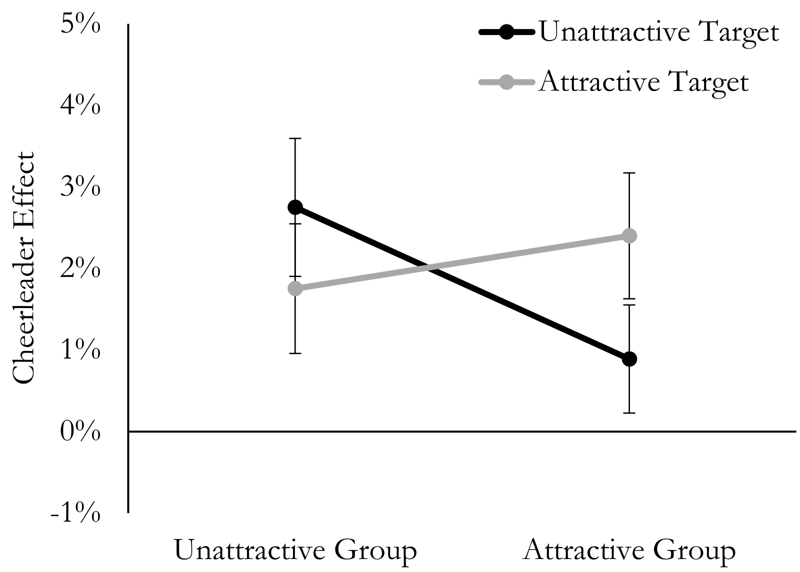

Figure 5.8. Cheerleader effect measures for unattractive [black] and attractive target faces [grey], presented in unattractive and attractive groups. Error bars show within-participants SEM (Cousineau, 2005).

A 2 x 2 within-participants ANOVA was used to investigate whether the size of the cheerleader effect was influenced by target attractiveness (unattractive, attractive) and group attractiveness (unattractive, attractive). The main effects of both target attractiveness, $F(1, 29) = 0.04$, $p = .842$, $\eta_p^2 = .001$, and group attractiveness were non-significant, $F(1, 29) = 0.78$, $p = .385$, $\eta_p^2 = .026$. The interaction between target attractiveness and group attractiveness was significant, $F(1, 29) = 5.64$, $p = .024$, $\eta_p^2 = .163$.

The significant interaction between target attractiveness and group attractiveness was analysed using post-hoc paired samples *t*-tests, with a Bonferroni corrected alpha of $a = .025$. The interaction was the result of a non-significant trend toward unattractive faces receiving a larger cheerleader effect in unattractive groups compared to attractive groups, $t(29) = 2.26$, *95%CI*[0.17, 3.54], $p = .032$, $d = 0.41$. The size of the cheerleader effect for attractive targets did not differ significantly between group attractiveness conditions, $t(29) = 0.71$, *95%CI*[-1.21, 2.50], $p = .481$, $d = 0.13$.

Average Faces. A 2 (target attractiveness: unattractive, attractive) x 3 (ensemble condition: unattractive group, original image, attractive group) within-participants ANOVA was used to investigate the effect that digital averaging had on the attractiveness ratings given to the target faces. Greenhouse-Geisser corrected degrees of freedom are reported in all analyses where sphericity was violated. The main effect of target attractiveness was significant, $F(1, 29) = 153.03$, $p < .001$, $\eta_p^2 = .841$, as was the main effect of ensemble condition, $F(1.41, 40.86) = 143.76$, $p < .001$, $\eta_p^2 = .832$ (see Figure 5.9). Both main effects were qualified by a significant interaction between the target attractiveness and ensemble condition, $F(2, 58) = 93.23$, $p < .001$, $\eta_p^2 = .763$. Separate one-way ANOVAs for each target attractiveness condition were used to investigate the significant interaction.

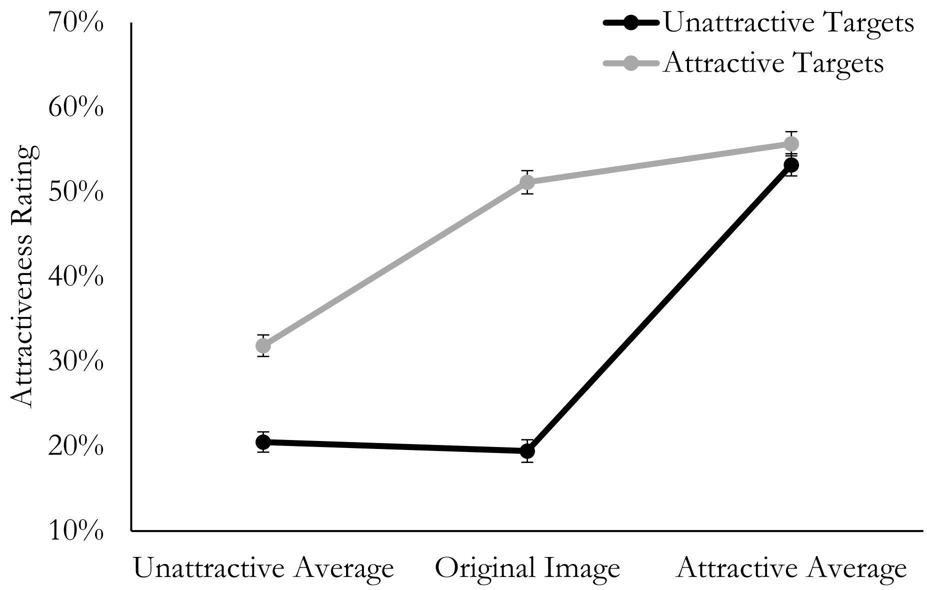

Figure 5.9. Attractiveness ratings for the original unattractive [black] and attractive [grey] faces when seen alone, and when averaged with unattractive and attractive groups. Error bars show within-subjects SEM (Cousineau, 2005).

A one-way within-participants ANOVA was used to analyse the attractiveness ratings for attractive target faces in each ensemble condition. The main effect of ensemble condition was significant, $F(1.27, 36.82) = 72.60, p < .001, \eta_p^2 = .715$. Attractive group averages ($M = 55.70$, $SD = 15.30$) were significantly more attractive than attractive target faces ($M = 51.16$, $SD = 15.14$), $t(29) = 4.40$, *95%CI*[2.43, 6.65], $p < .001$, $d = 0.80$. Attractive target faces were significantly more attractive than the unattractive group averages ($M = 31.87$, $SD = 14.28$), $t(29) = 7.87$, *95%CI*[14.27, 24.30], $p < .001$, $d = 1.44$. Finally, attractive group averages were significantly more attractive than unattractive group averages, $t(29) = 9.61$, *95%CI*[18.75, 28.90], $p < .001$, $d = 1.75$.

A one-way within-participants ANOVA was used to analyse the attractiveness ratings for unattractive target faces in each ensemble condition. The main effect of ensemble condition was significant, $F(1.13, 32.74) = 194.40, p < .001, \eta_p^2 = .870$. Attractive group averages ($M = 53.21$, $SD = 15.25$) were significantly more attractive than unattractive target faces ($M = 19.44$, $SD = 10.73$), $t(29) = 14.32$, *95%CI*[28.95, 38.59], $p < .001$, $d = 2.62$. Interestingly, the unattractive group averages ($M = 20.51$, $SD = 11.51$) were not significantly more attractive than the original unattractive target faces, $t(29) = 1.56$, *95%CI*[-0.33, 2.46], $p = .129$, $d = 0.29$. Finally, the attractive group averages were significantly more attractive than unattractive group averages, $t(29) = 14.14$, *95%CI*[27.98, 37.44], $p < .001$, $d = 2.58$.

Hierarchical Encoding. Finally, we performed the first direct test of the proposed hierarchical encoding mechanism in the cheerleader effect. Recall that we calculated a new change measure for this analysis, ensemble effect. Ensemble effect shows how much more attractive the group average face was than the target face being recalled from each group. In contrast to our prediction of a positive correlation between the two measures, the relationship between ensemble effect and the cheerleader effect was non-significant, $r(32) = -.178$, *95%CI*[-.496, .182], $p = .330$ (see Figure 5.10). Our results suggest that the size of the cheerleader effect is not related to the attractiveness of the group average.

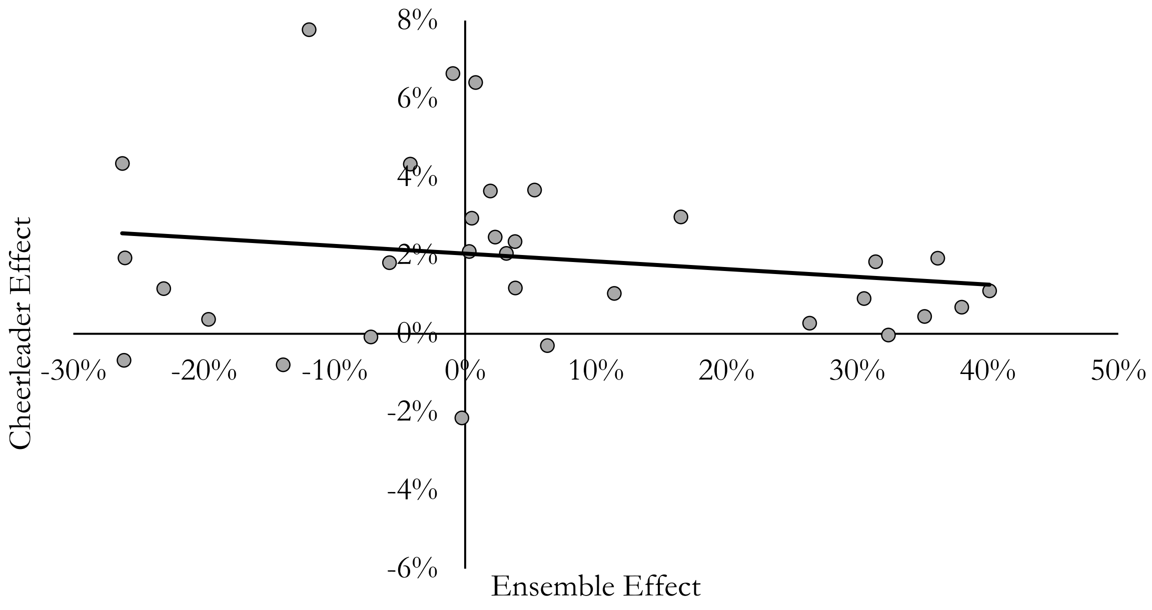

Figure 5.10. The relationship between the size of the cheerleader effect and the direction of ensemble effect for each target face. Faces with positive ensemble effect scores were less attractive than the group average, while faces with negative ensemble effect scores were more attractive than the group average.

Lastly, we performed a second correlation to investigate whether the attractiveness of the individual was related to the size of the cheerleader effect. As was the case in Experiment 1, the attractiveness of the individual was unrelated to the size of the cheerleader effect they experienced in a group, $r(32) = .152$, *95%CI*[-.207, .476], $p = .405$.

Trustworthiness

Trustworthiness Change. One sample *t*-tests found untrustworthy faces were significantly more trustworthy when shown in untrustworthy groups, $t(29) = 2.90$, *95%CI*[1.10, 6.34], $p = .007$, $d = 0.53$, but not trustworthy groups, $t(29) = 1.63$, *95%CI*[-0.54, 4.75], $p = .114$, $d = 0.30$ (see Figure 5.11). Trustworthy faces did not increase in trustworthiness when shown in either untrustworthy, $t(29) = 0.43$, *95%CI*[-1.78, 2.73], $p = .670$, $d = 0.08$, or trustworthy groups, $t(29) = 0.34$, *95%CI*[-1.87, 2.61], $p = .737$, $d = 0.06$.

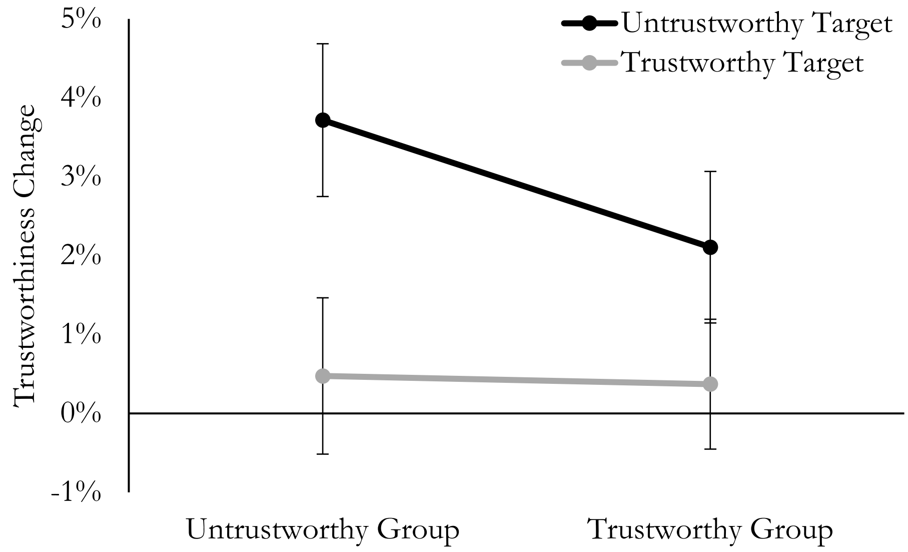

Figure 5.11. The change in perceived trustworthiness for untrustworthy [black] and trustworthy target faces [grey], when presented in unattractive and attractive groups. Error bars represent within-participants SEM (Cousineau, 2005).

A 2 x 2 within-participants ANOVA was used to examine whether the strength of the trustworthiness change was influenced by target trustworthiness (untrustworthy, trustworthy) and group trustworthiness (untrustworthy, trustworthy). The main effects of target trustworthiness, $F(1, 29) = 2.68$, $p = .113$, $\eta_p^2 = .084$, and group trustworthiness were non-significant, $F(1, 29) = .987$, $p = .329$, $\eta_p^2 = .033$. The interaction between target trustworthiness and group trustworthiness was similarly non-significant, $F(1, 29) = 1.24$, $p = .275$, $\eta_p^2 = .041$.

Average Faces. A 2 (target trustworthiness: untrustworthy, trustworthy) x 3 (ensemble condition: untrustworthy group, original image, trustworthy group) within-participants ANOVA was used to investigate the effect that digital averaging had on the trustworthiness ratings given to the target faces. The main effects of target trustworthiness, $F(1, 29) = 143.39$, $p < .001$, $\eta_p^2 = .832$, and ensemble condition were significant, $F(1.57, 45.52) = 78.71$, $p < .001$, $\eta_p^2 = .731$ (see Figure 5.12). The significant main effects were qualified by a significant interaction, $F(2, 58) =$

74.02, $p < .001$, $\eta_p^2 = .719$. Separate one-way ANOVAs for each target trustworthiness condition were used to investigate the significant interaction.

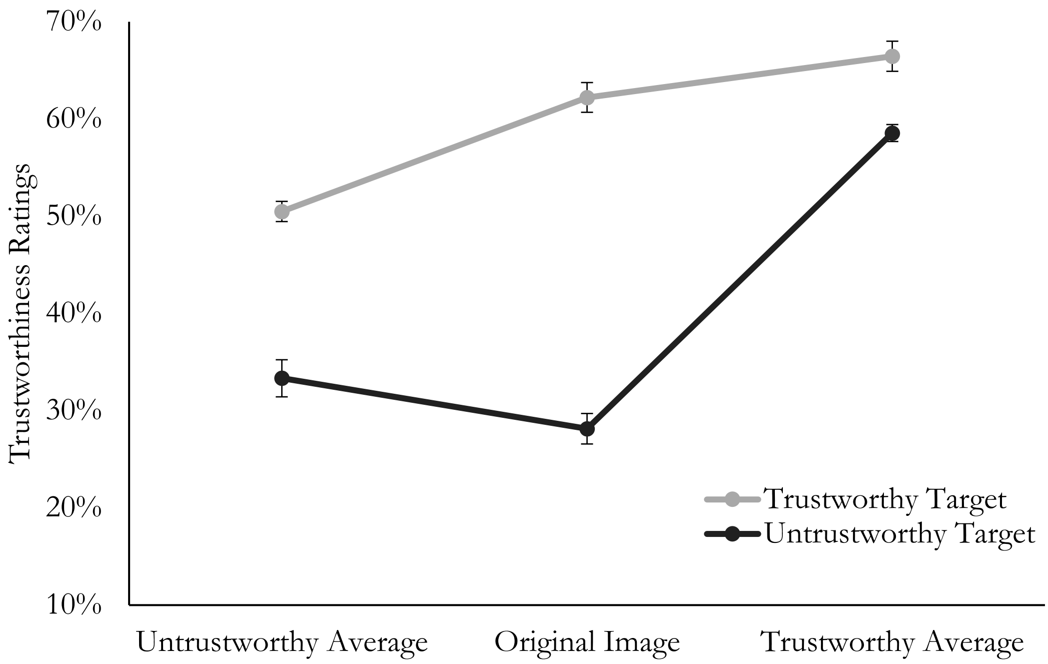

Figure 5.12. Trustworthiness ratings for the original untrustworthy [black] and trustworthy [grey] faces when seen alone, and when averaged with untrustworthy and trustworthy groups. Error bars show within-subjects SEM (Cousineau, 2005).

A one-way within-participants ANOVA was used to analyse the trustworthiness ratings for the trustworthy target faces in each ensemble condition. The main effect of ensemble condition was significant, $F(1.45, 42.03) = 36.17$, $p < .001$, $\eta_p^2 = .555$. Trustworthy group averages ($M = 66.45$, $SD = 13.01$) were significantly more trustworthy than the trustworthy target faces ($M = 62.22$, $SD = 13.18$), $t(29) = 3.52$, *95%CI*[1.77, 6.70], $p = .001$, $d = 0.64$. The trustworthy target faces were significantly more trustworthy than the untrustworthy group averages ($M = 50.49$, $SD = 14.92$), $t(29) = 5.28$, *95%CI*[7.18, 16.28], $p < .001$, $d = 0.96$. Finally, trustworthy group averages were significantly more trustworthy than untrustworthy group averages, $t(29) = 7.18$, *95%CI*[11.41, 20.52], $p < .001$, $d = 1.31$.

A one-way within-participants ANOVA was used to analyse the trustworthiness ratings for the untrustworthy faces in each ensemble condition. The main effect of ensemble condition was significant, $F(1.60, 46.33) = 109.74$, $p < .001$, $\eta_p^2 = .791$. Trustworthy group averages (M = 58.56, SD = 13.05) were significantly more trustworthy than untrustworthy target faces (M = 28.14, SD = 12.25), $t(29) = 13.30$, *95%CI*[25.74, 35.09], $p < .001$, $d = 2.43$. Untrustworthy group averages (M = 33.33, SD = 15.57) were also significantly more trustworthy than untrustworthy target faces, $t(29) = 3.25$, *95%CI*[1.92, 8.46], $p = .003$, $d = 0.59$. Finally, the trustworthy group averages were significantly more trustworthy than the untrustworthy group averages, $t(29)$ = 9.75, *95%CI*[19.94, 30.52], $p < .001$, $d = 1.78$.

Hierarchical Encoding. We tested the correlation between the magnitude of trustworthiness change in a group, and the ensemble effect score for each target face. The relationship between ensemble effect and trustworthiness change was non-significant, $r(32)$ = .244, *95%CI*[-.115, .546], $p = .179$ (see Figure 5.13). Our findings suggest the change to trustworthiness when seen in a group is unrelated to the trustworthiness of the group average.

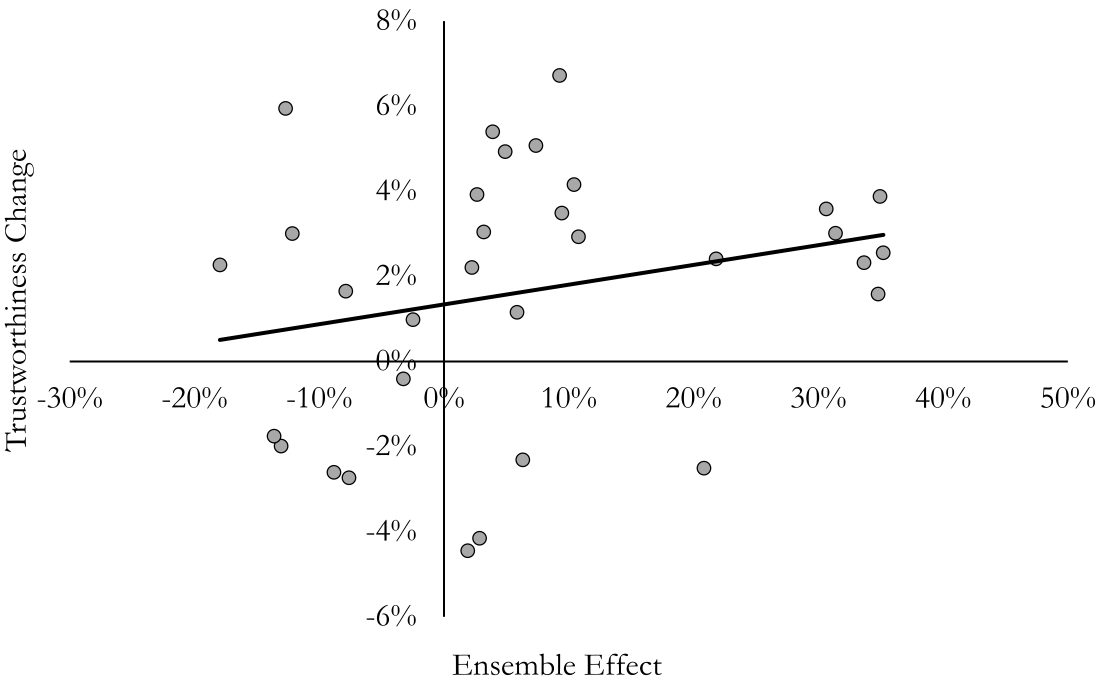

Figure 5.13. The relationship between the change in trustworthiness in a group and the direction of ensemble effect for each target face. Faces with positive ensemble effect scores were less trustworthy than the group average, while faces with negative ensemble effect scores were more trustworthy than the group average.

Finally, we performed a second correlation to investigate whether the trustworthiness of the individual was related to the change in perceived trustworthiness when seen in a group. There was a negative correlation between an individual's trustworthiness alone and the change to perceived trustworthiness in a group, $r(32) = -.500$, *95%CI*[-.723, -.183], $p = .004$, replicating the negative correlation reported in Experiment 1.

Discussion

Interestingly, our results suggest that hierarchical encoding does not contribute to the cheerleader effect. The cheerleader effect occurred when attractive targets were presented in attractive groups, and when unattractive targets were presented in unattractive groups. However, the faces created by averaging the unattractive targets with the unattractive groups were not significantly more attractive than the original unattractive targets. Moreover, the cheerleader effect did not occur when unattractive faces were presented in attractive groups, even though participants rated these average faces to be much more attractive than the original target face. Furthermore, if memory for the attractiveness of the individual was biased toward the attractiveness of the group average, then the attractive targets should have been recalled to be significantly less attractive in an unattractive group, which was not the case. Finally, there was no relationship between the size of the cheerleader effect each face experienced and the attractiveness of the average face of the group. As such, our findings are inconsistent with hierarchical encoding, and suggest that exploration of a new mechanism is warranted. Curiously, our results suggest that the mechanism causing the cheerleader effect has little to do with the attractiveness of the faces in the group at all.

Untrustworthy faces became significantly more trustworthy in a group, as was suggested by the results of Experiments 1 and 2. Importantly, we replicated the negative correlation reported in Experiment 1 between the trustworthiness of the individual and the change to their perceived trustworthiness in a group. As with attractiveness judgments, the change to the perceived trustworthiness of an individual was unrelated to the trustworthiness of the group average. It is noteworthy that trustworthy faces did not experience the predicted decrease in trustworthiness that would be typical of hierarchical encoding. Rather, appearing in either group condition had little influence on the change in trustworthiness for already trustworthy faces. Across three experiments, the perceived trustworthiness of the target face was the strongest determinant of the change in trustworthiness in a group. The largest increase in trustworthiness occurs for untrustworthy faces, while already trustworthy faces remain the same. Our results suggest that hierarchical encoding does not underlie this effect either.

General Discussion

Both attractiveness and trustworthiness are automatically evaluated from the face (Engell et al., 2007; Winston et al., 2007; Winston et al., 2002). Both are positive judgments that load on the same factor dimension in trait perception (Oosterhof & Todorov, 2008). Therefore, it is not surprising that we found attractiveness and trustworthiness ratings to be positively correlated in Experiment 1. The dissociation whereby the same faces became more attractive, but not necessarily more trustworthy when presented in a group, was surprising. It appears group presence has a unique and dissociable influence on the same faces, depending on the trait being evaluated. Although attractiveness and trustworthiness are strongly correlated, they are hypothesised to have different evolutionary origins and social functions. We propose that these functional differences between attractiveness and trustworthiness explain why group presence has a unique influence on each judgment.

Attractiveness

The cheerleader effect is a robust visual phenomenon, whereby faces are consistently perceived to be more attractive in a group than alone (Carragher et al., 2018; Walker & Vul, 2014). Unlike trustworthiness, attractiveness is not a "trait" judgment as such; it is a judgment of a physical characteristic of the face being evaluated. Facial attractiveness is hypothesised to have an evolutionary origin (Little et al., 2011b; Thornhill & Gangestad, 1999), as attractiveness ratings are similar across genders (Rhodes, 2006), and cultures (Cunningham et al., 1995). As both males and females seek attractive partners (Fletcher, Tither, O'Loughlin, Friesen, & Overall, 2004; Rhodes, 2006), it is possible that facial attractiveness might have been shaped by sexual selection (Grammer & Thornhill, 1994; Little et al., 2011b). Sex-typical traits (Cunningham, Barbee, & Pike, 1990; Perrett et al., 1998; Rhodes et al., 2003), symmetry (Grammer & Thornhill, 1994; Perrett et al., 1999), and averageness (Langlois & Roggman, 1990; Rhodes, 2006; Rhodes et al., 1999) are suggested to be highly attractive in faces because they can be honest cues about the health of a potential mate (Fink & Penton-Voak, 2002; Little et al., 2011b; Thornhill & Gangestad, 1999). From an evolutionary perspective, it is unlikely that social group membership directly influences the perceived attractiveness of an individual, because group membership is not a direct indicator of an individual's health. Nonetheless, observers consistently perceive faces to be more attractive in a group than alone (Carragher et al., 2018; Walker & Vul, 2014).

In keeping with an evolutionary perspective of facial attractiveness, the hierarchical encoding mechanism (Walker & Vul, 2014) did not treat the cheerleader effect to be a social phenomenon, but instead considered it to be a side effect of the way the visual system encodes group displays. Using predictions derived from a hierarchical encoding mechanism, we expected the size of the cheerleader effect would be related to the attractiveness of the ensemble average, which we approximated by creating a digital morph of each group image. However, we showed that the size of the cheerleader effect for each target face was unrelated to the attractiveness of the digitally average face of the group it was presented within. Furthermore, the cheerleader effect did not occur when unattractive faces where shown in attractive groups, even though the digitally averaged faces of those groups were significantly more attractive than the unattractive target. Walker and Vul (2014) also noted that the size of the cheerleader effect did not increase with increasing group size, nor with decreasing image quality, both of which were predicted to increase by a hierarchical encoding mechanism. Taken together, our findings do not support the idea that hierarchical encoding underlies the cheerleader effect. As such, we discuss potential other mechanisms that could contribute to the cheerleader effect.

The cheerleader effect is conceptually similarly to the *group attractiveness effect*, wherein observers overestimate the attractiveness of a whole group of faces, relative to the average of the attractiveness ratings given to each individual group member (van Osch et al., 2015). van Osch et al. (2015) showed that the group attractiveness effect occurs because observers only attend to a subset of the most attractive faces in the group. Attractiveness is a salient visual cue (Levy et al., 2008), and observers gaze longer at attractive than unattractive faces (Maner et al., 2003; van

Osch et al., 2015). It is possible that selective attention toward the most attractive face in the group could cause the cheerleader effect. Consistent with a selective attention mechanism, the cheerleader effect did not occur when the target was the most attractive face in the group (i.e., Experiments 2 and 3). However, the cheerleader effect was not larger for unattractive faces in attractive groups, even though they should be unattended. Furthermore, visual inspection of the correlation pattern in Figure 5.4 finds no obvious decline in the size of the cheerleader effect for highly attractive faces. Taken together, our findings indirectly suggest that the cheerleader effect is unlikely the result of selective attention toward the most attractive faces in the group. Future research should test the selective attention account directly by using eye-tracking to investigate whether observers preferentially attend to the most attractive faces in the group (Maner et al., 2003; van Osch et al., 2015).

Alternatively, the increase in attractiveness could be related to social inferences based on group membership. Appearing in a social group might suggest that the individual is friendly and likable (i.e., if other people enjoy the company of this person, perhaps there is something attractive about them). The distinction between a social effect and a perceptual effect could be identified by manipulating the type of images that are included in the group. If the cheerleader effect were truly social in nature, the effect should not occur when a human face is presented in a group with non-human stimuli (e.g., houses, cars). However, if one were to find the effect for a human face in a non-human group, it would provide evidence that the cheerleader effect is perceptual, and is related to the complexities of encoding group stimuli.

One potential limitation of our work is that the direct test of the hierarchical encoding mechanism was performed using digitally averaged computer-generated stimuli (Todorov et al., 2013). The attractiveness of average facial characteristics is hypothesised to be related to their signalling of underlying health characteristics (Lie et al., 2008; Little et al., 2011b; Rhodes, 2006), which are not present in computer-generated faces. Furthermore, the computer-generated stimuli are symmetrical and free from idiosyncrasies, which are removed when natural faces are averaged, and can also contribute to making the average face appear more attractive than the original faces. Therefore, it is possible that the relationship between the attractiveness of the group average face and the individual target is stronger when naturalistic faces are used. Despite the potential limitations of using computer-generated faces, the cheerleader effect was still observed, showing that naturalistic faces are not required to observe the effect. Future research should be undertaken to reconsider our conclusions about the hierarchical encoding mechanism by digitally averaging natural faces.

Trustworthiness

Interestingly, the cheerleader effect does not extend to perceptions of trustworthiness, even though identical stimuli were used in both judgment conditions (i.e., Experiment 1). Although trustworthiness was not systematically increased when faces were shown in a group, group presence did consistently modulate perceived trustworthiness of an individual seen in a group. This effect of group presence appears to disproportionately benefit untrustworthy individuals. Across three experiments, untrustworthy faces consistently became more trustworthy in a group, whereas there was no consistent change to trustworthy faces. The trustworthiness of the distractor faces in the group had little influence on this effect, suggesting that the trustworthiness of the target drives the effect. As with the cheerleader effect, the change to the trustworthiness of a face in a group is inconsistent with hierarchical encoding. Firstly, the change to an individual's trustworthiness when in a group was unrelated to the trustworthiness of the group average (i.e., Experiment 3). Moreover, trustworthy faces do not experience a significant decrease in trustworthiness, even when they are the most trustworthy face in the group. Unlike attractiveness judgments, which are evaluations of an individual's physical characteristics, trustworthiness judgments are inherently social in nature (Todorov, 2008). We propose that the influence of group presence on judgments of an individual's trustworthiness can be attributed to the social function of trustworthiness judgments.

Trustworthiness is a judgment made by an individual about the likelihood that a second individual will behave in a cooperative manner (Ben-Ner & Halldorsson, 2010). As such, trustworthiness judgments are argued to have evolved to enable social cooperation between individuals (Cosmides & Tooby, 1992), by helping us identify who to cooperate with, and who to avoid (Todorov, 2008). Indeed, individuals change their cooperative behaviour to reflect implicit judgments of facial trustworthiness, such that they prefer to cooperate with trustworthy looking individuals (Rezlescu et al., 2012), and avoid cooperating with untrustworthy individuals (Stirrat & Perrett, 2010). The social nature of trustworthiness judgments may mean that group membership can be used as a salient cue when evaluating the trustworthiness of an individual. Seeing an individual in a group could signal to the observer that an individual is trustworthy, because one can infer that the other members of the group do not find the individual threatening. As such, individuals might be perceived to be more trustworthy in a group than alone. However, this explanation does not explain why group membership only appears to benefit untrustworthy individuals.

While trustworthiness judgments are important social evaluations, these judgments also have a strong biological basis. Even when trustworthiness judgments are not task relevant, the trustworthiness of unfamiliar faces automatically activates the amygdala (Engell et al., 2007; Winston et al., 2002), a neural structure that is also activated by threatening stimuli (Öhman, 2005) and directly linked with fear conditioning (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998). Interestingly, the activation pattern in the amygdala is asymmetric, such that greater activation occurs for untrustworthy compared to trustworthy faces (Engell et al., 2007; Todorov, 2008). Todorov (2008) suggests that the asymmetric activation of the amygdala is evidence that it is more important to avoid potentially threatening untrustworthy individuals, than it is to cooperate with trustworthy individuals (Cosmides & Tooby, 1992). Perhaps group context does not benefit trustworthy individuals, simply because they were already considered trustworthy, and were already likely to be cooperated with. In contrast, group context might disproportionately benefit untrustworthy individuals, because they were likely to be avoided when seen alone. Appearing in a group might instead suggest to the observer that the individual can be trusted, because the other members in the group have not judged the individual to be an immediate threat.

A potential limitation concerns the generalisability of our findings to trustworthiness judgments of naturalistic male faces. In Experiments 1 and 2, the stimuli presented were happy female faces, which overall, were rated to be relatively trustworthy. The male faces that we used in Experiment 3 were computer-generated, and were selected because they appeared extremely trustworthy or untrustworthy. Therefore, our work does not speak to the effect of group presence on naturalistic male faces. Given that observers spontaneously trust female faces more than male faces (Carragher, Thomas, & Nicholls, 2017), there could be a pronounced difference in the effect of group presence on trustworthiness judgments, depending on the sex of the face being evaluated. Perhaps male faces, which are considered less trustworthy than females initially, experience a larger increase in trustworthiness when in a group than do female faces. Further research, which uses naturalistic male and female faces, is needed to address this suggestion directly.

Conclusion

Our work adds to a growing literature, which indicates that trait perception is influenced by social context (Carragher et al., 2018; van Osch et al., 2015; Walker & Vul, 2014). As the vast majority of social perception research has examined how we judge faces that are presented in isolation (Phillips et al., 2014), it is clear that further work is needed to understand how social presence influences trait perception. Future work should consider whether appearing in a group not only makes an individual more attractive, but whether this results in the attribution of other favourable stereotypes that are associated with attractiveness, such as competence or intelligence (Eagly et al., 1991; Langlois et al., 2000). Furthermore, Figure 5.5 showed that many highly trustworthy individuals were perceived to be less trustworthy within a group. Therefore, it should be considered that the influence of social presence might not always be beneficial to an individual. Our findings raise the interesting possibility that the influence of group presence on trait impressions differs depending on the trait being evaluated. Further research is needed to understand how social presence influences a multitude of trait judgments other than attractiveness and trustworthiness.

Chapter 6: Does Hierarchical Encoding Cause The Cheerleader Effect?

Citation

At the time of thesis submission, the entirety of this chapter has been resubmitted for peer-review: Carragher, D. J., Thomas, N. A., Gwinn, O. S., & Nicholls, M. E. R. (Resubmitted: 12/10/18). The cheerleader effect occurs in the absence of hierarchical encoding. *Journal of Experimental Psychology: General.* Authors NAT, OSG and MERN contributed to the conceptualisation of the current project and edited the manuscript.

Rationale

In the following chapter, I explicitly test whether Walker and Vul's (2014) hierarchical encoding mechanism causes the cheerleader effect. Across five experiments, I investigated whether the cheerleader effect would occur when faces were presented in groups that could not be mentally summarised to create an ensemble average with average facial characteristics. I also investigated whether the cheerleader effect relies on face specific processing mechanisms, and whether the cheerleader effect itself is specific to human faces. Finally, this chapter also contains a meta-analysis of seven control condition cheerleader effect measures that are reported, which identifies the size of the typical cheerleader effect for future research.

Abstract

"The cheerleader effect" occurs when an individual face is perceived to be more attractive in a group of other faces compared to alone. The visual system rapidly summarises groups of faces into an ensemble average, and the hierarchical structure of visual working memory causes observers to remember individual faces as being more alike the ensemble average than they truly were. The cheerleader effect might occur because the ensemble average has average facial characteristics, which are typically perceived to be highly attractive. Across five experiments, we tested whether hierarchical encoding is necessary for the cheerleader effect to be observed. Observers gave attractiveness ratings for target faces shown in a group and alone. Consistent with hierarchical encoding, the largest cheerleader effects occurred when the target face was presented in a group with human faces, which could be summarised to give the ensemble representation average characteristics. A meta-analysis of the data from seven control conditions indicated that the *typical* cheerleader effect increases the attractiveness of an individual by 1.6% (d = 0.69). Although some support for the hierarchical encoding mechanism was found, we also observed significant cheerleader effects in group conditions that could not be summarised to create an average face, suggesting hierarchical encoding is not necessary for the cheerleader effect to occur. Instead, our results indicate that the size of the cheerleader effect fluctuates with the value of the items in the group, suggesting the effect might be related to the socially desirable characteristics that are attributed to individuals in groups.

Introduction

Facial attractiveness is rapidly evaluated during first impressions (Olson & Marshuetz, 2005; Ritchie, Palermo, & Rhodes, 2017; Willis & Todorov, 2006), and is associated with many positive social stereotypes (Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000). Compared to unattractive individuals, attractive people are perceived to have more socially desirable personalities and occupations (Dion et al., 1972), to be more trustworthy (Andreoni & Petrie, 2008; Wilson & Eckel, 2006) and intelligent (Eagly et al., 1991; Langlois et al., 2000). Attractive individuals also receive shorter criminal sentences (Stewart, 1980), higher salaries (Hamermesh & Biddle, 1993), and go on more dates than unattractive individuals (Langlois et al., 2000). Facial attractiveness is influenced by specific facial characteristics such as symmetry, sexually dimorphic traits, and averageness (Little et al., 2011b; Rhodes, 2006; Thornhill & Gangestad, 1999). Because the majority of facial attractiveness research has focused on identifying the characteristics of attractive faces, these studies have primarily presented face stimuli in isolation (Phillips et al., 2014). Yet many of our daily interactions take place in social settings, wherein people are surrounded by others when we meet them for the first time. Recent findings indicate that the perceived attractiveness of an individual can fluctuate in response to social context (Carragher et al., 2018; Geiselman et al., 1984; Little et al., 2011c; Sigall & Landy, 1973; van Osch et al., 2015; Walker & Vul, 2014).

The notion that attractiveness can be influenced by the presence of others has colloquially been described as *the cheerleader effect*, which was popularised by the television show '*How I Met Your Mother*' (Rashid & Fryman, 2008). Walker and Vul (2014) provided the first empirical support for the cheerleader effect, reporting that the same individual is perceived to be more attractive when seen in a group, compared to when they are seen alone. The cheerleader effect is a robust visual phenomenon, observed for female and male faces, as well as for individual faces seen in groups of 4 to 16 faces (Walker & Vul, 2014). Carragher et al. (2018) replicated the cheerleader effect and found that it was remarkably consistent in size, regardless of the spatial position of the target face within the group. Overall, the same individual is perceived to be approximately 1.5-2% more attractive in a group, compared to when they are seen alone (Carragher et al., 2018). Walker and Vul (2014) hypothesised that the cheerleader effect was the result of a complex interaction between the way that visual scenes are summarised by the visual system (Alvarez, 2011; Whitney & Yamanashi Leib, 2018), the attractive characteristics of average faces (Langlois & Roggman, 1990; Rhodes, 2006), and the hierarchical structure of visual working memory (Brady & Alvarez, 2011, 2015a).

The cheerleader effect demonstrates that even though the observer is instructed to recall the attractiveness of a single face from the group, the task-irrelevant distractor faces within the group display are encoded and interfere with the subsequent judgement. When presented with a complex visual scene, such as a group of faces, the human visual system rapidly creates a summary representation of the scene through a process known as ensemble coding (Alvarez, 2011; Ariely, 2001; Whitney & Yamanashi Leib, 2018). Ensemble coding compresses redundant visual information into a single percept, which allows the visual system to minimise the cognitive load associated with encoding many individual items, whilst retaining access to the accurate summary statistics of the scene (Alvarez, 2011; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018). Through ensemble coding, observers can accurately identify the average characteristics of a scene, such as the average orientation of a group of tilted gabor patches (Parkes et al., 2001) or the average size of a group of circles (Ariely, 2001; Chong & Treisman, 2003, 2005b). Despite being instructed to focus on the size of the four individual circles in the display, observers incorrectly "remembered" seeing circles that were the average size of the group, even though no such circles were presented (Ariely, 2001). Ensemble coding occurs rapidly, if not automatically (Fischer & Whitney, 2011; Parkes et al., 2001), and summary statistics are extracted with remarkable accuracy (Alvarez, 2011; Whitney & Yamanashi Leib, 2018), even from unattended displays (Alvarez & Oliva, 2008, 2009). Importantly, ensemble coding also occurs for groups of faces (Haberman & Whitney, 2007, 2009, 2011).

Through ensemble coding, observers can accurately identify the average identity of a group (de Fockert & Wolfenstein, 2009; Neumann et al., 2013; Yamanashi Leib et al., 2014). After being presented with a group consisting of four different identities, observers are asked whether the subsequent memory probe showed a face that was a member of the preceding group. In addition to correctly recognising set exemplars, observers also mistakenly "recognise" the set average (i.e., a face that had been created by digitally averaging the individual faces in the group together). Crucially, observers do not endorse "foil averages" (i.e., faces made by averaging a different group of faces; Neumann et al., 2013), suggesting that the set average is not recognised because average faces are familiar (Langlois et al., 1994), but because the observer mentally summarises the group and extracts the average identity (de Fockert & Wolfenstein, 2009; Neumann et al., 2013).

Haberman and Whitney (2007) found that through ensemble coding, observers could also accurately identify the average expression shown by a group of four emotional faces. Remarkably, even when no individual face shows the average emotion of the group, observers can identify the group's average expression with the same precision that is shown when identifying the expression of a single emotional face (Haberman et al., 2009). The ensemble coding of expression occurs rapidly, as observers can accurately identify the average expression shown by a group of 16 faces that has been presented for only 50 ms (Haberman & Whitney, 2009). Ensemble coding can also be used to accurately identify the diversity of a group on dimensions including ethnicity, gender and dominance (Phillips et al., 2018). Interestingly, Luo and Zhou (2018) also found that observers could accurately identify the average attractiveness of a group of faces. Taken together, these findings demonstrate that the ensemble coding of faces is automatic, highly precise, and occurs across a variety of dimensions including identity, emotion, and attractiveness.

Despite allowing us to accurately identify the average emotion of a group of faces (Haberman & Whitney, 2007, 2009), ensemble coding also influences our memory for the individual items that were present in the group (Brady & Alvarez, 2011; Lew & Vul, 2015). When asked to recall the characteristics of a single item from a group, observers remember it to be more alike the ensemble average than it was. For example, Brady and Alvarez (2011) asked observers to recall the size of a single circle, which was presented in a display that simultaneously consisted of three sets of circles that were different sizes and colours. Observers recalled the same individual circle to be larger when it was the same colour as the larger circles in the display, and smaller when it was the same colour as the smaller circles. Hierarchical encoding also influences the recall of emotion shown by an individual face in a group (Griffiths et al., 2018). When asked to recall the emotion of a single target face from the group, observers were more likely to recall it to be slightly more alike the average expression of the group than it truly was. Target faces that were less happy than the group average were recalled as being happier than they really were, while faces that were happier than the group average were remembered as being less happy. These findings suggest that visual working memory has a hierarchical structure, wherein the individual items from the scene are encoded in reference to the summary statistics of the display (Lew & Vul, 2015), and when recalled, are incorrectly remembered as being more alike the group average than they really were (Brady & Alvarez, 2011; Griffiths et al., 2018).

Walker and Vul (2014) proposed that the cheerleader effect is also caused by hierarchical encoding, such that the recalled attractiveness of an individual face is biased toward the attractiveness of the ensemble average (Luo & Zhou, 2018). Crucially, instead of being perceived to be of "average" attractiveness, faces with average characteristics are actually perceived to be highly attractive (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001). In fact, average faces created by digitally averaging individual faces together are often perceived to be more attractive than the individual faces from which they are made (Langlois & Roggman, 1990; Rhodes et al., 1999; Rhodes et al., 2001). Walker and Vul (2014) suggested that much like a digitally averaged face, the ensemble average is also perceived to be highly attractive, because it has the average facial characteristics of the group. Therefore, Walker and Vul (2014) hypothesise that the cheerleader effect occurs because the individual target face is recalled as being more similar to the ensemble average (Brady & Alvarez, 2011), which, due to its average characteristics, is actually perceived to be highly attractive (Langlois & Roggman, 1990). Although the cheerleader effect is consistent with the hierarchical encoding mechanism proposed by Walker and Vul (2014), such a mechanism has not been directly tested. Our aim was to investigate the underlying cause of the cheerleader effect by first testing for an influence of hierarchical encoding and subsequently testing for the possible contribution of other factors.

Experiment 1

The hierarchical encoding mechanism proposed by Walker and Vul (2014) is predicated on the ensemble being mentally summarised to create an ensemble average that has average facial characteristics. To explore what role, if any, hierarchical encoding has in the cheerleader effect, we investigated whether the cheerleader effect would occur when individual faces were presented in a group condition that could not be summarised to create a face with average facial characteristics.

In a between-participants design, observers were randomly assigned to complete one of two cheerleader effect tasks. In both tasks, observers made attractiveness ratings for target faces that were seen once in a group with two unique distractor faces (i.e., control condition), and once alone as a portrait image. The two cheerleader effect tasks differed in the composition of a second group condition that the target face was also seen in. Observers in the *identical-distractors task* rated the attractiveness of the same target face when it was presented in a group where the two distractor images were identical images of the target face (i.e., identical-distractors condition). Observers in the *self-distractors task* rated the attractiveness of the same target face when it was presented in a group where the two distractor images were different photographs of the target face (i.e. self-distractors condition).

The key distinction between these distractor conditions (identical-distractors, self-distractors) was the amount of variance between the faces in the group image. Crucially, there is no variance between the faces in a group consisting of identical photographs of the same person. As such, identical faces cannot be summarised to create an ensemble average with average facial characteristics that are more attractive than the individual faces in the group. On the other hand, there can be considerable variance between different photographs of the same person, due to random factors between images such as lighting, head orientation, and expression. The variability between different photographs of the same person is such that observers often classify different photographs of the same person as being photographs of two different people (Jenkins, White, Van Montfort, & Burton, 2011), and the variability of trait judgments made for different photographs of the same person can often exceed the variability of trait judgments made for different individuals (Todorov & Porter, 2014). Because of this variability, different photographs of the same person can be digitally averaged to create a face with average characteristics (Burton et al., 2005).

We predicted the cheerleader effect would occur in the control condition, because the three unique faces in the group can be summarised to create an ensemble average with attractive, average facial characteristics (Carragher et al., 2018; Walker & Vul, 2014). If the cheerleader effect is caused by hierarchical encoding, the effect should not occur in the identical-distractors condition, because the ensemble average would be no more attractive than the target face. Finally, the variability between different photographs of the same person should produce an ensemble average with average characteristics, and the cheerleader effect should also occur in the self-distractors condition.

Method

We have previously found that the cheerleader effect occurs with a medium effect size ($d = .60$; Carragher et al., 2018). A power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample of 26 participants was required to achieve 85% power, with an alpha of $\alpha = .05$, in a two-tailed, one sample t-test. The one sample t-test is used as the primary test of the cheerleader effect, to assess whether the change in the perceived attractiveness of an individual seen in a group is statistically significant. Prior to analysis (Carragher et al., 2018), exclusion criteria were established to remove all data from participants with a cheerleader effect score further than 3*SD* from the condition mean, and from those participants who did not complete the task as instructed.

The Social and Behavioural Research Ethics Committee of Flinders University approved all research conducted in the current study. All participants gave their written informed consent prior to testing, were debriefed upon completion, and received course credit or $10 AUD for their participation. Each testing session required approximately 15-25 minutes to complete[12]. All participants in each experiment reported in the current study were recruited from Flinders University.

Participants

Sixty-four participants (48 females, M_{age} = 23.58, *SD* = 9.73) were recruited to participate in either the identical-distractors task or the self-distractors task. Participants with a cheerleader effect further than 3*SD* from the condition mean (*n* = 2), and those who failed to follow task instructions (*n* = 2) were excluded from analysis (Carragher et al., 2018). Additionally, participants with incomplete data due to technical failure were also excluded from analysis (*n* = 1). The final sample in the identical-distractors task consisted of 28 participants (20 females, M_{age} = 26.68, *SD* = 13.76). The final sample in the self-distractors task consisted of 31 participants (23 females, M_{age} = 21.13, *SD* = 3.46).

Apparatus

All experiments in the current study were programmed using E Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), and presented on a 22" monitor (1680 x 1050) running at 60 Hz. Stimuli were viewed at an approximate distance of 500 mm. In each experiment, individual

[12] Participants in the self-distractors task (Experiment 1), as well as all participants in Experiments 2 and 4, completed a single additional task that was unrelated to the aims of the current research. For practical purposes, additional tasks were given when the cheerleader effect task would have otherwise required only a brief testing session. The order of task administration was counterbalanced for these participants.

portraits were presented in the centre of the display (70 mm x 80 mm; 8.00° x 9.15°). Group images were created by presenting three individual portraits side by side (210 mm x 80 mm; 23.72° x 9.15°; Carragher et al., 2018). The position of the target and distractor faces in the group (i.e., left, centre, right) was randomly assigned. Target faces appeared with equal frequency in each position within the group. In all experiments that required the same target face to be presented in two or more group images, the target image was always presented in the same position within each group (left, centre, right).

Stimuli

Identical-Distractors. The face stimuli selected for the identical-distractors task came from the stimulus set created by Carragher et al. (2018). This stimulus set was created by initially collecting photographs of females in groups, by querying an online search engine with the search term "Bridesmaids". Individual portraits were created by closely cropping the individual faces from the group image. In total, the stimuli set contained 320 unique faces. All faces were female and showed happy or joyous expressions, as determined by the researchers. Most faces appeared to be of Caucasian ethnicity and between 20 and 40 years of age.

From this stimulus set, 105 faces were randomly selected to be presented as target faces in the identical-distractors task. The remaining faces from the stimulus set were randomly presented as distractor faces in the control condition (see Figure 6.1). Each group image was presented once, and a cheerleader effect measure was only calculated for the single target face in the group. Identical-distractor groups were created by simultaneously presenting the target image three times side by side. Trial presentation was fully randomised, with group and alone trials intermixed. The identical-distractors task consisted of 315 trials: 105 control trials, 105 identical-distractors trials, and 105 alone trials.

Self-Distractors. A new stimulus set was created for the self-distractors task. We collected 3 different images of 225 unique identities from online sources. The stimulus set consisted of images of female athletes, because many labelled images of these athletes are readily available, but were unlikely to be familiar to participants (Fink, 2015). Indeed, no participant spontaneously claimed to have recognised an identity among the stimulus set. The faces in this stimulus set were also all female, showed happy or joyous expressions, and most appeared to be of Caucasian ethnicity and between 20 and 40 years of age. All images were cropped to remove any cues to the individual's profession or identity (e.g., sporting equipment, sponsorship branding etc.).

From this stimulus set, 75 individuals were randomly selected as target identities. Of the three available images of each target identity, one image was randomly selected to be presented as the target face, while the remaining two images were presented as the distractor faces in the self-distractors condition (see Figure 6.1b). The distractor faces in the control condition were sampled by randomly selecting a single image from each of the 150 remaining identities in the stimulus set. The self-distractors task consisted of 225 trials: 75 control trials, 75 self-distractors trials, and 75 alone trials, which were intermixed and randomised. The self-distractors task had fewer trials than the identical-distractors task because there were fewer faces in the stimulus set; however, because the analyses were conducted using the average attractiveness ratings made by participants, the reduced number of trials was unlikely to affect the results.

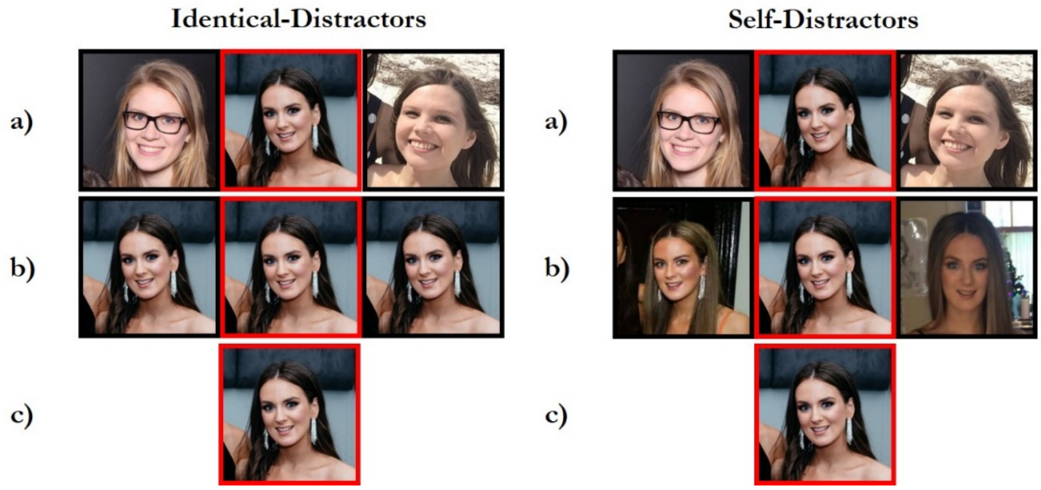

Figure 6.1. Stimulus configurations for the Identical-Distractors and Self-Distractors tasks a) control group trial b) identical-distractor trial/self-distractor trial c) alone trial. The *target* face was identified by a red frame [centre]. Please note, the faces in the figure are representative of those in the stimulus set, but were not themselves shown in the experiment.

Procedure

The procedure was the same for both tasks, and unless otherwise noted, was the same for all experiments reported in the current study. Participants were asked to rate the attractiveness of each target face, which would be identified by a red frame that appeared around a single face in the group. Participants completed six practice trials to familiarise themselves with the task. Initially, group images were presented without the target identified for 2000 ms [un-cued], and each face was surrounded by a black frame. During this time, participants were encouraged to look at each face in the group. A red frame then appeared around the target face for an additional 1000 ms [cued], after which time all images disappeared from the screen. During alone trials, the target face was presented for 1000 ms with a black frame and an additional 1000 ms with a red frame, before disappearing from the screen. Once all stimuli were removed from the screen, participants gave their attractiveness rating for the target face via mouse click along a visual analogue scale that ranged from "*Very Unattractive*" (0%) to "*Very Attractive*" (100%) (width

= 192 mm; 21.74°; Carragher et al., 2018). The trial procedure and stimulus presentation time replicates those used by Walker and Vul (Experiment 4; 2014; Carragher et al., 2018).

Analysis

A *cheerleader effect measure* was calculated separately for each group condition, by subtracting the attractiveness ratings given to the target faces when seen alone, from the attractiveness rating given to the same target faces seen in a group. A positive change score represents an increase in attractiveness when seen in a group (i.e., the cheerleader effect), whereas negative values indicate a decrease. This cheerleader effect measure was calculated in all experiments reported in the current study.

Data Availability

All data reported in the current study are available in the Open Science Framework repository, [https://osf.io/8sg62/?view_only=3e658c7391704afb9d9dba0e2ba48723].

Results

A 2 x 2 repeated-measures ANOVA with group condition (control, distractor manipulation) as a within-participants factor, and cheerleader effect task (identical-distractors, self-distractors) as a between-participants factor, was used to investigate whether the variance between the faces in the group influenced the size of the cheerleader effect. The main effect of group condition was non-significant, $F(1, 57) = 2.80$, $p = .100$, $\eta_p^2 = .047$, as was the main effect of cheerleader effect task, $F(1, 57) < .001$, $p = .983$, $\eta_p^2 < .001$. Crucially, the interaction between the group condition and cheerleader effect task was significant, $F(1, 57) = 6.13$, $p = .016$, $\eta_p^2 = .097$, (see Figure 6.2). Separate analyses were conducted for each cheerleader effect task to investigate the nature of the significant interaction.

Figure 6.2. The cheerleader effect in the control and distractor manipulation conditions, shown separately for participants in each cheerleader effect task in Experiment 1. Error bars represent the standard error of the mean (SEM).

Identical-Distractors Task

One sample *t*-tests were used to investigate whether the change to the attractiveness of the target face seen in a group was statistically significant. Consistent with the cheerleader effect, faces were perceived to be significantly more attractive in the control, $t(27) = 4.71$, *95%CI*[1.01, 2.56], $p < .001$, $d = 0.89$, and identical-distractors conditions, $t(27) = 3.18$, *95%CI*[0.28, 1.32], $p = .004$, $d = 0.60$. A paired-samples *t*-test was used to compare the size of the cheerleader effect between the two group conditions. The cheerleader effect was significantly larger in the control condition than in the identical-distractors condition, $t(27) = 2.59$, *95%CI*[0.20, 1.76], $p = .015$, $d = 0.49$.

Self-Distractors Task

One sample *t*-tests showed that faces were significantly more attractive in both the control, $t(30) = 3.24$, *95%CI*[0.45, 1.96], $p = .003$, $d = 0.58$, and self-distractors conditions, $t(30) = 3.21$, *95%CI*[0.51, 2.28], $p = .003$, $d = 0.58$. A paired-samples *t*-test indicated that the difference between the control and self-distractors conditions was non-significant, $t(30) = 0.65$, *95%CI*[-0.41, 0.79], $p = .520$, $d = 0.12$.

Discussion

Target faces were perceived to be significantly more attractive in the control condition in both tasks, replicating the cheerleader effect (Carragher et al., 2018; Walker & Vul, 2014). The increase in attractiveness was similar to the 1.5-2% increase observed by Carragher et al. (2018), suggesting that the cheerleader effect is robust. As predicted, the cheerleader effect also occurred in the self-distractors condition. Interestingly, the size of the effect did not differ significantly between the control and the self-distractors conditions, suggesting that the faces in the self-distractors group conditions contained a similar amount variability as the faces of different identities in the control condition (Burton et al., 2005; Jenkins et al., 2011; Todorov & Porter, 2014). Notably, the cheerleader effect was significantly reduced in the identical-distractors condition, which is consistent with the predictions of a hierarchical encoding mechanism. The reduction to the cheerleader effect in the identical-distractors condition cannot be attributed to the three faces in the group sharing the same identity, because that was also true of the self-distractors condition. Instead, our findings suggest that the cheerleader effect was significantly reduced in the identical-distractor condition because the group lacked the variance to create an attractive ensemble average.

Unexpectedly, a small increase in attractiveness still occurred in the identical-distractors condition, which is inconsistent with the predictions of a hierarchical encoding mechanism. This finding raises the question of whether the observed pattern is consistent with the predictions of a hierarchical encoding mechanism, or whether the cheerleader effect may, in fact, be due to another underlying mechanism. In Experiment 2, we explicitly tested whether the magnitude of the cheerleader effect is directly related to the characteristics of the average face of the group, as predicted by hierarchical encoding.

Experiment 2

In Experiment 1, we found mixed support for the notion that the cheerleader effect is caused by hierarchical encoding. To investigate whether the cheerleader effect is related to the characteristics of the ensemble average, we asked observers to make an attractiveness rating for the digitally averaged face of each group image. These "group average faces" were used to approximate the characteristics of the mentally summarised ensemble average for each group (de Fockert & Wolfenstein, 2009; Neumann et al., 2013).

Digitally averaging the faces in the group is commonly used to investigate the ensemble coding of identity (de Fockert & Wolfenstein, 2009; Neumann et al., 2013; Yamanashi Leib et al., 2014). Observers show remarkable sensitivity in these tasks, such that they will incorrectly "recognise" the set average as having been a member of the previous group, but not other digitally averaged faces (Kramer et al., 2015; Neumann et al., 2013). This finding strongly suggests that faces created by digitally averaging the individual faces in the group very closely approximate the ensemble average that is mentally summarised by the observer.

The attractiveness ratings given to the group average faces were used to test two explicit predictions derived from the hierarchical encoding mechanism (Walker & Vul, 2014). First, we tested whether the average face of each group was perceived to be significantly more attractive than the three individual faces used to create them (Langlois & Roggman, 1990; Rhodes et al., 1999). Second, we tested whether the size of the cheerleader effect was related to the difference in attractiveness between the target face and the average face of the group it was presented in. Consistent with hierarchical encoding (Griffiths et al., 2018), we predicted that individual faces that were much less attractive than the group average face would experience the largest increase in attractiveness in a group, whereas any individual face that was more attractive than the group average face would become less attractive in a group.

Method

Participants

Thirty-one participants (25 females, M_{age} = 19.97, *SD* = 4.12) were recruited for Experiment 2. All participants completed the task as instructed, and no cheerleader effect data fell further than 3*SD* from the condition mean. As such, no data were excluded from analysis.

Stimuli

Thirty female faces with neutral facial expressions were selected from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist et al., 1998). These 30 faces were randomly sorted to create 10 group images. To maximise the number of possible trials with 10 group images, we collected a cheerleader effect measure for each face in the group. As such, each group image was presented three times during the experiment, and a different face was identified as the target each time (Walker & Vul, 2014). The faces in each group appeared in the same location during repeated presentations of the same group image.

The average face for each of the 10 group images was created using Abrosoft FantaMorph Deluxe 5.4.8 (Abrosoft, 2016). The three faces in each group were weighted equally in the morphing procedure to create the average face for each group. The external features (e.g., hair, neck) of the original and average face stimuli were occluded by a black mask that was placed around each face, where evidence of digital averaging could potentially be noticed (see Figure 6.3). In total, the experiment consisted of 70 trials that were intermixed and randomised: 30 group trials, 30 alone trials, and 10 group average trials. The trial procedure was identical to Experiment 1.

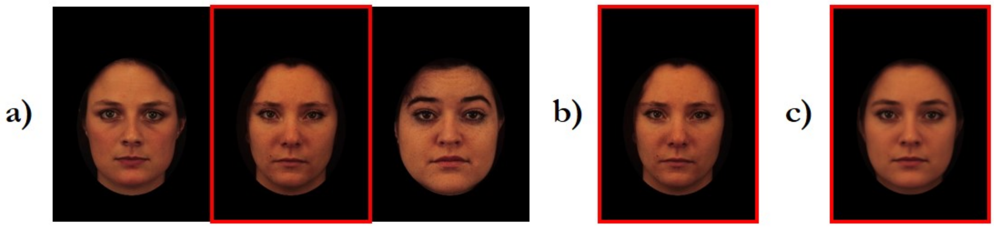

Figure 6.3. Example stimulus configurations in Experiment 2 a) group trial b) alone trial c) group average trial.

Analysis

To investigate whether hierarchical encoding causes the cheerleader effect, we created a second change measure, called the '*ensemble effect*'. The ensemble effect measure was calculated for each target face, by subtracting the attractiveness rating given to the individual face when it was seen alone (Figure 6.3b), from the attractiveness rating given to the average face of the group that the individual face was presented in (Figure 6.3c). Ensemble effect, therefore, shows how much more attractive the group average face was compared to each target face. We hypothesised that hierarchical encoding would be demonstrated by a *positive* correlation between ensemble effect and the size of the cheerleader effect. This positive correlation would show that target faces much less attractive than the group average (i.e., those with positive ensemble effect scores) receive the largest cheerleader effect, whereas any face more attractive than the group average (i.e., those with a negative ensemble effect) are perceived to be less attractive in a group.

Results

Group Average Faces

First, we examined whether the group average faces were perceived to be significantly more attractive than the individual faces from which they were composed. A paired samples *t*-test showed that participants perceived the group average faces to be significantly more attractive ($M = 52.53$, $SD = 13.54$) than the original target faces ($M = 35.96$, $SD = 11.49$), $t(30) = 9.96$, 95%CI[13.17, 19.97], $p < .001$, $d = 1.79$. Further examination of the attractiveness ratings showed that the 10 group average faces were perceived to be more attractive than all 30 original individual faces in the stimulus set ($M_{increase}$ = 16.57%, SD = 8.04; Range = 0.65%, 31.59%). Moreover, a series of within participants paired-samples *t*-tests were used to compare the attractiveness ratings given to each individual target face with those given to the average face of the group it appeared within. The group average faces were significantly more attractive than 26 of the 30 individual faces. These findings suggest that the mentally summarised ensemble average is likely perceived to be more attractive than each individual face in the group, as required by the hierarchical encoding mechanism.

Cheerleader Effect

A one sample *t*-test showed that the target faces were perceived to be significantly more attractive in a group compared to alone (M = 1.64, SD = 2.30), $t(30)$ = 3.97, 95%CI[0.79, 2.48], $p < .001$, d = .71, replicating the cheerleader effect.

Hierarchical Encoding

We tested whether hierarchical encoding contributes to the cheerleader effect by investigating the relationship between the size of the cheerleader effect and the measure of ensemble effect for each target face. In contrast to our predictions of a positive correlation between the two measures, the relationship between ensemble effect and the cheerleader effect was non-significant, $r(30)$ = .110, *95%CI*[-0.261, 0.452], p = .564, indicating the size of the cheerleader effect is not related to the attractiveness of the average face of the group (see Figure 6.4).

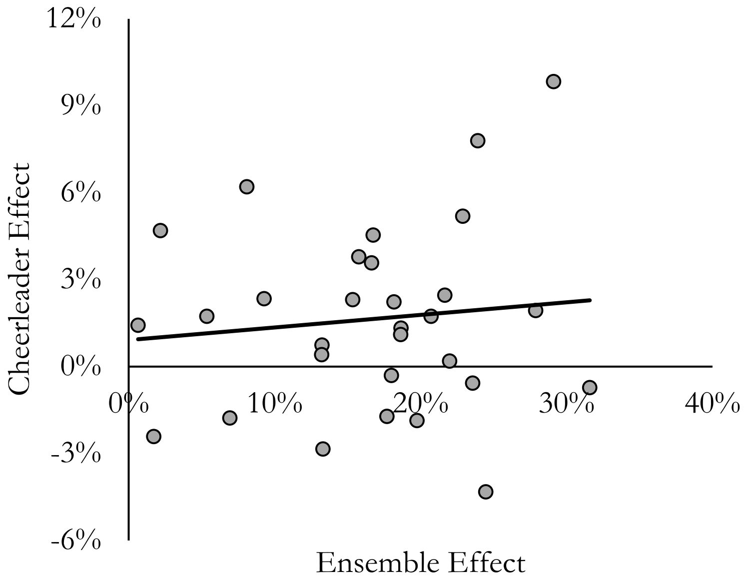

Figure 6.4. The relationship between the size of the cheerleader effect and the ensemble effect measure for each target face. Positive ensemble effect scores indicate that the group average was more attractive than the individual target face.

Discussion

The average face of each group was perceived to be significantly more attractive than the individual faces it was created from, replicating the findings of Langlois and Roggman (1990). Furthermore, the average increase in attractiveness for an individual face in a group was approximately 1.5-2%, replicating the size of the cheerleader effect in the two control conditions of Experiment 1 (Carragher et al., 2018). However, our data also clearly show that the size of the cheerleader effect for each target face was not related to the attractiveness of the average face of the group. Despite showing that the average face of each group was more attractive than each target face, and that the target faces were remembered to be more attractive in a group, our results suggest that the two phenomena are unrelated. Taken together, the findings of Experiments 1 and 2 do not support the proposition that the cheerleader effect is caused by hierarchical encoding (Walker & Vul, 2014), and suggest that exploration of an alternative mechanism is warranted. In considering the possible alternative causes of the cheerleader effect, we first sought to identify the boundaries of the effect. Consequently, we investigated whether the cheerleader effect is specific to groups of faces, or whether it would also occur when a human face was presented in a group of non-human stimuli.

Experiment 3

The primary aim of Experiment 3 was to test whether the cheerleader effect occurs for a face shown in a group with non-human distractor images. To this end, images of *houses* were presented in place of the distractor faces in the group (*house-distractors* condition). Houses were selected to replace the distractor faces because they are an easily identifiable, non-human stimulus that belong to a homogenous item category (Olson & Marshuetz, 2005). A significant cheerleader effect in the house-distractors condition would not only demonstrate that hierarchical encoding cannot cause the effect, because the ensemble representation of such a group would not be a human face, but also that the cheerleader effect does not rely on a face specific processing mechanism. On the other hand, a non-significant cheerleader effect in the house-distractors condition would suggest that the cheerleader effect relies on a mechanism that is specific to the processing of faces.

We also investigated whether the cheerleader effect itself is specific to human faces. To this end, participants also made attractiveness judgments about target *houses*, which were presented once in a group with two distractor *houses*, and once alone. Curiously, previous investigations have suggested that food products (Cooremans & Geuens, 2015) and oranges (Harp, Haberman, & Whitney, 2009) are perceived to be more attractive when they are seen within a group of similar items compared to alone. Therefore, we predicted that the cheerleader effect would also occur for images of houses. A significant cheerleader effect for target houses would provide further evidence that the cheerleader effect does not rely on a face specific processing mechanism.

Method

Participants

Twenty-nine participants (25 females, M_{age} = 22.52, *SD* = 9.28) were recruited for Experiment 3. All data were within 3*SD* of the condition mean and all participants completed the task as instructed. As such, no data were excluded from analysis.

Stimuli

A new stimulus set consisting of images of houses was collected from various real-estate websites. All houses included in the stimulus set were two-story, featured windows on the top level of the house, and were free of other notable stimuli (e.g., people, cars). From this stimulus set, 60 target houses were randomly selected to be presented twice: once in a group with two unique distractor houses (house-target condition), and once alone (alone house trial; see Figure 6.5). The procedure for responding to house targets was identical to that for faces.

From the stimulus set of Carragher et al. (2018), 60 target faces were randomly selected to be presented 3 times: once in a control group, once in a group where the distractor images were houses (*house-distractors* condition), and once alone. To discourage participants from fixating the only human face in a house-distractors group, 60 additional filler trials were included in the experiment, wherein a house was identified as the target in a group that contained a single human face and two houses (see Figure 6.5c). These filler trials were not of theoretical importance and were discarded prior to analysis. Both target *face* and target *house* trials were randomised and intermixed throughout the experiment, which consisted of 360 trials: 60 control trials (6.5a), 60 house-distractors trials (6.5b), 60 filler trials (6.5c), 60 alone face trials (6.5d), 60 house-target trials (6.5e) and 60 alone house trials (6.5f).

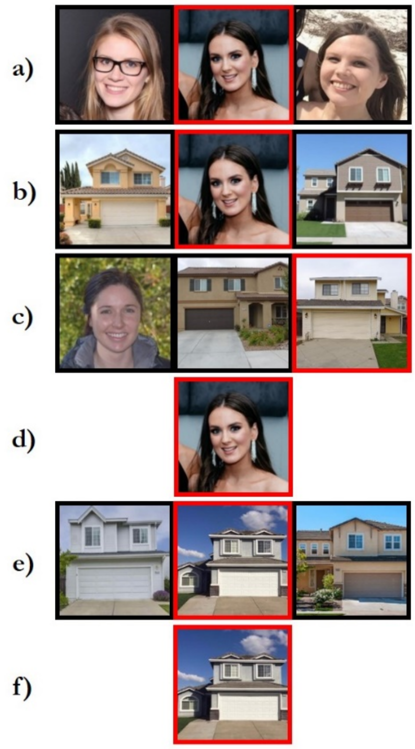

Figure 6.5. Stimulus configurations in Experiment 3 a) control trial b) house-distractors trial c) filler trial [target right] d) alone face trial e) house-target trial f) alone house trial.

Results

One sample *t*-tests showed that faces were significantly more attractive in the control, $t(28) = 3.27$, *95%CI*[0.55, 2.40], $p = .003$, $d = 0.61$, and house-distractors conditions, $t(28) = 2.42$, *95%CI*[0.12, 1.39], $p = .022$, $d = 0.45$ (see Figure 6.6). A paired-samples *t*-test showed that the cheerleader effect was significantly larger in the control condition than the house-distractors condition, $t(28) = 2.14$, *95%CI*[0.03, 1.42], $p = .041$, $d = 0.40$. Remarkably, a one sample *t*-test also showed that houses were also significantly more attractive when shown in a group of houses, compared to alone, $t(28) = 2.81$, *95%CI*[0.32, 2.02], $p = .009$, $d = 0.52$.

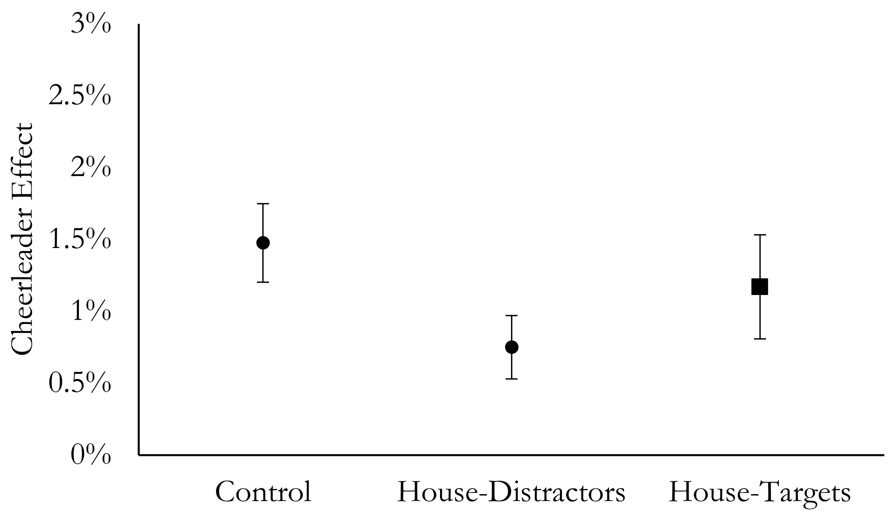

Figure 6.6. The cheerleader effect in the control, house-distractors, and house-target conditions [square] in Experiment 3. Error bars represent the within-participants SEM (Cousineau, 2005).

Discussion

The cheerleader effect occurred in the control condition, replicating the results of Experiments 1 and 2. Interestingly, the cheerleader effect was significantly reduced, but still statistically significant, when target faces were shown in the house-distractors condition. The results of the present study are conceptually similar to those of Experiment 1, wherein the cheerleader effect was significantly reduced, but not eliminated, when target faces were presented in the *identical-distractors* condition. Notably, neither the house-distractors or identical-distractors groups could be summarised to create an attractive ensemble average, suggesting that the cheerleader effect is unrelated to the average characteristics of the group.

Remarkably, target houses were also perceived to be significantly more attractive when shown in a group of houses compared to when they were seen alone. This finding is consistent with previous studies that have found food products (Cooremans & Geuens, 2015) and oranges (Harp et al., 2009) to be more attractive in a group compared to alone. Taken together, our findings not only demonstrate that the cheerleader effect is not caused by a face specific processing mechanism, but that the cheerleader effect occurs for non-human items.

Experiment 4

The cheerleader effect was significantly reduced, but not eliminated, when a face was presented with two houses. When combined with the results from Experiments 1 and 2, which show that the effect does not appear to be caused by hierarchical encoding, our findings strongly suggest that another mechanism underlies the cheerleader effect. One possibility might be that an individual is perceived to be more attractive when shown with houses, because houses are socially desirable stimuli that are associated with significant monetary value and social status. Both males and females who possess luxurious or valuable items are treated favourably in social interactions (Nelissen & Meijers, 2011), and are perceived to have favourable personal attributes (Hennighausen, Hudders, Lange, & Fink, 2016). Perhaps an individual is rated to be more attractive when presented with two houses, because the houses signal that the individual is successful or wealthy, and observers infer that they have positive personal attributes. To test whether the cheerleader effect is related to the desirability of the items in the group, we investigated whether a face would be perceived to be more attractive in a group that consisted of socially neutral images. Target faces were presented once in a control group, once in a group with two house-distractors (*socially desirable*), and once in a group with two chair-distractors (*socially neutral*).

We predicted a replication of Experiment 3, whereby the cheerleader effect would be significantly larger in the control condition than in the house-distractors condition. We also predicted that the cheerleader effect would be significantly larger in the control condition than in the chair-distractors condition. If the cheerleader effect occurs when the target face is presented with house-distractors (replicating Experiment 3), but not with chair-distractors, our results would be consistent with the suggestion that the cheerleader effect is related to the desirability of the group that the individual is seen within. However, a significant cheerleader effect in the chair-distractors condition would be inconsistent with this notion.

Method

Participants

Thirty-one participants (25 females, M_{age} = 21.68, *SD* = 5.55) were recruited for Experiment 4. Participants with a cheerleader effect further than 3*SD* from the condition mean (*n* = 1) were excluded from analysis. The final sample consisted of 30 participants.

Stimuli

A new stimulus set consisting of images of chairs was collected from the websites of online furniture retailers. All chair images featured a single chair that was the focal point of the image, and were free of other notable or desirable stimuli (e.g., people, computers). We selected chairs as a suitable non-desirable stimulus category because, like houses, chairs are easily identifiable and belong to a discrete item class. Most importantly, being seen with a chair does not typically signal that somebody has desirable attributes, such as being friendly or likable (i.e., faces), or signal wealth or status (i.e., houses).

From the stimulus set of Carragher et al. (2018), 36 target faces were randomly selected to be presented 4 times throughout the experiment; once in a control group, once in a group with two house-distractors (socially desirable), once in a group with two chair-distractors (socially neutral), and once alone (see Figure 6.7). As in Experiment 3, dummy group trials were included for house-distractors (*n* = 9) and chair-distractors (*n* = 9) trials. Dummy targets were also presented once alone, to give them the appearance of critical trials. All dummy trials were discarded prior to analysis. All trial conditions were intermixed and randomised. In total, each participant completed 180 trials: 36 control trials (6.7a), 36 house-distractors trials (6.7b), 36 chair-distractors trials (6.7c), 36 alone trials (6.7e), and 36 dummy trials (6.7d).

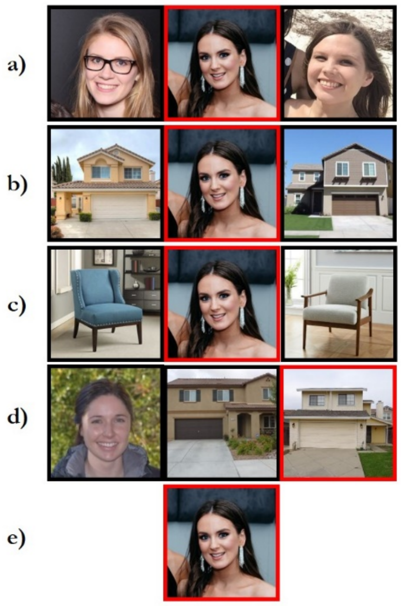

Figure 6.7. Example stimulus configurations Experiment 4 a) control trial b) house-distractors trial c) chair-distractors trial d) filler trial e) alone trial.

Results

One sample *t*-tests showed that target faces were significantly more attractive in the control condition, $t(29) = 4.27$, *95%CI*[0.76, 2.14], $p < .001$, $d = 0.78$, but not in the house-distractors, $t(29) = 1.18$, *95%CI*[-0.34, 1.28], $p = .246$, $d = 0.22$, or chair-distractors conditions, $t(29) = 0.72$, *95%CI*[-1.15, 0.55], $p = .477$, $d = 0.13$ (see Figure 6.8).

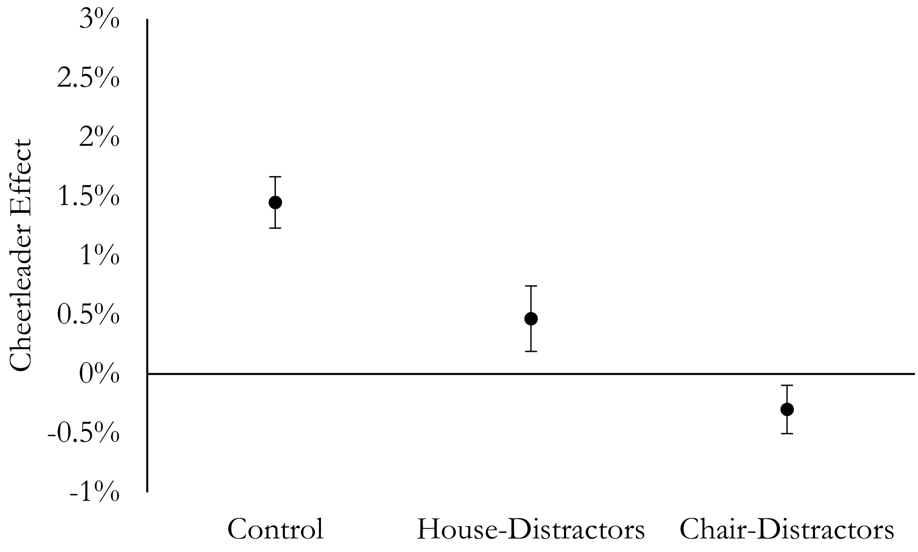

Figure 6.8. The cheerleader effect for each distractor condition in Experiment 4. Error bars represent the within-participants SEM (Cousineau, 2005).

A repeated-measures ANOVA with distractor condition (control, house-distractors, chair-distractors) as a within-participants factor was used to investigate whether the cheerleader effect was modulated by distractor condition. The main effect of distractor condition was significant, $F(2, 58) = 9.31, p < .001, \eta_p^2 = .243$. Planned comparisons using paired samples *t*-tests showed that the cheerleader effect was significantly larger in the control condition than in the house-distractors condition, $t(29) = 2.17$, *95%CI*[0.05, 1.91], $p = .039$, $d = 0.40$, and the chair-distractors condition, $t(29) = 5.51$, *95%CI*[1.10, 2.40], $p < .001$, $d = 1.01$. The change in attractiveness was not significantly different between the house-distractors and chair-distractors conditions, $t(29) = 1.76$, *95%CI*[-0.12, 1.66], $p = .088$, $d = 0.32$.

Comparison with Experiment 3

Unlike Experiment 3, the cheerleader effect did not occur for target faces in the house-distractors condition. To clarify whether a significant increase in attractiveness occurs when the distractor images in the group are houses, we compared the data from the control and house-distractors conditions in Experiments 3 and 4.

A 2 x 2 repeated-measures ANOVA with distractor condition (control, house-distractors) as a within-participants factor and Experiment (3, 4) as a between-participants factor was used to investigate whether the size of the cheerleader effect differed between the control and house-distractor conditions. Crucially, the main effect of distractor condition was significant, $F(1, 57) = 8.99$, $p = .004$, $\eta_p^2 = .136$. The main effect of Experiment was non-significant, $F(1, 57) = .118$, $p = .733$, $\eta_p^2 = .002$, as was the interaction between the two factors, $F(1, 57) = .203$, $p = .654$, $\eta_p^2 = .004$. As the ANOVA showed that the data did not differ between experiments, we collapsed the data from each experiment into single control and houses-distractors measures.

One samples *t*-tests showed that faces were significantly more attractive in the control condition, $t(58) = 5.25$, *95%CI*[0.90, 2.02], $p < .001$, $d = 0.68$, and in the house-distractors condition, $t(58) = 2.42$, *95%CI*[0.11, 1.11], $p = .019$, $d = 0.32$. Finally, a paired samples *t*-test showed that the increase in attractiveness was significantly larger in the control condition ($M = 1.46$, $SD = 2.14$) compared to the house-distractor condition ($M = 0.61$, $SD = 1.93$), $t(58) = 3.03$, *95%CI*[0.29, 1.42], $p = .004$, $d = 0.39$. These results are consistent with those reported in Experiment 3, and suggest that a significant increase in attractiveness does occur when a single face is presented with two houses, but that the increase is significantly smaller than the typical cheerleader effect.

Discussion

Interestingly, our findings were consistent with a cheerleader effect that is modulated by the relative social value of the distractor images in the group. The cheerleader effect was largest when the target face was surrounded by other faces, reduced for houses, and was eliminated when the distractor images were chairs. These results suggest that social desirability is a plausible alternative mechanism to hierarchical encoding (Walker & Vul, 2014). The basis of the social desirability mechanism and its application to the cheerleader effect are explored in greater detail in the general discussion.

While the results of the current experiment are consistent with the newly proposed social desirability mechanism, we conducted one final experiment to investigate whether presentation time, which is also known to influence the attractiveness of faces (Rashidi, Pazhoohi, & Hosseinchari, 2012), influences the size of the cheerleader effect. Specifically, we investigated whether a discrepancy between the time that the target face is presented for in the group and alone trials might contribute to the size of the cheerleader effect.

Experiment 5

Each experiment in the present study has used the same stimulus presentation timing, which was first introduced by Walker and Vul (Experiment 4; 2014). During group image presentations, the group is initially presented without the target face identified for 2000 ms (un-cued), before the target face is cued for an additional 1000 ms. If we assume that the observer disperses their gaze equally across the 3 faces in the group during the un-cued presentation time, and then exclusively fixates the target face once cued, the target face can be viewed for approximately 1667 ms during a group trial. Yet during alone trials, the target is presented for 2000 ms. Increased presentation time has previously been shown to decrease the liking of faces (Gerger, Forster, & Leder, 2017). Crucially, the same face is also perceived to be less attractive after being seen for 1000 ms compared to 20 ms (Saegusa & Watanabe, 2016; Willis & Todorov, 2006), or having been viewed for 5000 ms compared to 200 ms (Rashidi et al., 2012). Together, these findings raise the possibility that within the cheerleader effect paradigm, faces might be perceived to be more attractive in the group condition than alone, simply because they are seen for less time. To investigate this possibility, we manipulated the amount of time that the target was presented for in the group condition.

In their original investigation of the cheerleader effect, Walker and Vul (2014) also investigated whether a difference in presentation time might contribute to the effect. Reducing the presentation time of the alone condition to match that of the group condition did not influence the size of the cheerleader effect (Walker & Vul, 2014). However, each group image was presented three times, so that a cheerleader effect measure could be collected for each face in the group. Walker and Vul's (2014) presentation time manipulation was not adjusted to account for the repeated presentations of each group image (Walker & Vul, 2014). It is possible that observers might alter their gaze pattern across a group image that has already been seen, to avoid fixating a face that has already been identified as a target. Consequently, the amount of time an observer spent examining any one target face in the group might have been influenced by the number of times that the same group image had previously been presented. The current study overcomes this potential limitation as each group image was only presented once, meaning that observers do not have an incentive to distribute their gaze unevenly across the faces in the group.

In addition to addressing the existing presentation time discrepancy between the group and alone conditions, the current experiment was also designed to investigate whether any manipulation to the presentation time of the group image would modulate the size of the cheerleader effect. In a between-participants design, observers could experience *moderate* or *extreme* changes to the un-cued presentation duration of the group image. For participants in the *moderate* task, the un-cued presentation time of the group images was manipulated to be 1000 ms (short condition), 2000 ms (control condition), or 3000 ms (long condition). Crucially, in the long presentation time condition, if the observer gazed equally across the three faces in the group during the un-cued presentation time (3000 ms), and then exclusively fixated the target face once cued (1000 ms), the target face could be examined for the same amount of time as in the alone condition (2000 ms). Therefore, a significant reduction in the size of the cheerleader effect in the long condition compared to control, would indicate that the discrepancy between the presentation time of the group image and the alone image contributes to the cheerleader effect.

A second sample of participants experienced *extreme* manipulations to the un-cued group presentation duration, to explore whether the cheerleader effect is limited by any temporal constraints. For these participants, the un-cued presentation time of the group images was further decreased to 300 ms (extremely-short condition), and increased to 7000 ms (extremely-long condition). We selected 300 ms for the un-cued presentation time in the extremely-short condition to allow participants the opportunity for saccadic eye movements across the group image, so that each face could be briefly examined (Crouzet et al., 2010; Rayner, 1998). We selected 7000 ms as the presentation time of the extremely-long condition because the observer could potentially gaze at the target face for 3334 ms, which is twice the amount of time available in the control condition. Consistent with previous findings (Rashidi et al., 2012; Saegusa & Watanabe, 2016), we expected attractiveness ratings would decrease as observers were given more time to examine the faces in the group. If presentation time influences the cheerleader effect, the largest effect should occur for the shortest presentations and decrease as the presentation duration of the group image increases.

Method

Participants

Sixty-six participants (54 females, M_{age} = 22.71, *SD* = 6.77) were randomly allocated to either the *moderate* or *extreme* presentation time tasks. Participants with a cheerleader effect score further than 3*SD* from the condition mean (*n* = 1) were excluded from analysis. Additionally, participants with incomplete data due to technical failure were excluded from analysis (*n* = 1). The final sample consisted of 64 participants, evenly divided between the *moderate* (27 females, M_{age} = 23.78, *SD* = 8.78) and *extreme* (26 females, M_{age} = 21.78, *SD* = 4.06) presentation time tasks.

Stimuli

From the stimulus set of Carragher et al. (2018), 99 faces were randomly selected as targets for both presentation time tasks. Each target face was presented twice; once in a group and once alone. Observers saw each target face in one presentation time condition only; however, target images were balanced between-participants, such that each target face was rated in all presentation time conditions. The experiment was blocked according to presentation time condition, and the alone and group trials for each target image were presented within the same experimental block. Each presentation time block contained 66 trials: 33 group trials and 33 alone trials. In total, each participant completed 198 trials: 99 group trials [33 (extremely-)short condition, 33 control condition, 33 (extremely-)long condition], and 99 alone trials.

Procedure

Except for the manipulation to the un-cued presentation time of the group, the trial procedure was identical to previous experiments. Regardless of the un-cued presentation duration of the group image, each target face was cued for an additional 1000 ms (as in Experiments 1-4). As always, the target was presented for 2000 ms in the alone condition.

Results

One sample *t*-tests showed that faces were significantly more attractive in each *moderate* presentation time condition: short, $t(31) = 3.50$, *95%CI*[0.72, 2.73], $p = .001$, $d = 0.62$, control, $t(31) = 3.74$, *95%CI*[0.87, 2.95], $p = .001$, $d = 0.66$, and long, $t(31) = 3.13$, *95%CI*[0.43, 2.05], $p = .004$, $d = 0.55$ (see Figure 6.9). One planned paired samples *t*-test was used to investigate whether the cheerleader effect was significantly larger in the control condition compared to the long condition. The difference between the cheerleader effects in the control and long presentation time conditions was not significant, $t(31) = 1.17$, *95%CI*[-0.50, 1.82], $p = .252$, $d = 0.21$.

One sample *t*-tests also showed that faces were significantly more attractive in each *extreme* presentation time condition: extremely-short, $t(31) = 2.16$, *95%CI*[0.08, 2.58], $p = .038$, $d = 0.38$, control, $t(31) = 3.77$, *95%CI*[0.76, 2.54], $p < .001$, $d = 0.67$, and extremely-long, $t(31) = 3.08$, *95%CI*[0.41, 2.01], $p = .004$, $d = 0.54$.

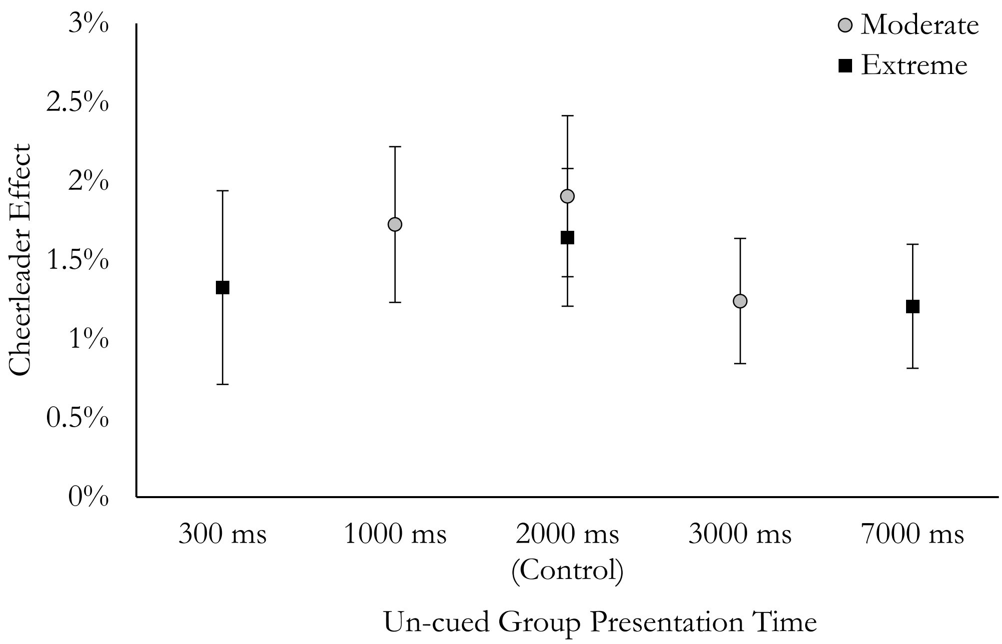

Figure 6.9. The cheerleader effect for each presentation time condition in Experiment 5. Error bars represent the SEM.

A mixed-measures ANOVA with presentation time (short, control, long) as a within-participants factor, and task manipulation (moderate, extreme) as a between-participants factor, was used to investigate whether the duration of the un-cued group presentation influenced the size of the cheerleader effect. The main effect of presentation time was non-significant, $F(2, 124) = .948$, $p = .390$, $\eta_p^2 = .015$, as was the main effect of task manipulation, $F(1, 62) = .217$, $p = .643$, $\eta_p^2 = .003$. The interaction between the two factors was similarly non-significant, $F(2, 124) = .107$, $p = .899$, $\eta_p^2 = .002$. Our findings strongly suggest that the duration of un-cued presentation time does not modulate the size of the cheerleader effect.

Discussion

The cheerleader effect was observed in each presentation time condition, regardless of the duration of the un-cued presentation time. Crucially, the non-significant difference between the cheerleader effects in the long and control presentation time conditions demonstrates that the existing discrepancy between the presentation time of the target in the group and alone conditions does not meaningfully contribute to the cheerleader effect. Furthermore, we found that the cheerleader effect occurred both when the group was presented un-cued for an extremely brief time (300 ms), and when the available viewing time of the target far outweighed that in the alone presentation time condition (7000 ms). Taken together, our results suggest that presentation time has very little influence on the size of the cheerleader effect.

Cheerleader Effect Meta-Analysis

Each experiment in the current study contained an identical control condition. As such, we performed a meta-analysis to compare the size of the cheerleader effect across experiments, with the aim of identifying the size of the *typical* cheerleader effect for future research. The control condition was formally defined to be: a group image consisting of a human target face and the faces of two unique identities as distractor images, presented initially with an un-cued duration of 2000 ms, and a target cue duration of 1000 ms. As Experiments 1 and 5 used between-participants designs [1a) identical-distractors, 1b) self-distractors, 5a) moderate presentation time, 5b) extreme presentation time], there were seven unique control data sets available for analysis.

In addition to investigating the size of the typical cheerleader effect, we also performed an exploratory analysis to investigate whether the sex of the participant influenced the size of the cheerleader effect. To date, there has been no investigation of the influence of participant sex on the size of the cheerleader effect (Carragher et al., 2018; Walker & Vul, 2014). Because we did not recruit our participant samples to balance sex, only the meta-analysis contained enough male participants to conduct a reliable analysis. Across five experiments, control condition cheerleader effect measures were available for 213 unique participants (170 females, M_{age} = 22.45, SD = 7.85).

A factorial ANOVA with Experiment (1a, 1b, 2, 3, 4, 5a, 5b) and participant sex (female, male) as between-participants factors, was used to investigate whether the size of the cheerleader effect differed across experiments or participant sex. The main effect of Experiment was non-significant, $F(6, 199) = .338$, $p = .916$, $\eta_p^2 = .010$, (see Figure 6.10). Similarly, the main effect of participant sex was also not significant, $F(1, 199) = .248, p = .619, \eta_p^2 = .001$. The cheerleader effect occurred among both female, ($M = 1.63$, $SD = 2.32$), $t(169) = 9.19$, *95%CI*[1.28, 1.99], $p < .001$, $d = 0.71$, and male observers, ($M = 1.40$, $SD = 2.22$), $t(42) = 4.14$, *95%CI*[0.72, 2.08], $p < .001$, $d = 0.63$. Finally, the interaction between Experiment and participant sex was also non-significant, $F(6, 199) = .521, p = .792, \eta_p^2 = .015$. As such, all data were collapsed into a single cheerleader effect measure.

A one sample *t*-test was used to identify the typical effect size for the cheerleader effect. Across all experiments, the same face was perceived to be approximately 1.6% more attractive in a group compared to alone ($M = 1.59$, $SD = 2.30$, Range = -3.61, 10.04), $t(212) = 10.09$, *95%CI*[1.28, 1.90], $p < .001$, $d = 0.69$.

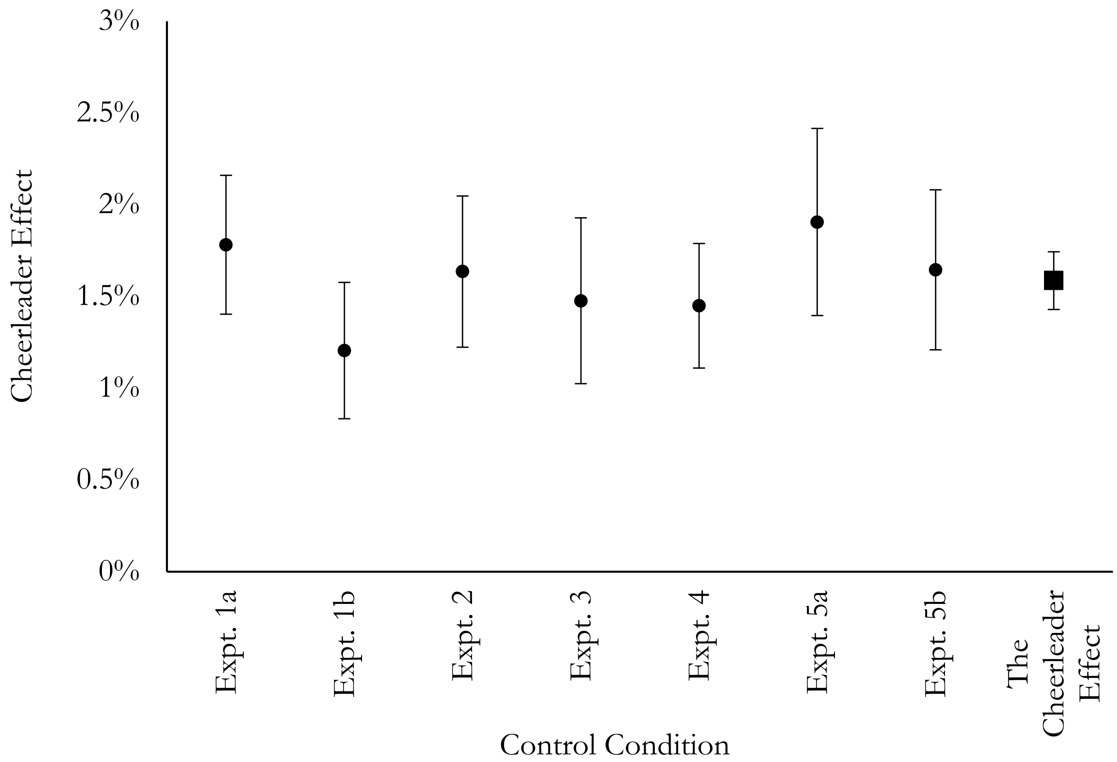

Figure 6.10. The cheerleader effect in the control condition of each experiment. The square [far right] shows the average size cheerleader effect. Error bars show SEM.

General Discussion

Our meta-analysis provides the first estimate of the typical increase to the attractiveness of an individual when seen within a group, which is approximately 1.6%. The increase in attractiveness identified in the meta-analysis is very similar to the 1.5-2% increase reported by Carragher et al. (2018), suggesting that the cheerleader effect is remarkably consistent. We have also shown, for the first time, that the cheerleader effect occurs consistently for male and female observers. We anticipate that the results of our meta-analysis will be useful for future research, both to establish whether the cheerleader effect has been observed, and to identify whether manipulations are successful in modulating the size of the cheerleader effect.

Previous investigations of the cheerleader effect have proposed that an individual is perceived to be more attractive in a group because of hierarchical encoding, whereby the individual is recalled as being more alike a highly attractive ensemble average (Walker & Vul, 2014). Consistent with hierarchical encoding, we found that the cheerleader effect was significantly larger for group conditions that could be summarised to create an ensemble average with facial characteristics more average than the component faces, compared to group conditions that could not (e.g., identical-distractors, house-distractors). In this respect, our results provide partial support for the role of a hierarchical encoding mechanism in the cheerleader effect (Walker & Vul, 2014). However, we also found that the size of the cheerleader effect for each target face was not related to the attractiveness ratings given to the average face of the group. A hierarchical encoding mechanism also cannot explain why the cheerleader effect was only reduced, but not eliminated, in group conditions where the summary representation of the group should not be highly attractive, or even a human face. Thus, our results suggest that we must reconsider the role of hierarchical encoding in the cheerleader effect and explore alternative explanations. Interestingly, Luo and Zhou (2018) also suggested that hierarchical encoding might not underlie the cheerleader effect, after showing in an ensemble coding task that observers extracted an ensemble average that was actually slightly *less* attractive than true average of the group.

Having demonstrated that hierarchical encoding is not necessary for the cheerleader effect to occur, we turned our attention toward identifying other possible causes of the effect. Our findings provide support for the idea that the cheerleader effect is modulated by the social value of the images in the group. The proposed social desirability mechanism is supported by an evolutionary phenomenon called "mate choice copying", which occurs when an individual increases their preference for a potential mate upon learning that the potential mate is desired by others (Gouda-Vossos et al., 2018; Little et al., 2011c). By preferring potential mates who have already been endorsed by others (i.e., their current partners), mate choice copying is a strategy that is used to maximise the chances of choosing a high quality mate who has desirable qualities (Gouda-Vossos et al., 2018; Little et al., 2011c). While mate choice copying behaviours are displayed by both male and female observers (Little et al., 2008; Place et al., 2010), the phenomenon has primarily been investigated among females, who place greater value on traits in male partners that are not clearly shown in the face (e.g., education, good financial prospects; Buss & Barnes, 1986). Mate choice copying is displayed when female observers rate male faces to be more attractive when they are seen with a female partner compared to alone (Gouda-Vossos et al., 2018), or when seen with an attractive, compared to an unattractive, female partner (Little et al., 2008; Little et al., 2011a).

While mate choice copying has previously been shown to occur in specific circumstances related to opposite-sex mate selection (Gouda-Vossos et al., 2018; Little et al., 2011a), our meta-analysis demonstrated that both male and female observers perceived a female target face to be more attractive in a group of female faces than alone. Although uncommon in the mate choice copying literature (Gouda-Vossos et al., 2018), males and female observers provide similar attractiveness ratings when faces of either sex are seen in isolation (Langlois et al., 2000), suggesting that similarities between the sexes should not be unexpected when evaluating faces in social groups. The similar outcomes of mate choice copying and the cheerleader effect (i.e., an increase in attractiveness in a group) raises the possibility that the two effects are caused by similar mechanisms.

Rodeheffer et al. (2016) showed that female observers engaged in mate choice copying because they attributed more desirable qualities (e.g., trustworthy, caring) to the same male when he was seen with a female partner compared to alone. Males are also attributed more favourable personality characteristics when they are seen with attractive, compared to unattractive, females (Bar-Tal & Saxe, 1976; Sigall & Landy, 1973; Winegard, Winegard, Reynolds, Geary, & Baumeister, 2017). Together, these findings have been used to suggest that mate choice copying occurs because the value of the potential mate's current partner, is used to infer the likely personal qualities of the potential mate (Rodeheffer et al., 2016). When applied to the current investigation, observers might infer that an individual who is seen within a social group has been accepted into the group by the other members. Therefore, being seen in a group might suggest that the individual possesses desirable characteristics, such as being friendly or likable, which, in turn, could make them appear to be more attractive (Rodeheffer et al., 2016).

A social desirability mechanism, whereby individuals are attributed favourable characteristics based on the value of the other images in the group, can also explain why individuals were rated to be more attractive when shown with houses, but not chairs. In addition to attributing favourable characteristics to individuals based on the characteristics of their current mate (Hennighausen et al., 2016; Rodeheffer et al., 2016), favourable characteristics, such as attractiveness, are also attributed to both males and females who own material possessions that signal wealth or social status (Hennighausen et al., 2016; Hudders, De Backer, Fisher, & Vyncke, 2014; Nelissen & Meijers, 2011). Notably, this effect is much larger for male than female faces (Wang et al., 2018), as males, but not females, are perceived to be more attractive when shown inside luxury apartments (Dunn & Hill, 2014), sitting in luxury cars (Dunn & Searle, 2010), or wearing luxury clothes (Townsend & Levy, 1990). Perhaps an individual is perceived to be more attractive when presented with two houses, because the observer infers that they own the houses, which signals the high social status of the individual (Dunn & Hill, 2014). On the other hand, the attractiveness of an individual was unchanged when seen with chairs, perhaps because chairs are unlikely to signal social status.

Of course, the newly proposed social desirability should initially be considered with caution. We did not measure the perceived desirability of friends, houses, and chairs, nor did we investigate the personal characteristics that they are perceived to signal about an individual. While future research is needed to test this social desirability mechanism directly, such a mechanism can account for a larger proportion of the current results than the previously suggested hierarchical encoding mechanism (Walker & Vul, 2014). The results of the current study suggest that social desirability offers a viable alternative explanation for the cheerleader effect.

There are, however, results in the current study that are potentially inconsistent with the newly proposed social desirability mechanism. It is not clear why an effect of social desirability would cause the cheerleader effect to be reduced when the group consists of identical images of the same person, but not different images of the same person (although one possibility might be that observers did not realise that these different images were of the same person, e.g., Jenkins et al., 2011). Furthermore, social desirability also cannot convincingly explain why a house is perceived to be more attractive when seen in a group of other houses. Although the simplest explanation for the cheerleader effect is one that relies on a single mechanism, neither hierarchical encoding nor social desirability appear to adequately explain the cheerleader effect alone. Therefore, the possibility that two (or more) separate mechanisms contribute to the cheerleader effect must also be considered.

Although our results are inconsistent with a cheerleader effect that is caused by a single hierarchical encoding mechanism, it should be considered that ensemble coding, the prerequisite process for hierarchical encoding, occurs automatically upon viewing groups of faces (Fischer &

Whitney, 2011; Haberman & Whitney, 2011). Therefore, it is likely that the individual faces within each group are summarised to create an ensemble representation, which could influence the observer's memory for the target face (Brady & Alvarez, 2011; Griffiths et al., 2018). One possibility that cannot be discounted, is that our results might be consistent with a two-mechanism account of the cheerleader effect; wherein hierarchical encoding is one mechanism, which is accompanied by a second mechanism that also increases the attractiveness of images seen in groups. In support of this proposition, our results show that the size of the cheerleader effect fluctuates when hierarchical encoding is possible. An increase in attractiveness of approximately 1.6% is observed in group conditions that can be summarised to create an attractive ensemble average (i.e., control and self-distractors conditions), whereas a significantly smaller, but still statistically significant, increase in attractiveness of 0.6-0.8% is observed in group conditions that should not be subject to hierarchical encoding (i.e., identical-distractors and house-distractors conditions). If we assume that two underlying mechanisms exist, these results suggest that approximately only half of increase in attractiveness caused by the cheerleader effect can be attributed to hierarchical encoding. Notably, since an increase in attractiveness was found for faces in groups of non-human stimuli, this second mechanism would not necessarily be specific to the processing of faces. Although speculative, future research might consider whether the cheerleader effect is caused by two separate mechanisms.

While investigating the possible causes of the cheerleader effect, we examined whether the longer presentation of the target in the alone condition relative to the group condition contributed to the magnitude of the cheerleader effect. Allowing observers more time to examine the target face in the group image did not influence the size of the cheerleader effect, suggesting that presentation time is not likely contributing to the size of the effect in the control condition. Rather, our results are consistent with Walker and Vul (2014), who also reported that the cheerleader effect was not influenced by matching the presentation time of the group and alone trials. Interestingly, we found that the cheerleader effect persisted when the presentation time of the group images was substantially increased, beyond the viewing time available in the alone condition. These findings are inconsistent with the notion that increasing presentation duration decreases the perceived attractiveness of faces (Rashidi et al., 2012; Saegusa & Watanabe, 2016). Instead, our results raise the interesting possibility that the cheerleader effect might persist regardless of how long a group is viewed.

The evaluations made during first impressions are long lasting (Chang et al., 2010), and can influence social decision-making (Carragher et al., 2017; Rezlescu et al., 2012). Attractive individuals are often treated favourably (Dion et al., 1972), such that they are perceived to be intelligent (Eagly et al., 1991), are more likely to be trusted (Wilson & Eckel, 2006), and have more dating experience than unattractive individuals (Langlois et al., 2000). It is imperative that future research investigates the many social implications that may arise from the cheerleader effect. If an individual is perceived to be more attractive when they are seen in a group due to the cheerleader effect, are they also attributed the positive stereotypes that are associated with attractiveness? If the cheerleader effect does result in consistently positive stereotypes for an individual seen in a group, then it might be advantageous for individuals to present themselves in groups more often. For instance, an individual could select group photographs for their social media profiles, to increase their perceived attractiveness, which in turn, might increase the success of their online dating (Langlois et al., 2000) or professional profiles (Eagly et al., 1991).

The cheerleader effect is a robust visual phenomenon that consistently results in an individual being rated to be approximately 1.6% more attractive when seen in a group compared to when seen alone. This effect is not gender specific, with both male and female observers demonstrating the cheerleader effect to the same degree. Furthermore, the cheerleader effect is not likely caused by a face specific processing mechanism, as it also occurs during the evaluation of non-human items. We have shown that the cheerleader effect is not consistent with a single hierarchical encoding mechanism. Rather, our results suggest that the cheerleader effect might be related to the positive characteristics that can be attributed to an individual when they appear to be endorsed by others or are associated with items that signal social status. The robust nature of the cheerleader effect suggests that social context is an important factor to consider in the study of facial attractiveness, which raises the possibility that other trait impressions might also be influenced by social context.

Chapter 7: General Discussion

7.1 Summary

This thesis has advanced our understanding of the influence that social context has on judgments of attractiveness and trustworthiness; a line of research that has long been overlooked in the field of social perception. Chapter 3, however, reflects the original line of inquiry that this thesis was intended to follow, which was to investigate whether asymmetries in emotional expressions would result in the left cheek being perceived as more trustworthy than the right cheek (Zaidel et al., 2003). Contrary to the predictions, the cheek shown most prominently by each model did not influence the amount of money that they received in an economic game, suggesting that trustworthiness is not lateralised in the face. There was, however, a robust female trustworthiness premium, whereby all participants sent female models approximately 20% more money than the male models. The clear absence of a trustworthiness asymmetry using this experimental approach led me to reconsider the viability of this original project. Ultimately, I redirected my attention toward the cheerleader effect.

Despite the widespread popularity of the effect in popular culture (Rashid & Fryman, 2008), the cheerleader effect had not been replicated since Walker and Vul's (2014) original study. In Chapter 4, I investigated whether the cheerleader effect could be replicated, and explored whether the effect was modulated by known visual asymmetries. Most importantly, this study not only replicated the basic cheerleader effect, but found it to be remarkably consistent. Individual faces were rated to be significantly more attractive in a group, regardless of whether the target or distractor faces were presented in the left or right visual fields. The absence of a visual field asymmetry in the cheerleader effect was inconsistent with the prediction that observers would over-attend to the left visual field due to the effect of pseudoneglect (Bowers & Heilman, 1980; Guo et al., 2009; McGeorge et al., 2007). It is consistent, however, with the hypothesised function of ensemble coding, which is to rapidly extract summary statistics from large visual scenes (Alvarez, 2011; Alvarez & Oliva, 2008; Whitney & Yamanashi Leib, 2018).

In Chapter 5, I investigated whether faces also become more trustworthy when seen in a group. These experiments showed that social context had a dissociable influence on judgments of attractiveness and trustworthiness. Regardless of their attractiveness when seen alone, nearly all faces became more attractive when they were in a group. On the other hand, only extremely untrustworthy faces became more trustworthy in a group, whereas many highly trustworthy faces were perceived to be less so. These results suggest that the systematic increase in attractiveness that is typical of the cheerleader effect, might be a pattern that is specific to attractiveness judgments only. The final experiment of Chapter 5 showed that the size and direction of the cheerleader effect that each target face experienced was not related to the attractiveness of the digitally averaged face of the group that it appeared within. These findings were inconsistent with the predictions derived from a hierarchical encoding mechanism (Walker & Vul, 2014), which led me to directly investigate the underlying cause of the cheerleader effect in Chapter 6.

In Chapter 6, I conducted a series of experiments to investigate whether the cheerleader effect is caused by hierarchical encoding, as initially proposed by Walker and Vul (2014). These experiments showed that cheerleader effect was largest when a human face was shown alongside the faces of two different individuals, reduced when a face was shown in a group with images of houses or identical photographs of the target, and non-existent when the group consisted of images of chairs. Because hierarchical encoding can only cause an increase in attractiveness when the images in the group can be summarised to create an average face, these results demonstrate that a single hierarchical encoding mechanism cannot account for the cheerleader effect. Instead, I proposed an alternative explanation for the cheerleader effect, wherein favourable personal characteristics, such as attractiveness, are attributed to an individual based on the perceived value or desirability of the other images in the group. Finally, a combined analysis of the seven unique control conditions reported in Chapter 6 provided the first estimate of the typical size of the cheerleader effect, which is an increase in attractiveness of approximately 1.6% for an individual seen in a group ($d = 0.69$).

In the following discussion, I will briefly reiterate the shortcomings of Walker and Vul's (2014) hierarchical encoding mechanism, and discuss several alternative explanations for the cheerleader effect that are also inconsistent with these data. I will expand upon the discussion in Chapter 6 to consider whether the cheerleader effect is an extension of mate choice copying, and discuss whether the effect could be caused by the newly proposed social desirability mechanism. I will then further explore whether the cheerleader effect might be caused by two separate mechanisms, and identify two additional factors embedded within the cheerleader effect paradigm that could be contributing to the increase in attractiveness for faces in a group. I also describe how a conventional hierarchical encoding paradigm could be used to clarify whether hierarchical encoding has any role in the cheerleader effect. Finally, I will outline several promising avenues for future research into the cheerleader effect and the influence of social context on social perception in general.

7.2 Hierarchical Encoding and The Cheerleader Effect

Walker and Vul (2014) proposed that the cheerleader effect was the result of a complex interaction between ensemble coding, the attractiveness of average faces, and the hierarchical structure of visual working memory. The central premise of the hierarchical encoding mechanism is that the target face is recalled as being highly similar to the ensemble average, which itself is perceived to be highly attractive, because it possesses average facial characteristics. Indeed, several of the current results are consistent with such a mechanism. The size of the cheerleader effect was significantly reduced in group conditions that could not be summarised to create an average face, including when a target face was presented among groups of houses or chairs (Experiments 3 & 4, Chapter 6). Most importantly, the cheerleader effect was also significantly reduced when the group consisted of identical images of the target face, but not different photographs of the target (Experiment 1, Chapter 6). The key difference between these two conditions is that only the group consisting of different photographs of the target (self-distractors) contains any variability between the faces in the group, which is required to create a face with average facial characteristics (Burton et al., 2005). Because these two group conditions were otherwise the same (e.g., a group consisting of three images of the same identity), this finding provides the strongest evidence that the cheerleader effect is related to the characteristics of the ensemble average. However, despite explaining the reduced cheerleader effect for groups that cannot be averaged, many of the other findings in this thesis are inconsistent with a hierarchical encoding mechanism (Walker & Vul, 2014).

If the cheerleader effect is caused by hierarchical encoding, the size of the cheerleader effect that each face experiences in a group should be directly related to the attractiveness of the ensemble average. Digitally averaged faces were created to approximate the mentally summarised ensemble average (Experiment 3, Chapter 5; Experiment 2, Chapter 6). Although the average face of each group was rated to be more attractive than each individual group member, and the cheerleader effect was observed, the size of the cheerleader effect for each target face was not correlated with the attractiveness of the average face of the group (Experiment 2, Chapter 6). A similar result was found when digitally averaged faces were created from group images in which the attractiveness of the target face and the distractor faces had been manipulated (Experiment 3, Chapter 5). The average face of a group that consisted of two unattractive distractors and an attractive target face was rated to be significantly less attractive than the attractive target face when it was seen alone; however, the attractive target face was not rated to be significantly less attractive when seen in this unattractive group. On the other hand, the average face of a group that consisted of two attractive distractors and an unattractive target face, was rated to be significantly more attractive than the unattractive target face when it was seen alone; yet, a significant cheerleader effect did not occur when the unattractive target face was seen in an attractive group. Finally, it should also be considered that a smaller cheerleader effect still occurred when target faces were shown with identical-distractors (Experiment 1, Chapter 6) and house-distractors (Experiments 3 & 4, Chapter 6), even though neither group could be mentally summarised to create an average face. Therefore, these results strongly suggest that hierarchical encoding alone cannot cause the cheerleader effect.

The cheerleader effect also appears to be inconsistent with both contrast and assimilation effects. Assimilation effects occur when the attractiveness of an individual is raised or lowered to match the attractiveness of the other faces in the group (Bar-Tal & Saxe, 1976; Geiselman et al., 1984; Sigall & Landy, 1973), whereas contrast effects occur when the attractiveness of an individual is biased away from the attractiveness of the other group members (Kenrick & Gutierres, 1980; Melamed & Moss, 1975). Crucially, if the cheerleader effect was caused by either effect, individuals would have been rated to be significantly *less* attractive in specific group configurations. A contrast effect would cause unattractive target faces to become significantly less attractive in a group with attractive distractors. On the other hand, an effect of assimilation would cause attractive faces to become significantly less attractive in an unattractive group. Yet, an increase in attractiveness occurred regardless of the attractiveness of the target or the distractor faces in the group (Figure 5.4, Experiment 1, Chapter 5). Notably, even if the increase in attractiveness failed to reach significance (Experiments 2 & 3, Chapter 5), the target faces were never rated to be significantly *less* attractive in a group. These results demonstrate that neither contrast nor assimilation can explain the cheerleader effect, which is best characterized as a consistent increase in attractiveness for individuals seen in a group.

Before discussing the factors that might contribute to the cheerleader effect (sections 7.3, 7.4), I will briefly outline several additional alternative explanations for the effect that are inconsistent with the data collected thus far. First, it is unlikely that the cheerleader effect is due to increasing familiarity with the target faces over repeated presentations (Peskin & Newell, 2004), because the effect occurs when the target face is shown twice in a randomised order (group, alone; e.g., Experiment 1, Chapter 5). Second, the cheerleader effect occurred when the left-right orientation of the response scale anchors ("very unattractive/very attractive") was counterbalanced between participants (Chapter 4), indicating that the effect is not caused by a spatial congruency between the target location and the rating scale (Nicholls et al., 2006; Simpson & Thomas, 2018). Third, the cheerleader effect occurred when the presentation time of the group image (7000 ms) greatly exceeded that of the alone condition (2000 ms; Experiment 5, Chapter 6), suggesting that the effect does not occur because the observer has less time to examine the target face in the group image (Rashidi et al., 2012; Saegusa & Watanabe, 2016). Fourth, the cheerleader effect is not likely to be driven by a subset of images (see Figure 5.4), because new target and distractor faces were selected randomly before each experiment. Moreover, faces were never rated to be significantly less attractive in a group, discounting the possibility of regression to the mean, or that participants were simply unable to use the response scale consistently. Finally, social context had a dissociable effect on judgments of attractiveness and trustworthiness, suggesting that the cheerleader effect is simply not a generalised positive response bias that occurs for faces in groups. Although the underlying cause of the cheerleader effect remains unknown, the results of these experiments will be valuable to future researchers investigating cause of the effect.

7.3 Mate Choice Copying, Social Desirability and The Cheerleader Effect

Beneath the pop-culture inspired name, the cheerleader effect simply shows that observers perceive a face to be more attractive in a group of faces compared to alone. As discussed in Chapter 6, the cheerleader effect shares many similarities with mate choice copying (Gouda-Vossos et al., 2018; Little et al., 2011c; Place et al., 2010; Waynforth, 2007), which raises the question, is it possible that the two effects are related? One key difference between the two effects is that mate choice copying appears to only occur under specific circumstances (Gouda-Vossos et al., 2018; Little et al., 2011a). For example, Rodeheffer et al. (2016) found that females only increased their ratings of desirability for a male face when the accompanying female face was described as his current partner, but not his sister. Similarly, Waynforth (2007) also showed that females only increased their attractiveness ratings when the male face was presented with an attractive, but not an unattractive, female. The increase in attractiveness for these male faces is not consistent with a generalised phenomenon, because the basic composition of the image was the same in each circumstance (i.e., a male and female face pair). Instead, these findings suggest that mate choice copying only occurs in contexts wherein the value of the potential mate's current partner can offer clues about the desirability of the potential mate to others (Rodeheffer et al., 2016). On the other hand, the cheerleader effect appears to occur whenever a face is presented in a group, regardless of the attractiveness of the target face, or the attractiveness of the group. Therefore, unlike mate choice copying, the cheerleader effect appears to be a generalisable phenomenon that occurs due to the mere presence of the distractor faces in the group.

However, it should be considered that the research paradigms used to test mate choice copying are often designed to be highly specific. Most studies only recruit heterosexual participants, who are asked to evaluate opposite-sex faces, often within an explicitly stated mate selection scenario (e.g., potential short-term or long-term partner; Little et al., 2008). Therefore, the generalisability of mate choice copying has not yet been clearly established. To my knowledge, only one study has investigated whether heterosexual observers perceive a same-sex face to be more attractive when paired with an opposite-sex face. After initially showing that the heterosexual observers displayed conventional mate choice copying behaviours toward opposite-sex faces that were paired with attractive same-sex faces, Little et al. (2011a) asked the same observers to rate the attractiveness of the same-sex face in the image pair. The attractiveness ratings for same-sex faces did not differ depending on the attractiveness of the opposite-sex partner, suggesting that the increase in attractiveness that is typical of mate choice copying does not generalise beyond mate choices.

In contrast to the findings of Little et al. (2011a), the meta-analysis reported in Chapter 6 showed that the typical cheerleader effect occurred for both female and male observers. These findings demonstrate that female observers rated female faces to be significantly more attractive in a group with female faces than when alone. Importantly, the participants in the present studies were not recruited based on their sexuality, nor were they asked about their mate preferences. As such, the participant samples reported here are assumed to be representative of the majority heterosexual general population. Because most of these observers were unlikely to be making mate selections, the cheerleader effect cannot strictly be an example of mate choice copying. However, the robust nature of the cheerleader effect clearly demonstrates that social context has a much more generalised influence on attractiveness judgments than has previously been suggested by mate choice copying.

Even if they are not intrinsically linked, the similar outcome to the cheerleader effect and mate choice copying (i.e., an increase of attractiveness for an individual seen in a group) might suggest that the two effects have a similar underlying cause. Indeed, the social desirability mechanism of the cheerleader effect that was proposed in Chapter 6 is based upon the mechanism proposed to cause mate choice copying. If favourable personal qualities are attributed to potential mates based upon the quality of the others who desire them (Bar-Tal & Saxe, 1976; Rodeheffer et al., 2016; Sigall & Landy, 1973; Winegard et al., 2017), then it is possible that being accepted into a social group also signals than an individual has desirable qualities. However, as stated in Chapter 6, the newly proposed social desirability mechanism must be treated cautiously until it can be directly tested.

To test plausibility of the social desirability mechanism, it will be necessary to investigate whether observers actually perceive groups of friends and houses to be desirable, but perceive groups of chairs to be socially neutral. Second, future research must also investigate whether observers attribute favourable characteristics, such as increased attractiveness or social status, to individuals that are presented in groups with faces or houses (Rodeheffer et al., 2016). Finally, future research might consider whether appearing in a group of socially undesirable images makes an individual significantly less attractive compared to when they are seen alone. A challenge for such a design will be to identify categories of stimuli that are undesirable to most observers. One potential approach might be to compare the size of the cheerleader effect for faces shown among groups of houses that are either picturesque or dilapidated, which presumably differ on their value and perceived desirability. A significant decrease in attractiveness for faces shown groups of dilapidated houses would be consistent with the social desirability mechanism. Although the social desirability mechanism is undoubtedly speculative, the findings in Chapter 6 offer preliminary evidence that the cheerleader effect is related to the positive attributes that might be inferred about an individual based upon the desirability of the group they are seen within.

7.4 Could Two Mechanisms Contribute to The Cheerleader Effect?

As noted in Chapter 6, there are findings reported in this thesis that the social desirability mechanism cannot explain. Notably, the same faces become more attractive, but not necessarily more trustworthy in a group (Experiment 1, Chapter 5). This dissociation is difficult to understand within a social desirability framework, since one would assume that trustworthiness is a favourable trait that is signalled by group membership (Bar-Tal & Saxe, 1976; Rodeheffer et al., 2016). Therefore, much like the hierarchical encoding mechanism (Walker & Vul, 2014), it appears that the proposed social desirability mechanism also cannot account for the complete pattern of results reported here. To the best of my knowledge, no single mechanism can.

As discussed in Chapter 6, the possibility that a second mechanism contributes to the cheerleader effect must be considered. Below, I expand upon this suggestion and outline two methodological factors within the cheerleader effect paradigm, which might inadvertently increase the attractiveness ratings given to individuals seen in groups. Importantly, either suggested factor could contribute to the cheerleader effect independently of a hierarchical encoding or a social desirability mechanism. Therefore, finding evidence in favour of either proposed factor is not, in and of itself, additional evidence that either hierarchical encoding or social desirability underlies the cheerleader effect.

7.4.1 Sample Size Bias and Ensemble Amplification

A significant increase in attractiveness occurred for faces shown in identical-distractors and house-distractors groups, when an average summary face could not be created (Experiments 1 & 3, Chapter 6). Having shown that this increase in attractiveness was not likely caused by the presentation time of the stimulus (Experiment 5, Chapter 6), one of the only remaining factors common to both conditions was simply that a human face was presented in a group with two distractor images. One possibility is that faces are perceived to be more attractive in a group compared to alone, simply because the group contains more images. Sample size bias is a pervasive phenomenon, whereby observers give larger estimates of the average characteristics of the group as the number of items in the group increases (Price, 2001; Price, Kimura, Smith, & Marshall, 2014; Price, Smith, & Lench, 2006; Smith & Price, 2010). Smith and Price (2010) suggest that the number of items in the group primes a general magnitude heuristic (Pelham, Sumarta, & Myaskovsky, 1994), which causes observers to overestimate the characteristics of the group on dimensions that are themselves unrelated to the magnitude of the group. For example, participants who were required to draw long pencil lines on a piece of paper before an experiment, subsequently gave longer estimates of the length of the Mississippi river, and higher estimates of the average temperature of Honolulu, compared to those who initially drew short lines (Oppenheimer, LeBoeuf, & Brewer, 2008).

Sample size bias influences estimates of average risk of heart attack (Price, 2001), height (Price et al., 2006), duration of time (Smith, Rule, & Price, 2017), and even arithmetic means (Smith & Price, 2010). For example, observers perceive the average height of identical stick figures to be taller when they are presented in a group of 12, compared to a group of 4 (Price et al., 2006). Observers also estimate the arithmetic mean of 20 numbers to be significantly greater than that of a set of 5 numbers, even though both have the same true mean (Smith & Price, 2010). Crucially, observers have also been found to be affected by sample size bias in conventional ensemble coding tasks (Price et al., 2014). When presented with sets of 3, 6, 9, and

12 identical squares, estimates of the mean size of the group increased linearly as a function of group size (Price et al., 2014). This sample size bias persisted when the squares were identically sized, and when the sets were presented sequentially, demonstrating that it is the number of items in the set that biases responses, rather than observers mistakenly attending to the largest squares in each set, or the total surface area of the display (Price et al., 2014).

In a similar phenomenon, Kanaya, Hayashi, and Whitney (2018) reported that ensemble coding is prone to "ensemble amplification", wherein observers give larger estimates of average circle size as the number of circles in the set increases. Unlike sample size bias, which is directly related to the numerosity of the group (Smith & Price, 2010), Kanaya et al. (2018) show that ensemble amplification occurs as a result of the way that the visual system extracts summary statistics from the display. The most recent findings suggest that summary statistics are formed by subsampling and integrating a number of items that is approximately equal to the square root of the total number of items in the display (Whitney & Yamanashi Leib, 2018). However, rather than sampling these items randomly, Kanaya et al. (2018) found that observers preferentially sampled the largest circles in the set, suggesting that these large circles had an inherent saliency to the observers. These findings suggest that because the visual system only subsamples the most salient items in the group to extract summary statistics, estimates of the group's average characteristics will become less accurate as the size of the group increases. When considered together, both sample size bias and ensemble amplification suggest that the number of items in the set can directly influence the judgment made by the observer.

Despite the many examples of sample size bias occurring in a wide variety of judgments (Price, 2001; Price et al., 2014; Price et al., 2006; Smith & Price, 2010; Smith et al., 2017), Walker and Vul (2014) reported that a similar sized cheerleader effect for groups of 4, 9, and 16 faces. However, it is possible that Walker and Vul (2014) did not find an effect of group size because the size of the group was increased by adding novel faces to each group image. Adding novel faces to the group not only makes the group larger, but also makes the ensemble average more "average". Although this method could potentially increase the likelihood of observing a larger cheerleader effect, the procedure confounds any effect that sample size bias might have with the increasing averageness of the set.

To investigate whether sample size bias itself contributes to the cheerleader effect, it is necessary to increase the number of faces in the group without also increasing the averageness of the group. One possible approach would be to test whether increasing attractiveness judgments are given to a target face that is presented in sets of increasing size (e.g., 1, 3, 6, 9, 12), which consist only of identical photographs of the face itself. Because these sets would only contain one face that is presented multiple times, any increase in attractiveness could only be attributed to the increasing number of faces in the set. The results of this sample size bias experiment could then be used to establish the increase in attractiveness that is attributable to the size of the group. An effect of sample size bias could explain why a small increase in attractiveness is found when a face is presented with identical-distractors (Experiment 1, Chapter 6) or house-distractors (Experiments 3 & 4, Chapter 6), and why houses are also perceived to be more attractive when presented with other houses (Experiment 3, Chapter 6). However, it should be noted that the trustworthiness of a face was not necessarily increased in a group, which is potentially inconsistent with the predicted effect of sample size bias (Experiment 1, Chapter 5). Nonetheless, given the number of paradigms and judgments that are influenced by sample size bias, one cannot help but think that sample size bias might contribute to the cheerleader effect.

7.4.2 Attentional Cueing

In the basic cheerleader effect paradigm, the group image is presented for the observer to examine freely, before a red frame is presented around a single target face that will be evaluated. The red frame serves as a cue to direct the observer's attention toward the face that requires evaluation. However, simply cueing attention toward a face might also make the face appear more attractive (Störmer & Alvarez, 2016). In a forced choice task, two identical faces appeared on either side of a fixation cross for 58 ms, and observers indicated which of the faces was more attractive. Prior to the stimulus presentation, a cue was flashed for 70 ms at the location of one of the two faces. Even though the faces were identical, Störmer and Alvarez (2016) found that the face that had been cued was selected as being more attractive more often than the face that was not cued. Although there are notable differences between Störmer and Alvarez's (2016) paradigm and the cheerleader effect paradigm, such as the cue preceding stimulus onset, these findings raise the possibility that simply cuing attention toward the target face in the group image might inadvertently make it appear to be more attractive. Of course, a red cue was also flashed around the target face in the alone condition of the cheerleader effect paradigm; however, this cue is likely redundant because attention is already directed toward the single face on the screen. An effect of cued attention might offer an alternative explanation for why an increase in attractiveness occurs in group conditions that are incompatible with the suggested mechanisms of hierarchical encoding or social desirability.

Cuing attention is necessary to investigate the cheerleader effect; after all, without a cue, the observer will not know which face they are supposed to rate. Removing the cue from the cheerleader effect paradigm provides an interesting methodological challenge for future research. One option would be to modify the trial procedure to retroactively cue the target face, by presenting a cue at the spatial location of the target face after the group image is removed from the screen (see Figure 7.1; Griffiths et al., 2018). Alternatively, the procedure could be modified to remove the un-cued viewing of the group entirely. Rather than cueing a single target face from the group, a different coloured cue could be presented around each face in the group sequentially (see Figure 7.1). After each face in the group has been cued, the group would be removed from the display, and a colour prompt would signal which face in the group was the target. One additional benefit of this colour-cueing procedure would be to ensure that the observer fixates each face in the group, which is not required in the current paradigm. While both of these suggestions have the potential to eliminate any attention induced increase in attractiveness, any task with a retroactive cue would likely place a much higher working memory load on the participant, who would be required to correctly remember the characteristics of the retroactively cued face. Consequently, it is not clear how eliminating a possible attentional cueing effect, while simultaneously increasing working memory load, might inadvertently affect task performance.

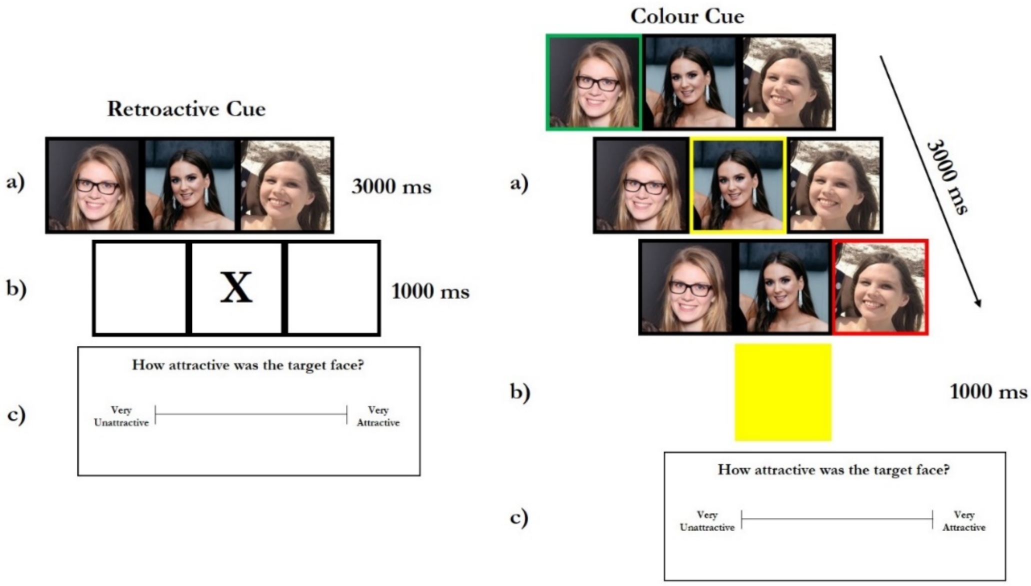

Figure 7.1. Two hypothetical trial procedures for cheerleader effect tasks with retroactive target cues. Retroactive Cue a) the group is presented for 3000ms b) a cue is presented at the location of the target face c) an attractiveness rating is made. Colour Cue a) each face in the group is presented with a different colour frame for 1000 ms b) a colour prompt identifies the target face to be rated c) an attractiveness rating is made.

7.5 Measuring The Cheerleader Effect with a Hierarchical Encoding Paradigm

Despite proposing that the cheerleader effect is caused by ensemble coding and the hierarchical structure of visual working memory, the paradigm introduced by Walker and Vul (2014) does not explicitly test the occurrence of either process. Instead, the increased attractiveness rating for an individual within a group is taken as evidence that hierarchical encoding has occurred. Although the experiments reported in this thesis have advanced our understanding of the cheerleader effect, they were conducted using the same limited paradigm (Walker & Vul, 2014). Despite showing that hierarchical encoding is unlikely to be the only cause of the cheerleader effect, it is still unclear whether it has any role in the phenomenon. As noted in Chapter 6, the automatic nature of ensemble coding (Fischer & Whitney, 2011; Haberman & Whitney, 2011), means that the group image is still likely summarised into an ensemble representation, which could then influence the recall of the target face from the group (Griffiths et al., 2018). Determining whether hierarchical encoding has *any* role in the cheerleader effect will require the implementation of a new paradigm that can measure the effect of hierarchical encoding on the recall of a single face (Griffiths et al., 2018; Haberman & Whitney, 2007, 2009; Luo & Zhou, 2018).

The ensemble coding of identity is often measured using a paradigm introduced by de Fockert and Wolfenstein (2009), wherein the observer is presented with a set of four faces, which is immediately followed by a single memory probe (see Figure 2.3). The memory probe can be a set exemplar, a foil exemplar, a set average, or a foil average. Incorrectly "recognising" the set average (which has never been seen) as a member of the preceding group is taken as evidence that ensemble coding has occurred. This measure of ensemble coding could be integrated into the existing cheerleader effect paradigm by presenting observers with a memory probe for the identity of the target face, after the observer has made their attractiveness rating for the target. Hierarchical encoding would be suggested if observers simultaneously demonstrated the cheerleader effect and incorrectly endorsed the set average as the target face they were asked to rate; however, because a separate response is required to measure the cheerleader effect and hierarchical encoding, this design could only show that the two effects co-occur, but not that they are causally related.

Despite being quite easy to implement, de Fockert and Wolfenstein's (2009) task represents a relatively coarse measure of ensemble coding, because ensemble coding can only be demonstrated by endorsing the set average. This method would be particularly limited in the cheerleader effect paradigm because the observers are cued to attend toward a single target face within the group. Directing attention toward the target face not only makes the ensemble average more alike the cued item (de Fockert & Marchant, 2008), but can also increase memory for the attended target face (LaRocque et al., 2015). Therefore, observers might correctly reject the set average as having been the target face; but would still recall the target face to be *slightly* more alike the set average if the task required a more sensitive discrimination. This limitation could be addressed by employing a task that uses a morph sequence of faces to capture the range of possible identities between the individual target face and the set average (Haberman & Whitney, 2007, 2009; Luo & Zhou, 2018).

Morph sequences have been extensively used by Haberman and Whitney (2007, 2009) to investigate the ensemble coding of emotion in faces. Luo and Zhou (2018) used a similar technique to create a continuum of attractiveness, by morphing between a very unattractive face and a very attractive face. I propose that a similar approach could be taken to investigate the cheerleader effect. First, several extremely unattractive faces would be averaged together to create an unattractive composite (e.g., 0% attractiveness; see Figure 7.2a), and several extremely attractive faces would be averaged to create an attractive composite (e.g., 100%). Notably, even though it is average, an average face created from many unattractive faces is still perceived to be significantly less attractive than an average face created from many attractive faces (Perrett et al., 1994). Initially creating unattractive and attractive average faces improves upon Luo and Zhou's (2018) approach, since averaging together many very unattractive faces will remove the idiosyncratic factors that make each face unattractive, while exemplifying the shared traits that make all faces unattractive (Sutherland et al., 2013). A sequence of morphed faces could then be interpolated between these extremely unattractive and extremely attractive average faces, ultimately creating a complete sequence of faces that varies on attractiveness (see Figure 7.2a). By employing a stimulus set that systematically varies on attractiveness, this proposed design removes the need to collect an explicit attractiveness rating from the observer about each target face. Instead, both the cheerleader effect and hierarchical encoding can be measured by having observers match the identity of the memory probe to the target face, to investigate whether probes that are more attractive than the target are endorsed as having been the target face.

Creating a single morph sequence of attractiveness also addresses one of the most elusive elements of the cheerleader effect: the non-average attractiveness of average faces. As seen in Experiment 2 of Chapter 6, a digitally averaged face was rated to be more attractive than each of the three original individuals it was created from (Langlois & Roggman, 1990; Rhodes et al., 1999). However, because the faces in an attractiveness sequence already represent the different levels of attractiveness between two fixed anchor faces, there should be no additional increase in attractiveness when these faces are averaged together. Therefore, the average face of the group will simply be the average attractiveness value of the group. Notably, if the group can no longer be summarised to create a highly attractive ensemble average, hierarchical encoding would be demonstrated by a bidirectional change in attractiveness. Faces less attractive than the group average would increase in attractiveness, whereas faces more attractive than the group average would decrease in attractiveness.

The experimental procedure required to measure the cheerleader effect within a hierarchical encoding paradigm would closely follow that used by Griffiths et al. (2018) to measure hierarchical encoding in emotional expressions. A set of four faces would be randomly selected from the attractiveness continuum, which would vary around a mean value not shown within the group (e.g., a set composed of 50%, 55%, 65% and 70% has a mean attractiveness of 60%; see Figure 7.2b). A single face from the group would be randomly cued as the target face (e.g., 50%). After a brief interval, a single memory probe face would follow, and observers would make a forced choice response to indicate if the probe showed the target face (i.e., yes/no). These memory probes would be sampled from the faces either side of the target on the attractiveness continuum (i.e., ranging from 15 faces more attractive than the target to 15 less attractive). The same trial procedure would occur for single face presentations, which would be used to establish baseline accuracy in target discrimination (Figure 7.2c).

Luo and Zhou's (2018) findings suggest that the "yes" responses in the alone condition should be normally distributed around the target face, demonstrating that observers can accurately identify the attractiveness of the face when presented alone. Hierarchical encoding would be demonstrated if the "yes" response distributions for target faces less attractive than the group average were shifted toward endorsing probes more attractive than the target, whereas target faces more attractive than the group average would be recalled to be less attractive (Figure 7.2d). On the other hand, if the most attractive faces in each group were still recalled to be more attractive in a group (which would be consistent with the findings reported in this thesis), this result would demonstrate that the cheerleader effect is not caused by hierarchical encoding. To understand what role, if any, ensemble coding plays in the cheerleader effect, future research should implement a new paradigm that can measure both the cheerleader effect and hierarchical encoding.

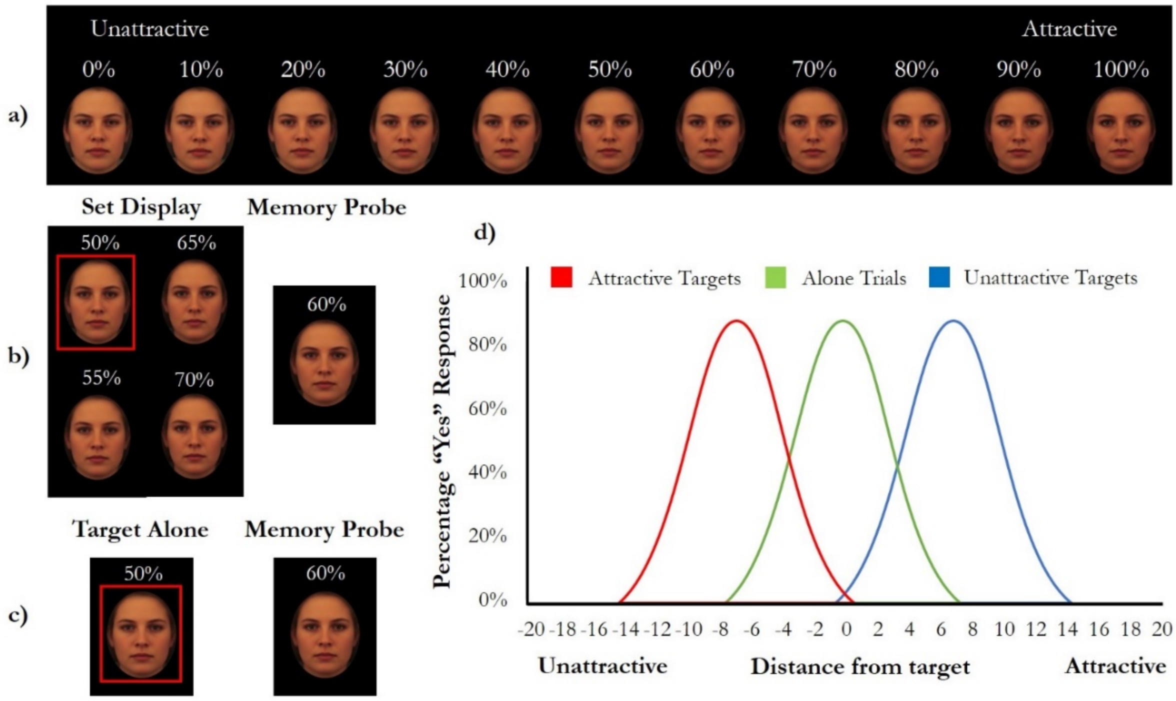

Figure 7.2. A hypothetical experiment that uses a hierarchical encoding paradigm to measure the cheerleader effect a) a sequence of faces that represent an attractiveness continuum. The least attractive face (0%) and most attractive face (100%) were created by averaging together the six least or six most attractive faces in the KDEF database[13] (Lundqvist et al., 1998) b) the group presentation condition, with cue, and subsequent memory probe c) the alone presentation condition and subsequent memory probe d) the hypothesized pattern of results that would demonstrate the hierarchical encoding of facial attractiveness.

[13] Attractiveness ratings for the KDEF were acquired from the website of Alexander Todorov's Social Perception Laboratory at Princeton University (https://tlab.princeton.edu/databases/).

7.6 Outstanding Questions and Future Directions

The cheerleader effect is a relatively new phenomenon that raises questions not only about the influence of social context on judgments of facial attractiveness, but about the influence of social context on all trait impressions that are made from the face. As the cheerleader effect has only been investigated by Walker and Vul (2014) and in the present thesis, there are many questions that still need to be addressed. In the section below, I outline several of these outstanding questions that can be addressed with future research.

7.6.1 Are The Cheerleader Effect and The Group Attractiveness Effect Related?

The cheerleader effect (Walker & Vul, 2014) and the group attractiveness effect (van Osch et al., 2015) are similar phenomena, which demonstrate that social context can modulate perceptions of facial attractiveness. The cheerleader effect refers to an individual being perceived to be more attractive in a group than alone, whereas the group attractiveness effect refers to observers overestimating the attractiveness of a whole group, compared to the average of the attractiveness ratings given to the individual group members. Published in quick succession (van Osch et al., 2015; Walker & Vul, 2014), and without having been replicated since, it is not clear whether the cheerleader effect and group attractiveness effect might actually describing the same phenomenon. Given the similarities between the two effects, one might speculate that observers overestimate the attractiveness of a whole group relative to the average attractiveness of the individual group members (demonstrating the group attractiveness effect), because each individual is already perceived to be more attractive in a group than they are alone (demonstrating the cheerleader effect). On the other hand, the relationship between the two effects might be reversed; perhaps individuals are more attractive in a group than alone (the cheerleader effect), because the group itself is perceived to be highly attractive (the group attractiveness effect) and the observer increases the attractiveness of the individuals to fit with their perception of a highly attractive group. An interesting avenue for future research would be to examine to what extent, if any, the cheerleader and group attractiveness effects are related.

Despite their apparent similarities, it is entirely possible that the group attractiveness and cheerleader effects are unrelated. Currently, the two effects are understood to occur due to two separate mechanisms. Using eye-tracking, van Osch and colleagues (2015) found that observers spent longer gazing at the most attractive faces within each group, and were more likely to remember the most attractive face in the group compared to the least attractive face. van Osch et al. (2015) hypothesised that observers overestimated the average attractiveness of the whole group because they were only attending to the most attractive faces. Notably, while the group attractiveness effect appears to be caused by selective attention (van Osch et al., 2015), a significant cheerleader effect was found when the group was composed of identical images of the same person. This finding suggests that selective attention is unlikely to cause of the cheerleader effect, because groups of identical images do not have a single, highly attractive individual to fixate.

However, before concluding that the two effects are unrelated, the two paradigms must be brought closer in alignment to make justified comparisons between them. For example, the group attractiveness effect was previously measured using a between-participants design, in which one sample of observers rated the attractiveness of the individual faces in the group and a second sample of participants rated attractiveness of the whole group; whereas the cheerleader effect relies on the same observers to make attractiveness ratings for an individual in a group and alone. Furthermore, because the observers made their attractiveness ratings while the faces remained on screen, the group attractiveness effect paradigm does not rely on visual working memory as the cheerleader effect does. Ideally, future research will investigate the relationship between these two effects by testing them within the same experimental procedure. A relatively simple design would be to integrate a measure of the group attractiveness effect into the conventional cheerleader effect design. The within-participants design of the cheerleader effect would be more sensitive to detect small changes to the perceived attractiveness of the same individual in different group conditions. In this proposed design, a cheerleader effect measure would be collected for each face in the group (e.g., Experiment 2, Chapter 6), and observers would also provide an estimate of the average attractiveness of the whole group (see Figure 7.3b). With such a design, it would be possible to identify whether the overestimated attractiveness of the whole group is due to each individual being perceived to be more attractive than they are alone.

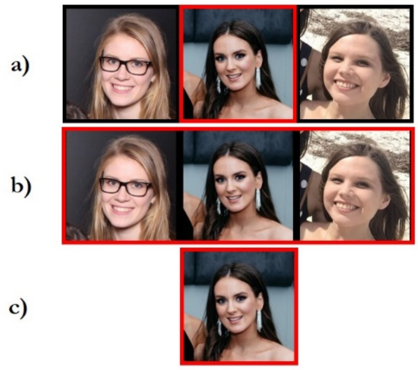

Figure 7.3. An example of the trial conditions in a cheerleader effect experiment that incorporates a measure of the group attractiveness effect a) target face in a group b) group attractiveness condition c) alone condition.

7.6.2 Negative Trait Judgments

Arguably, the most important finding within this thesis concerns the dissociable influence of social context on judgments of attractiveness and trustworthiness (Chapter 5). This finding was surprising because attractiveness and trustworthiness are often highly correlated with one another (Oosterhof & Todorov, 2008); indeed, the attractiveness and trustworthiness ratings given to the same target faces were strongly correlated when they were presented alone (r = .63; Experiment 1, Chapter 5). However, when presented in a group, the attractiveness of all faces increased, whereas an increase in trustworthiness only occurred for the most untrustworthy faces. Crucially, the same stimuli were used in both judgment tasks, suggesting that the dissociation between the judgments was caused by the judgments themselves. These findings raise the interesting possibility that the cheerleader effect might be specific to attractiveness judgments. Furthermore, the dissociable influence of group presence on these two highly related judgments clearly demonstrates that social context cannot be assumed to influence similar judgments in similar ways.

To date, the cheerleader effect has only been investigated among positive trait evaluations. Future research should consider the influence of social context on negative trait judgments, such as aggressiveness (Oosterhof & Todorov, 2008). Griffiths et al. (2018) demonstrated that hierarchical encoding causes observers to recall the emotional expression of an individual to be more alike the group average. A mildly angry face in a group with other angry faces is recalled as being angrier than it really was (Griffiths et al., 2018). Because trait impressions are viewed as temporal extensions of emotional expressions (Knutson, 1996; Secord, 1958), faces that appear angrier in a group will also likely be perceived to be more aggressive or dominant in a group, compared to when they are seen alone. However, due to the nature of hierarchical encoding, Griffiths et al. (2018) also found that faces angrier than the group average were remembered as being less angry in a group. Therefore, it remains to be seen whether appearing in a group of angry faces will cause all faces to be rated as being more aggressive than they are alone, or whether the change to an individual's perceived aggression would depend on the intensity of the other expressions in the group.

7.6.3 Sex Differences

The cheerleader effect is a remarkably general phenomenon that occurs for natural female and male faces (Walker & Vul, 2014), as well as for computer generated faces (Experiment 3, Chapter 5). To date, the cheerleader effect has only been investigated with groups composed of faces that are the same gender. Yet, findings in the mate choice copying literature suggest that it is possible that the sex (and sexuality) of the participant will interact with the sex of the target and distractor faces in the group image, to result in sex differences in the cheerleader effect. Hill and Buss (2008) found that female observers rated a male face to be significantly more desirable when seen in a group with female faces compared to male faces, and to be more desirable in a male group than alone. Conversely, while male observers did not discriminate between female faces that were seen in a group with females or alone, they rated a female to be significantly less desirable in a group with other males. Hill and Buss (2008) hypothesised that males find females in a group with other males to be less desirable because pursuing the female might lead to a physical confrontation with the other males in a display of intra-sex competition.

When applied to the cheerleader effect, Hill and Buss's (2008) findings suggest that the largest cheerleader effects might occur when female observers evaluate a male target face in a group with female distractors. On the other hand, a significant decrease in attractiveness might occur when male observers evaluate a female target face in a group with male distractors. However, Hill and Buss's (2008) findings have not always been replicated. Males have also been shown to engage in mate choice copying (Gouda-Vossos et al., 2018), by increasing their attractiveness ratings for a female after watching her interact with an interested male date (Place et al., 2010), or when she is seen with an attractive male face (Little et al., 2008). Despite the inconsistent reports of mate choice copying in males, these findings demonstrate the need to investigate whether the size of the cheerleader effect is modulated depending on the sex (and sexuality) of the observer and that of the faces in the group.

7.6.4 A Social Attractiveness Halo

Facial attractiveness is a valuable social commodity, as attractive individuals receive better academic grades (Landy & Sigall, 1974), more job offers (Gilmore et al., 1986), higher salaries (Hamermesh & Biddle, 1993) and shorter criminal sentences (Sigall & Ostrove, 1975; Stewart, 1980) than their unattractive counterparts. An interesting avenue for future research will be to examine whether the increase in attractiveness that is typical of the cheerleader effect can also result in more favourable life outcomes for an individual. One clear application of this point is to consider the emerging role of social media and dating applications in modern day romances. On these platforms, users are often able to select photographs of themselves to present to prospective partners. Perhaps it is possible that by selecting photographs of themselves in groups, individuals could use the cheerleader effect to their own advantage. Attractive individuals report more dating experience than unattractive individuals (Langlois et al., 2000), and both males and females use facial attractiveness to guide mate choices (Buss & Barnes, 1986; Rhodes et al., 2005). If the same individual is perceived to be more attractive in a group, would they experience more dating success by selecting and displaying images of themselves in groups to potential dates? Future research would be well directed to investigate whether the small increase in attractiveness that is typical of the cheerleader effect can result in the favourable impressions that are typical of the attractiveness halo (Dion et al., 1972).

7.7 Concluding Remarks

In 1963, Levy and Richter cautioned that models of social perception needed to be expanded to describe the effect that social context has on the trait judgments that are made from the face. However, this early warning went largely unheeded; the influence of social context on trait impressions has continued to be overlooked or ignored in the social perception literature (Barrett et al., 2011; Phillips et al., 2014). Despite being related to the physical qualities of the face, my findings have demonstrated that facial attractiveness is reliably modulated by social context. Natural faces are perceived to be more attractive in a group than alone, regardless of their own attractiveness, or the attractiveness of the group in which they appear. My research has shown the cheerleader effect to be a remarkably robust phenomenon; however, these findings have also raised many questions about the underlying cause of the effect. My findings are inconsistent with a cheerleader effect that is the result of a single hierarchical encoding mechanism. Instead, I have proposed that the cheerleader effect is consistent with observers attributing positive characteristics to individuals that are seen in socially desirable groups.

Crucially, my research raises further questions about the influence of social context on traits other than attractiveness. Despite becoming more attractive, the same faces did not necessarily become more trustworthy in a group, even though attractiveness and trustworthiness are highly similar judgments. The dissociable influence of social context on such closely related judgments highlights the need for the field of social perception to consider the role of social context on all trait judgments that are made from the face. In conclusion, my research has not only demonstrated that social context reliably modulates perceptions of facial attractiveness, but also raises the possibility that social context has idiosyncratic influences on each of the trait impressions that we make from the face at first glance.

References

Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion, 30*(2), 177-186. doi: 10.1007/s11031-006-9020-2

Adams, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion, 5*(1), 3-11. doi: 10.1037/1528-3542.5.1.3

Alley, T. R., & Cunningham, M. R. (1991). Article Commentary: Averaged faces are attractive, but very attractive faces are not average. *Psychological Science, 2*(2), 123-125. doi: 10.1111/j.1467-9280.1991.tb00113.x

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122-131. doi: 10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392-398. doi: 10.1111/j.1467-9280.2008.02098.x

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, 106*(18), 7345-7350. doi: 10.1073/pnas.0808981106

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology, 70*(4), 394-400. doi: 10.1037/h0022280

Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology, 29*(1), 73-93. doi: 10.1016/j.joep.2007.07.008

Andrzejewski, S. A., Hall, J. A., & Salib, E. R. (2009). Anti-Semitism and identification of Jewish group membership from photographs. *Journal of Nonverbal Behavior, 33*(1), 47-58. doi: 10.1007/s10919-008-0060-z

Aniulis, E., Churches, O., Thomas, N. A., & Nicholls, M. E. (2016). Representational pseudoneglect for detecting changes to Rey–Osterrieth figures. *Experimental Brain Research, 234*(11), 3381-3387. doi: 10.1007/s00221-016-4735-0

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323*(5918), 1183-1183. doi: 10.1126/science.1167748

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162. doi: 10.1111/1467-9280.00327

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559. doi: 10.1126/science.1736359

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*(46), 17948-17953. doi: 10.1073/pnas.0705435104

Bar-Tal, D., & Saxe, L. (1976). Perceptions of similarly and dissimilarly attractive couples and individuals. *Journal of Personality and Social Psychology, 33*(6), 772-781. doi: 10.1037/0022-3514.33.6.772

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278. doi: 10.1037/1528-3542.6.2.269

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286-290. doi: 10.1177/0963721411422522

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17*(3), 613-620. doi: 10.1016/s0926-6410(03)00174-5

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology, 31*(1), 64-79. doi: 10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142. doi: 10.1006/game.1995.1027

Bond, J., Charles F, Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15*(4), 523-534. doi: 10.1207/s15324834basp1504_8

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., . . . Whalen, J. R. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology, 12*(3), 446-458. doi: 10.1037//0894-4105.12.3.446

Borod, J. C., Kent, J., Koff, E., Martin, C., & Alpert, M. (1988). Facial asymmetry while posing positive and negative emotions: Support for the right hemisphere hypothesis. *Neuropsychologia, 26*(5), 759-764. doi: 10.1016/0028-3932(88)90013-9

Bourne, V. J. (2005). Lateralised processing of positive facial emotion: Sex differences in strength of hemispheric dominance. *Neuropsychologia, 43*(6), 953-956. doi: 10.1016/j.neuropsychologia.2004.08.007

Bourne, V. J. (2010). How are emotions lateralised in the brain? Contrasting existing hypotheses using the chimeric faces test. *Cognition & Emotion, 24*(5), 903-911. doi: 10.1080/02699930903007714

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia, 18*(4), 491-498. doi: 10.1016/0028-3932(80)90151-7

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384-392. doi: 10.1177/0956797610397956

Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision, 15*(15), 6-6. doi: 10.1167/15.15.6

Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 921-929. doi: 10.1037/xlm0000075

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 4-4. doi: 10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325-14329. doi: 10.1073/pnas.0803390105

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review, 120*(1), 85-109. doi: 10.1037/a0030779

Brooks, J. L., Della Sala, S., & Darling, S. (2014). Representational pseudoneglect: A review. *Neuropsychology Review, 24*(2), 148-165. doi: 10.1007/s11065-013-9245-2

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization, 68*(3-4), 466-476. doi: 10.1016/j.jebo.2007.10.006

Bulakowski, P. F., Post, R. B., & Whitney, D. (2011). Re-examining the possible benefits of visual crowding: Dissociating crowding from ensemble percepts. *Attention, Perception, & Psychophysics, 73*(4), 1003-1009. doi: 10.3758/s13414-010-0086-2

Burt, D. M., & Perrett, D. I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia, 35*(5), 685-693. doi: 10.1016/s0028-3932(96)00111-x

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. Calder, G. Rhodes, M. Johnson & J. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 287-306). Oxford, UK: Oxford University Press.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*(3), 256-284. doi: 10.1016/j.cogpsych.2005.06.003

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*(1), 1-14. doi: 10.1017/s0140525x00023992

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50*(3), 559-570. doi: 10.1037//0022-3514.50.3.559

Butler, S., Gilchrist, I., Burt, D., Perrett, D., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia, 43*(1), 52-59. doi: 10.1016/j.neuropsychologia.2004.06.005

Carragher, D. J., Lawrence, B. J., Thomas, N. A., & Nicholls, M. E. R. (2018). Visuospatial asymmetries do not modulate the cheerleader effect. *Scientific Reports, 8*(1), 2548. doi: 10.1038/s41598-018-20784-5

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. (2017). Is trustworthiness lateralized in the face? Evidence from a trust game. *Laterality: Asymmetries of Body, Brain and Cognition, 23*(1), 20-38. doi: 10.1080/1357650X.2017.1298120

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1651), 2651-2656. doi: 10.1098/rspb.2008.0873

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*(10), 1194-1198. doi: 10.1111/j.1467-9280.2009.02423.x

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*(2), 87-105. doi: 10.1016/j.cogpsych.2010.03.001

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology, 48*(4), 879-894. doi: 10.1080/14640749508401421

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393-404. doi: 10.1093/acprof:osobl/9780199734337.003.0031

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception, & Psychophysics, 67*(1), 1-13. doi: 10.3758/bf03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891-900. doi: 10.1016/j.visres.2004.10.004

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. *PLoS ONE, 9*(7), e99629. doi: 10.1371/journal.pone.0099629

Cooremans, K., & Geuens, M. (2015). *The group effect: how food products look more attractive in group than individually.* Paper presented at the Society for Consumer Psychology Annual Winter Conference.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163-228). Oxford & New York: Oxford University Press.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42-45. doi: 10.20982/tqmp.01.1.p042

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*(4), 16-16. doi: 10.1167/10.4.16

Cunningham, M. R., Barbee, A. P., & Pike, C. L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59*(1), 61. doi: 10.1037//0022-3514.59.1.61

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68*(2), 261-279. doi: 10.1037//0022-3514.68.2.261

Dakin, S., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181-3192. doi: 10.1016/s0042-6989(97)00133-8

Darwin, C. (1973). *The autobiography of Charles Darwin, 1809-1882.* New York, NY: WW Norton & Company.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27*(3), 403-424. doi: 10.1207/s15326942dn2703_6

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology, 62*(9), 1716-1722. doi: 10.1080/17470210902811249

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70*(5), 789-794. doi: 10.3758/pp.70.5.789

DeBruine, L. M., Jones, B. C., Unger, L., Little, A. C., & Feinberg, D. R. (2007). Dissociating averageness and attractiveness: Attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1420-1430. doi: 10.1037/0096-1523.33.6.1420

Dion, K. (1972). Physical attractiveness and evaluation of children's transgressions. *Journal of Personality and Social Psychology, 24*(2), 207-213. doi: 10.1037/h0033372

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290. doi: 10.1037/h0033731

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1*(1), 0001. doi: 10.1038/s41562-016-0001

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199. doi: 10.1016/j.econlet.2007.06.027

Dunn, M. J., & Hill, A. (2014). Manipulated luxury-apartment ownership enhances opposite-sex attraction in females but not males. *Journal of Evolutionary Psychology, 12*(1), 1-17. doi: 10.1556/jep.12.2014.1.1

Dunn, M. J., & Searle, R. (2010). Effect of manipulated prestige-car ownership on both sex attractiveness ratings. *British Journal of Psychology, 101*(1), 69-80. doi: 10.1348/000712609x417319

Dunstan, C. J., & Lindell, A. K. (2012). Hemifacial preferences for the perception of emotion and attractiveness differ with the gender of the one beheld. *Cognition & Emotion, 26*(5), 907-915. doi: 10.1080/02699931.2011.621931

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*(1), 109-128. doi: 10.1037//0033-2909.110.1.109

Eckel, C. C., & Petrie, R. (2011). Face value. *The American Economic Review, 101*(4), 1497-1513. doi: 10.1257/aer.101.4.1497

Ekman, P. (1971). *Universals and cultural differences in facial expressions of emotion.* Paper presented at the Nebraska Symposium on Motivation, Nebraska.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129. doi: 10.1037/h0030377

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6*(4), 238-252. doi: 10.1007/bf00987191

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712-717. doi: 10.1037//0022-3514.53.4.712

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203-235. doi: 10.1037/0033-2909.128.2.203

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology, 85*(2), 276-290. doi: 10.1037/0022-3514.85.2.276

Elias, E., Dyer, M., & Sweeny, T. D. (2016). Ensemble perception of dynamic emotional groups. *Psychological Science, 28*(2), 193-203. doi: 10.1177/0956797616678188

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19*(9), 1508-1519. doi: 10.1162/jocn.2007.19.9.1508

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception, 39*(7), 931-941. doi: 10.1068/p6633

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science, 18*(2), 327-334. doi: 10.1111/desc.12218

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628-634. doi: 10.1037//0096-1523.21.3.628

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/bf03193146

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59*(5), 981-993. doi: 10.1037//0022-3514.59.5.981

Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*(5), 154-158. doi: 10.1111/1467-8721.00190

Fink, J. S. (2015). Female athletes, women's sport, and the sport media commercial complex: Have we really “come a long way, baby”? *Sport Management Review, 18*(3), 331-342. doi: 10.1016/j.smr.2014.05.001

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106*(3), 1389-1398. doi: 10.1152/jn.00904.2010

Fletcher, G. J., Tither, J. M., O’Loughlin, C., Friesen, M., & Overall, N. (2004). Warm and homely or cold and beautiful? Sex differences in trading off traits in mate selection. *Personality & Social Psychology Bulletin, 30*(6), 659-672. doi: 10.1177/0146167203262847

Franklin, R. G., & Adams, R. B. (2010). The two sides of beauty: Laterality and the duality of facial attractiveness. *Brain and Cognition, 72*(2), 300-305. doi: 10.1016/j.bandc.2009.10.002

Furl, N. (2016). Facial-attractiveness choices are predicted by divisive normalization. *Psychological Science, 27*(10), 1379-1387. doi: 10.1177/0956797616661523

Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology, 15*(2), 73-85. doi: 10.1016/0162-3095(94)90018-3

Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology, 20*(5), 409-424. doi: 10.1016/0022-1031(84)90035-0

Gerger, G., Forster, M., & Leder, H. (2017). It felt fluent but I did not like it: Fluency effects in faces versus patterns. *The Quarterly Journal of Experimental Psychology, 70*(4), 637-648. doi: 10.1080/17470218.2016.1145705

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201-210. doi: 10.1016/j.cognition.2010.11.002

Gilmore, D. C., Beehr, T. A., & Love, K. G. (1986). Effects of applicant sex, applicant physical attractiveness, type of rater and type of job on interview decisions. *Journal of Occupational Psychology, 59*(2), 103-109. doi: 10.1111/j.2044-8325.1986.tb00217.x

Gouda-Vossos, A., Nakagawa, S., Dixson, B. J., & Brooks, R. C. (2018). Mate choice copying in humans: A systematic review and meta-analysis. *Adaptive Human Behavior and Physiology*, 1-23. doi: 10.1007/s40750-018-0099-y

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108*(3), 233-242. doi: 10.1037//0735-7036.108.3.233

Griffiths, S., Rhodes, G., Jeffery, L., Palermo, R., & Neumann, M. F. (2018). The average facial expression of a crowd influences impressions of individual expressions. *Journal of Experimental Psychology: Human Perception and Performance, 44*(2), 311-319. doi: 10.1037/xhp0000446

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology, 23*(1), 49-66. doi: 10.1016/s0167-4870(01)00067-8

Guo, K., Meints, K., Hall, C., Hall, S., & Mills, D. (2009). Left gaze bias in humans, rhesus monkeys and domestic dogs. *Animal Cognition, 12*(3), 409-418. doi: 10.1007/s10071-008-0199-3

Guo, K., Smith, C., Powell, K., & Nicholls, K. (2012). Consistent left gaze bias in processing different facial cues. *Psychological Research, 76*(3), 263-269. doi: 10.1007/s00426-011-0340-9

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015a). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General, 144*(2), 432-446. doi: 10.1037/xge0000053

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1-1. doi: 10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015b). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15*(4), 16-16. doi: 10.1167/15.4.16

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751-R753. doi: 10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718-734. doi: 10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72*(7), 1825-1838. doi: 10.3758/app.72.7.1825

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review, 18*(5), 855-859. doi: 10.3758/s13423-011-0125-6

Hamermesh, D. S., & Biddle, J. E. (1993). Beauty and the labor market. *84*(5), 1174-1194. doi: 10.3386/w4518

Harp, T., Haberman, J., & Whitney, D. (2009). Grouping oranges affects their overall appeal. *Journal of Vision, 9*(8), 925-925. doi: 10.1167/9.8.925

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837-852. doi: 10.1037//0022-3514.78.5.837

Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception. In A. Calder, G. Rhodes, M. Johnson & J. Haxby (Eds.), *The Oxford Handbook of Face Perception*. Oxford, UK: Oxford University Press.

Hennighausen, C., Hudders, L., Lange, B. P., & Fink, H. (2016). What if the rival drives a Porsche? Luxury car spending as a costly signal in male intrasexual competition. *Evolutionary Psychology, 14*(4), 1474704916678217. doi: 10.1177/1474704916678217

Hill, S. E., & Buss, D. M. (2008). The mere presence of opposite-sex others on judgments of sexual and romantic desirability: Opposite effects for men and women. *Personality and Social Psychology Bulletin, 34*(5), 635-647. doi: 10.1177/0146167207313728

Hudders, L., De Backer, C., Fisher, M., & Vyncke, P. (2014). The rival wears Prada: Luxury consumption as a female competition strategy. *Evolutionary Psychology, 12*(3), 147470491401200306. doi: 10.1177/147470491401200306

Indersmitten, T., & Gur, R. C. (2003). Emotion processing in chimeric faces: Hemispheric asymmetries in expression and recognition of emotions. *The Journal of Neuroscience, 23*(9), 3820-3825. doi: 10.1523/jneurosci.23-09-03820.2003

Jeffries, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1571), 1671-1683. doi: 10.1098/rstb.2010.0379

Jenkins, R., White, D., Van Montfort, X., & Burton, M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323. doi: 10.1016/j.cognition.2011.08.001

Jewell, G., & McCourt, M. E. (2000). Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia, 38*(1), 93-110. doi: 10.1016/s0028-3932(99)00045-7

Jones, B. C., DeBruine, L. M., Little, A. C., Burriss, R. P., & Feinberg, D. R. (2007). Social transmission of face preferences among humans. *Proceedings of the Royal Society of London B: Biological Sciences, 274*(1611), 899-903. doi: 10.1098/rspb.2006.0205

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences, 285*(1879), 20172770. doi: 10.1098/rspb.2017.2770

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311. doi: 10.1523/jneurosci.17-11-04302.1997

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 361*(1476), 2109-2128. doi: 10.1098/rstb.2006.1934

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology, 38*(1), 131-140. doi: 10.1037//0022-3514.38.1.131

Kernis, M. H., & Wheeler, L. (1981). Beautiful friends and ugly strangers: Radiation and contrast effects in perceptions of same-sex pairs. *Personality and Social Psychology Bulletin, 7*(4), 617-620. doi: 10.1177/014616728174017

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica, 33*, 193-201. doi: 10.1016/0001-6918(70)90132-0

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*(3), 165-182. doi: 10.1007/bf02281954

Kovacs-Balint, Z., Bereczkei, T., & Hernadi, I. (2013). The telltale face: possible mechanisms behind defector and cooperator recognition revealed by emotional facial expression metrics. *British Journal of Psychology, 104*(4), 563-576. doi: 10.1111/bjop.12007

Kramer, R. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision, 15*(4), 1-1. doi: 10.1167/15.4.1

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*(4), 730-735. doi: 10.1037/1528-3542.7.4.730

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron, 20*(5), 937-945. doi: 10.1016/s0896-6273(00)80475-4

Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*(3), 299-304. doi: 10.1037/h0036018

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390-423. doi: 10.1037//0033-2909.126.3.390

Langlois, J. H., Ritter, J. M., Roggman, L. A., & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology, 27*(1), 79-84. doi: 10.1037//0012-1649.27.1.79

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115-121. doi: 10.1111/j.1467-9280.1990.tb00079.x

Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and what is not average about attractive faces? *Psychological Science, 5*(4), 214-220. doi: 10.1111/j.1467-9280.1994.tb00503.x

LaRocque, J. J., Eichenbaum, A. S., Starrett, M. J., Rose, N. S., Emrich, S. M., & Postle, B. R. (2015). The short-and long-term fates of memory items retained outside the focus of attention. *Memory & Cognition, 43*(3), 453-468. doi: 10.3758/s13421-014-0486-y

Lefevre, C. E., Lewis, G. J., Bates, T. C., Dzhelyova, M., Coetzee, V., Deary, I. J., & Perrett, D. I. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior, 33*(6), 623-627. doi: 10.1016/j.evolhumbehav.2012.03.002

Leonards, U., & Scott-Samuel, N. E. (2005). Idiosyncratic initiation of saccadic face exploration in humans. *Vision Research, 45*(20), 2677-2684. doi: 10.1016/j.visres.2005.03.009

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635-654. doi: 10.1016/j.visres.2007.12.009

Levy, B., Ariely, D., Mazar, N., Chi, W., Lukas, S., & Elman, I. (2008). Gender differences in the motivational processing of facial beauty. *Learning and Motivation, 39*(2), 136-145. doi: 10.1016/j.lmot.2007.09.002

Levy, L. H., & Richter, M. L. (1963). Impressions of groups as a function of the stimulus values of their individual members. *The Journal of Abnormal and Social Psychology, 67*(4), 349-354. doi: 10.1037/h0040100

Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision, 15*(4), 10-10. doi: 10.1167/15.4.10

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in Psychology, 7*, 1332. doi: 10.3389/fpsyg.2016.01332

Lie, H. C., Rhodes, G., & Simmons, L. W. (2008). Genetic diversity revealed in human faces. *Evolution, 62*(10), 2473-2486. doi: 10.1111/j.1558-5646.2008.00478.x

Lindell, A. K. (2013a). Continuities in emotion lateralization in human and non-human primates. *Front Hum Neurosci, 7*, 464. doi: 10.3389/fnhum.2013.00464

Lindell, A. K. (2013b). The silent social/emotional signals in left and right cheek poses: A literature review. *Laterality: Asymmetries of Body, Brain and Cognition, 18*(5), 612-624. doi: 10.1080/1357650X.2012.737330

Little, A. C., Burriss, R. P., Jones, B. C., DeBruine, L. M., & Caldwell, C. A. (2008). Social influence in human face preference: Men and women are influenced more for long-term than short-term attractiveness decisions. *Evolution and Human Behavior, 29*(2), 140-146. doi: 10.1016/j.evolhumbehav.2007.11.007

Little, A. C., Caldwell, C. A., Jones, B. C., & DeBruine, L. M. (2011a). Effects of partner beauty on opposite-sex attractiveness judgments. *Archives of Sexual Behavior, 40*(6), 1119-1127. doi: 10.1007/s10508-011-9806-5

Little, A. C., & Hancock, P. J. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology, 93*(4), 451-464. doi: 10.1348/000712602761381349

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011b). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1571), 1638-1659. doi: 10.1098/rstb.2010.0404

Little, A. C., Jones, B. C., DeBruine, L. M., & Caldwell, C. A. (2011c). Social learning and human mate preferences: A potential mechanism for generating and maintaining between-population diversity in attraction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1563), 366-375. doi: 10.1098/rstb.2010.0192

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391-400. doi: 10.1016/j.tics.2013.06.006

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, 91-630.

Luo, A. X., & Zhou, G. (2018). Ensemble perception of facial attractiveness. *Journal of Vision, 18*(8), 7-7. doi: 10.1167/18.8.7

Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., & Neuberg, S. L. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology, 85*(6), 1107-1120. doi: 10.1037/0022-3514.85.6.1107

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion, 5*(1), 119-124. doi: 10.1037/1528-3542.5.1.119

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America A: Optics and Image Science, 31*(4), A93-A102. doi: 10.1364/josaa.31.000a93

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215-238. doi: 10.1037//0033-295x.90.3.215

McGeorge, P., Beschin, N., Colnaghi, A., Rusconi, M. L., & Della Sala, S. (2007). A lateralized bias in mental imagery: Evidence for representational pseudoneglect. *Neuroscience Letters, 421*(3), 259-263. doi: 10.1016/j.neulet.2007.05.050

McManus, I. C., & Humphrey, N. (1973). Turning the left cheek. *Nature, 243*(5405), 271-272. doi: 10.1038/243271a0

Mealey, L., Bridgstock, R., & Townsend, G. C. (1999). Symmetry and perceived facial attractiveness: A monozygotic co-twin comparison. *Journal of Personality and Social Psychology, 76*(1), 151-158. doi: 10.1037//0022-3514.76.1.151

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35. doi: 10.1037//1076-8971.7.1.3

Melamed, L., & Moss, M. K. (1975). The effect of context on ratings of attractiveness of photographs. *The Journal of Psychology, 90*(1), 129-136. doi: 10.1080/00223980.1975.9923935

Møller, A. P., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: A meta-analysis. *The American Naturalist, 151*(2), 174-192. doi: 10.1086/286110

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*(4), 237-254. doi: 10.1023/a:1027332800296

Nelissen, R. M., & Meijers, M. H. (2011). Social benefits of luxury brands as costly signals of wealth and status. *Evolution and Human Behavior, 32*(5), 343-355. doi: 10.1016/j.evolhumbehav.2010.12.002

Neumann, M. F., Ng, R., Rhodes, G., & Palermo, R. (2017). Ensemble coding of face identity is not independent of the coding of individual identity. *The Quarterly Journal of Experimental Psychology, 71*(6), 1357-1366. doi: 10.1080/17470218.2017.1318409

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128*(1), 56-63. doi: 10.1016/j.cognition.2013.03.006

Nicholls, M., Clode, D., Lindell, A. K., & Wood, A. G. (2001). Which cheek to turn? The effect of gender and emotional expressivity on posing behavior. *Brain and Cognition, 48*(2-3), 480-484.

Nicholls, M. E., Bradshaw, J. L., & Mattingley, J. B. (1999a). Free-viewing perceptual asymmetries for the judgement of brightness, numerosity and size. *Neuropsychologia, 37*(3), 307-314. doi: 10.1016/s0028-3932(98)00074-8

Nicholls, M. E., Clode, D., Wood, S. J., & Wood, A. G. (1999b). Laterality of expression in portraiture: Putting your best cheek forward. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1428), 1517-1522. doi: 10.1098/rspb.1999.0809

Nicholls, M. E., Orr, C. A., Okubo, M., & Loftus, A. (2006). Satisfaction guaranteed: The effect of spatial biases on responses to likert scales. *Psychological Science, 17*(12), 1027-1028. doi: 10.1111/j.1467-9280.2006.01822.x

Nicholls, M. E., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The Flinders handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex, 49*(10), 2914-2926. doi: 10.1016/j.cortex.2013.02.002

Nicholls, M. E., Wolfgang, B. J., Clode, D., & Lindell, A. K. (2002). The effect of left and right poses on the expression of facial emotion. *Neuropsychologia, 40*(10), 1662-1665. doi: 10.1016/s0028-3932(02)00024-6

Nicholls, M. E. R., Ellis, B. E., Clement, J. G., & Yoshino, M. (2004). Detecting hemifacial asymmetries in emotional expression with three-dimensional computerized image analysis. *Proceedings of the Royal Society B: Biological Sciences, 271*(1540), 663-668. doi: 10.1098/rspb.2003.2660

Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7*(1-3), 93-106. doi: 10.1080/135062800394702

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition, 26*(1), 146-160. doi: 10.3758/bf03211378

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology, 30*(10), 953-958. doi: 10.1016/j.psyneuen.2005.03.019

Ojiro, Y., Gobara, A., Nam, G., Sasaki, K., Kishimoto, R., Yamada, Y., & Miura, K. (2015). Two replications of "Hierarchical encoding makes individuals in a group seem more attractive (2014; Experiment 4)". *The Quantitative Methods for Psychology, 11*(2), r8-r11. doi: 10.20982/tqmp.11.2.r008

Okubo, M., Ishikawa, K., & Kobayashi, A. (2013). No trust on the left side: Hemifacial asymmetries for trustworthiness and emotional expressions. *Brain and Cognition, 82*(2), 181-186. doi: 10.1016/j.bandc.2013.04.004

Okubo, M., Ishikawa, K., Kobayashi, A., & Suzuki, H. (2016). Can I trust you? Laterality of facial trustworthiness in an economic game. *Journal of Nonverbal Behavior, 41*(1), 1-14. doi: 10.1007/s10919-016-0242-z

Okubo, M., Kobayashi, A., & Ishikawa, K. (2012). A fake smile thwarts cheater detection. *Journal of Nonverbal Behavior, 36*(3), 217-225. doi: 10.1007/s10919-012-0134-9

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*(11), 566-570. doi: 10.1016/j.tics.2014.09.007

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83-110. doi: 10.1007/s10919-009-0082-1

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Re-examining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*(2), 315-324. doi: 10.1016/j.jesp.2009.12.002

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion, 5*(4), 498-502. doi: 10.1037/1528-3542.5.4.498

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105*(32), 11087-11092. doi: 10.1073/pnas.0805664105

Oppenheimer, D. M., LeBoeuf, R. A., & Brewer, N. T. (2008). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition, 106*(1), 13-26. doi: 10.1016/j.cognition.2006.12.008

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research, 79*, 8-16. doi: 10.1016/j.visres.2012.12.004

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia, 49*(5), 1226-1235. doi: 10.1016/j.neuropsychologia.2011.02.021

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739-744. doi: 10.1038/89532

Pegors, T. K., Mattar, M. G., Bryan, P. B., & Epstein, R. A. (2015). Simultaneous perceptual and response biases on sequential face attractiveness judgments. *Journal of Experimental Psychology: General, 144*(3), 664-673. doi: 10.1037/xge0000069

Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology, 26*(2), 103-133. doi: 10.1006/cogp.1994.1004

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18(4), 445-451. doi: 10.1016/j.conb.2008.09.008

Penton-Voak, I., Jones, B. C., Little, A., Baker, S., Tiddeman, B., Burt, D., & Perrett, D. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London B: Biological Sciences, 268*(1476), 1617-1623. doi: 10.1098/rspb.2001.1703

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior, 25*(4), 229-241. doi: 10.1016/j.evolhumbehav.2004.04.003

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 607-640. doi: 10.1521/soco.2006.24.5.607

Perrett, D., Lee, K., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D., . . . Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*(6696), 884-887. doi: 10.1038/29772

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20*(5), 295-307. doi: 10.1016/s1090-5138(99)00014-8

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature, 368*(6468), 239-242. doi: 10.1038/368239a0

Peskin, M., & Newell, F. N. (2004). Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces. *Perception, 33*(2), 147-158. doi: 10.1068/p5028

Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology, 114*(5), 766-785. doi: 10.1037/pspi0000120

Phillips, L. T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior, 34*, 101-127. doi: 10.1016/j.riob.2014.10.001

Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2010). Humans show mate copying after observing real mate choices. *Evolution and Human Behavior, 31*(5), 320-325. doi: 10.1016/j.evolhumbehav.2010.02.001

Porter, S., England, L., Juodis, M., ten Brinke, L., & Wilson, K. (2008). Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 40*(3), 171-177. doi: 10.1037/0008-400x.40.3.171

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*(6), 477-491. doi: 10.1080/10683160902926141

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience, 17*(7), 1043-1057. doi: 10.1162/0898929054475181

Powell, W. R., & Schirillo, J. A. (2009). Asymmetrical facial expressions in portraits and hemispheric laterality: A literature review. *Laterality: Asymmetries of Body, Brain and Cognition, 14*(6), 545-572. doi: 10.1080/13576500802680336

Praino, R., Stockemer, D., & Ratis, J. (2014). Looking good or looking competent? Physical appearance and electoral success in the 2008 congressional elections. *American Politics Research, 42*(6), 1096-1117. doi: 10.1177/1532673x14532825

Price, P. C. (2001). A group size effect on personal risk judgments: Implications for unrealistic optimism. *Memory & Cognition, 29*(4), 578-586. doi: 10.3758/bf03200459

Price, P. C., Kimura, N. M., Smith, A. R., & Marshall, L. D. (2014). Sample size bias in judgments of perceptual averages. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1321-1331. doi: 10.1037/a0036576

Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality and Social Psychology, 90*(3), 382-398. doi: 10.1037/0022-3514.90.3.382

Proverbio, A. M., Brignone, V., Matarazzo, S., Del Zotto, M., & Zani, A. (2006). Gender differences in hemispheric asymmetry for face processing. *BMC Neuroscience, 7*(1), 44. doi: 10.1186/1471-2202-7-44

Rashid, R. (Writer) & P. Fryman (Director). (2008). Not a father's day [Television series episode]. In C. Bays & C. Thomas (Producer), *How I met your mother*. New York, NY: CBS.

Rashidi, M., Pazhoohi, F., & Hosseinchari, M. (2012). Effect of facial stimuli exposure time on evaluation of facial attractiveness. *Australian Journal of Psychology, 64*(3), 164-168. doi: 10.1111/j.0004-9530.2011.00050.x

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037//0033-2909.124.3.372

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE, 7*(3), e34293. doi: 10.1371/journal.pone.0034293

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*(1), 199-226. doi: 10.1146/annurev.psych.57.102904.190208

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London B: Biological Sciences, 270*(Suppl 1), S93-S95. doi: 10.1098/rsbl.2003.0023

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology, 91*(1), 125-140. doi: 10.1348/000712600161718

Rhodes, G., Neumann, M., Ewing, L., Bank, S., Read, A., Engfors, L. M., . . . Palermo, R. (2017). Ensemble coding of faces occurs in children and develops dissociably from coding of individual faces. *Developmental Science, 21*(2), e12540. doi: 10.1111/desc.12540

Rhodes, G., Neumann, M. F., Ewing, L., & Palermo, R. (2015). Reduced set averaging of face identity in children and adolescents with autism. *The Quarterly Journal of Experimental Psychology, 68*(7), 1391-1403. doi: 10.1080/17470218.2014.981554

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review, 5*(4), 659-669. doi: 10.3758/bf03208842

Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior, 26*(2), 186-201. doi: 10.1016/j.evolhumbehav.2004.08.014

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science, 10*(1), 52-58. doi: 10.1111/1467-9280.00106

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science, 7*(2), 105-110. doi: 10.1111/j.1467-9280.1996.tb00338.x

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception, 30*(5), 611-625. doi: 10.1068/p3123

Ritchie, K. L., Palermo, R., & Rhodes, G. (2017). Forming impressions of facial attractiveness is mandatory. *Scientific Reports, 7*(1), 469. doi: 10.1038/s41598-017-00526-9

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11*(12), 18-18. doi: 10.1167/11.12.18

Robson, M. K., Palermo, R., Jeffery, L., & Neumann, M. F. (2018). Ensemble coding of face identity is present but weaker in congenital prosopagnosia. *Neuropsychologia, 111*, 377-386. doi: 10.1016/j.neuropsychologia.2018.02.019

Rodeheffer, C. D., Proffitt Leyva, R. P., & Hill, S. E. (2016). Attractive female romantic partners provide a proxy for unobservable male qualities: The when and why behind human female mate choice copying. *Evolutionary Psychology, 14*(2), 1474704916652144. doi: 10.1177/1474704916652144

Rossion, B., Dricot, L., Devolder, A., Bodart, J.-M., Crommelinck, M., De Gelder, B., & Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 12*(5), 793-802. doi: 10.1162/089892900562606

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage, 20*(3), 1609-1624. doi: 10.1016/j.neuroimage.2003.07.010

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356-374. doi: 10.1016/j.jmp.2012.08.001

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology, 44*(4), 1100-1105. doi: 10.1016/j.jesp.2007.12.001

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS ONE, 5*(1), e8733. doi: 10.1371/journal.pone.0008733

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology, 45*(6), 1245-1251. doi: 10.1016/j.jesp.2009.07.010

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*(3), 409-426. doi: 10.1037/a0031050

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102-141. doi: 10.1037//0033-2909.115.1.102

Sackeim, H. A., Gur, R. C., & Saucy, M. C. (1978). Emotions are expressed more intensely on the left side of the face. *Science, 202*(4366), 434-436. doi: 10.1126/science.705335

Saegusa, C., & Watanabe, K. (2016). Judgments of facial attractiveness as a combination of facial parts information over time: Social and aesthetic factors. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 173. doi: 10.1037/xhp0000149

Said, C. P., Baron, S. G., & Todorov, A. (2009a). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience, 21*(3), 519-528. doi: 10.1162/jocn.2009.21041

Said, C. P., Sebe, N., & Todorov, A. (2009b). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260-264. doi: 10.1037/a0014681

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science, 22*(9), 1183-1190. doi: 10.1177/0956797611419169

Scharlemann, J. P., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*(5), 617-640. doi: 10.1016/s0167-4870(01)00059-9

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1431), 1913-1917. doi: 10.1098/rspb.1999.0866

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person Perception and Interpersonal Behavior* (pp. 300-315). Stanford, California, USA: Stanford University Press.

Sigall, H., & Landy, D. (1973). Radiating beauty: Effects of having a physically attractive partner on person perception. *Journal of Personality and Social Psychology, 28*(2), 218-224. doi: 10.1037/h0035740

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology, 31*(3), 410-414. doi: 10.1037/h0076472

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267. doi: 10.1016/s1364-6613(97)01080-2

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*(1), 16-20. doi: 10.1016/j.tics.2004.11.006

Simpson, A., & Thomas, N. A. (2018). Neuroticism, schizotypy, and scale anchors influence eye movement behaviour in the visual exploration of abstract art: An exploratory study. *Acta Psychologica, 183*, 85-98. doi: 10.1016/j.actpsy.2017.12.008

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development, 21*(2), 345-354. doi: 10.1016/s0163-6383(98)90011-x

Slepian, M. L., & Ames, D. R. (2015). Internalized impressions the link between apparent facial trustworthiness and deceptive behavior is mediated by targets' expectations of how they will be judged. *Psychological Science*, 282-288. doi: 10.1177/0956797615594897

Smith, A. R., & Price, P. C. (2010). Sample size bias in the estimation of means. *Psychonomic Bulletin & Review, 17*(4), 499-503. doi: 10.3758/pbr.17.4.499

Smith, A. R., Rule, S., & Price, P. C. (2017). Sample size bias in retrospective estimates of average duration. *Acta Psychologica, 176*, 39-46. doi: 10.1016/j.actpsy.2017.03.008

Smith, M. L., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., . . . Hillier, S. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society of London B: Biological Sciences, 273*(1583), 135-140. doi: 10.1098/rspb.2005.3296

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14), 19-19. doi: 10.1167/10.14.19

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7710-7715. doi: 10.1073/pnas.1014345108

Stewart, J. E. (1980). Defendant's attractiveness as a factor in the outcome of criminal trials: An observational study1. *Journal of Applied Social Psychology, 10*(4), 348-361. doi: 10.1111/j.1559-1816.1980.tb00715.x

Stins, J. F., Roelofs, K., Villan, J., Kooijman, K., Hagenaars, M. A., & Beek, P. J. (2011). Walk to me when I smile, step back when I'm angry: Emotional faces modulate whole-body approach–avoidance behaviors. *Experimental Brain Research, 212*(4), 603-611. doi: 10.1007/s00221-011-2767-z

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349-354. doi: 10.1177/0956797610362647

Stockemer, D., & Praino, R. (2015). Blinded by beauty? Physical attractiveness and candidate selection in the US House of Representatives. *Social Science Quarterly, 96*(2), 430-443. doi: 10.1111/ssqu.12155

Störmer, V. S., & Alvarez, G. A. (2016). Attention alters perceived attractiveness. *Psychological Science, 27*(4), 563-571. doi: 10.1177/0956797616630964

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology, 23*(3), 423-429. doi: 10.1016/j.conb.2012.12.011

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*(1), 105-118. doi: 10.1016/j.cognition.2012.12.001

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 329-337. doi: 10.1037/a0028712

Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental Science, 18*(4), 556-568. doi: 10.1111/desc.12239

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*(2), 225-245. doi: 10.1080/14640749308401045

Thomas, N. A., Wignall, S. J., Loetscher, T., & Nicholls, M. E. (2014). Searching the expressive face: Evidence for both the right hemisphere and valence-specific hypotheses. *Emotion, 14*(5), 962-977. doi: 10.1037/a0037033

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty. *Human Nature, 4*(3), 237-269. doi: 10.1007/bf02692201

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452-460. doi: 10.1016/S1364-6613(99)01403-5

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior, 27*(2), 131-144. doi: 10.1016/j.evolhumbehav.2005.06.001

Thornhill, R., & Møller, A. P. (1997). Developmental stability, disease and medicine. *Biological Reviews, 72*(4), 497-548. doi: 10.1111/j.1469-185x.1997.tb00022.x

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124*(1), 208-224. doi: 10.1196/annals.1440.012

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. New Jersey, NJ: Princeton University Press.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3*(2), 119-127. doi: 10.1093/scan/nsn009

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13*(4), 724-738. doi: 10.1037/a0032335

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience, 3*(4), 303-312. doi: 10.1093/scan/nsn033

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623-1626. doi: 10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*(1), 519-545. doi: 10.1146/annurev-psych-113011-143831

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813-833. doi: 10.1521/soco.2009.27.6.813

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science, 25*(7), 1404-1417. doi: 10.1177/0956797614532474

Townsend, J. M., & Levy, G. D. (1990). Effects of potential partners' costume and physical attractiveness on sexuality and partner selection. *The Journal of Psychology, 124*(4), 371-389. doi: 10.1080/00223980.1990.10543232

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*(1), 72-86. doi: 10.1177/1088868312461308

Vakirtzis, A. (2011). Mate choice copying and nonindependent mate choice: A critical review. *Annales Zoologici Fennici, 48*(2), 91-107. doi: 10.5735/086.048.0202

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*(4), 471-491. doi: 10.1111/j.2044-8295.1988.tb02747.x

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology, 43*(2), 161-204. doi: 10.1080/14640749108400966

Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review, 11*(3), 482-487. doi: 10.3758/bf03196599

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796-803. doi: 10.1016/j.cognition.2008.07.002

van Osch, Y., Blanken, I., Meijs, M. H., & van Wolferen, J. (2015). A group's physical attractiveness is greater than the average attractiveness of its members: The Group Attractiveness Effect. *Personality and Social Psychology Bulletin, 41*(4), 559-574. doi: 10.1177/0146167215572799

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior, 28*(4), 260-271. doi: 10.1016/j.evolhumbehav.2007.04.006

Walker, D., & Vul, E. (2014). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science, 25*(1), 230-235. doi: 10.1177/0956797613497969

Wang, G., Cao, M., Sauciuvenaite, J., Bissland, R., Hacker, M., Hambly, C., . . . Speakman, J. R. (2018). Different impacts of resources on opposite sex ratings of physical attractiveness by males and females. *Evolution and Human Behavior, 39*(2), 220-225. doi: 10.1016/j.evolhumbehav.2017.12.008

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*(5), 931-941. doi: 10.1016/0042-6989(92)90036-i

Waynforth, D. (1998). Fluctuating asymmetry and human male life-history traits in rural Belize. *Proceedings of the Royal Society of London B: Biological Sciences, 265*(1405), 1497-1501. doi: 10.1098/rspb.1998.0463

Waynforth, D. (2007). Mate choice copying in humans. *Human Nature, 18*(3), 264-271. doi: 10.1007/s12110-007-9004-2

Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. *Journal of the Optical Society of America A: Optics and Image Science, 31*(4), A283-A292. doi: 10.1364/josaa.31.00a283

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23*(3), 230-249. doi: 10.1016/0022-1031(87)90034-5

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS ONE, 2*(8), e710. doi: 10.1371/journal.pone.0000710

Whitney, D., Haberman, J., & Sweeny, T. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. Werner & L. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 695-710).

Whitney, D., & Levi, D. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160-168. doi: 10.1016/j.tics.2011.02.005

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105-129. doi: 10.1146/annurev-psych-010416-044232

Wiggins, J. S., Phillips, N., & Trapnell, P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology, 56*(2), 296-305. doi: 10.1037//0022-3514.56.2.296

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. doi: 10.1111/j.1467-9280.2006.01750.x

Willis, M. L., Palermo, R., & Burke, D. (2011). Judging approachability on the face of it: The influence of face and body expressions on the perception of approachability. *Emotion, 11*(3), 514-523. doi: 10.1037/a0022571

Willis, R. (1960). Stimulus pooling and social perception. *The Journal of Abnormal and Social Psychology, 60*(3), 365-373. doi: 10.1037/h0048912

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325-1331. doi: 10.1177/0956797615590992

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science, 7*(4), 331-338. doi: 10.1177/1948550615624142

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*(2), 189-202. doi: 10.1177/106591290605900202

Wincenciak, J., Dzhelyova, M., Perrett, D. I., & Barraclough, N. E. (2013). Adaptation to facial trustworthiness is different in female and male observers. *Vision Research, 87*, 30-34. doi: 10.1016/j.visres.2013.05.007

Winegard, B., Winegard, B., Reynolds, T., Geary, D. C., & Baumeister, R. F. (2017). One's better half: Romantic partners function as social signals. *Evolutionary Psychological Science, 3*(4), 294-305. doi: 10.1007/s40806-017-0095-7

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia, 45*(1), 195-206. doi: 10.1016/j.neuropsychologia.2006.05.009

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277-283. doi: 10.1038/nn816

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision, 16*(15), 28-28. doi: 10.1167/16.15.28

Yamanashi Leib, A., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision, 14*(8), 26-26. doi: 10.1167/14.8.26

Yamanashi Leib, A., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia, 50*(7), 1698-1707. doi: 10.1016/j.neuropsychologia.2012.03.026

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145. doi: 10.1037/h0027474

Yovel, G., Tambini, A., & Brandman, T. (2008). The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. *Neuropsychologia, 46*(13), 3061-3068. doi: 10.1016/j.neuropsychologia.2008.06.017

Zaidel, D. W., Bava, S., & Reis, V. A. (2003). Relationship between facial asymmetry and judging trustworthiness in faces. *Laterality: Asymmetries of Body, Brain and Cognition, 8*(3), 225-232. doi: 10.1080/13576500244000120

Zaidel, D. W., Chen, A. C., & German, C. (1995). She is not a beauty even when she smiles: Possible evolutionary basis for a relationship between facial attractiveness and hemispheric specialization. *Neuropsychologia, 33*(5), 649-655. doi: 10.1016/0028-3932(94)00135-c

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28*(2), 238-249. doi: 10.1177/0146167202282009

Zebrowitz, L. A., & Montepare, J. M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. *Evolution and Social Psychology*, 81-113.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497. doi: 10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-Eyed" and "Crooked-Faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin, 22*(12), 1258-1269. doi: 10.1177/01461672962212006

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235. doi: 10.1038/nature06860