How Extraneous Facial Markings Affect Recognition

By

© Victoria A. J. Kavanagh

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

Memorial University of Newfoundland and Labrador

January 2024

St. John’s, Newfoundland and Labrador, Canada

Abstract

The ability to recognize faces accurately is of significant concern across an array of fields, being critical to our legal system by playing a pivotal role in eyewitness identification and even beginning to be incorporated into artificial intelligence (e.g., facial recognition software). It is important, then, to understand when facial recognition is accurate and when it is not. Most research on facial recognition is limited to the impact of internal facial features (e.g., eyes, nose, and mouth) on accuracy. In the current study, I tested the accuracy of human recognition for faces containing extraneous markings (e.g., moles, scars, tattoos). In Experiment 1, I had participants study a series of faces, some of which were altered to include a mole or a scar, and then completed an old-new recognition task. I found that unaltered faces were more discriminable than faces in either altered condition; there were no differences between the altered conditions. In Experiment 2, I used a similar study phase but tested memory using two-alternative-forced-choice between the studied target and an alternative version of the same face. I once again found better discriminability for unaltered faces compared to faces with scars but only marginal differences compared to those with moles. In Experiment 3, I compared unaltered faces to faces with moles and faces altered to be more traditionally distinct (e.g., altered eye size); an old-new recognition test was used. I found better discriminability for the distinct faces in comparison to the other conditions; no difference was found between the other conditions. Finally, in Experiments 4a and 4b, using a different set of face stimuli, I compared old-new recognition for faces with and without tattoos, with the latter also using inversion at test. I found similar discriminability between faces with tattoos and those without was found in 4a, while better discriminability was found for the faces with tattoos in 4b. Together, these results indicate that the effect of extraneous markings depends more on how different markings are from each other than simply on the presence of one. Additionally, results also indicate that the presence of the markings shifts from processing the markings as extensions of faces and toward processing them as objects. Implications for the memorability of faces with extraneous markings and the multi-dimensional space framework are discussed.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Kathleen Hourihan. This long but worthwhile journey would not have been possible without her incredible mentorship and guidance. Her knowledge is endless, her ability to teach seems effortless, and the care with which she treats all of her students is truly awe-inspiring. She trusts in her students' abilities, allows them to explore both academically and professionally, and ensures that they have an accepting and open space to thrive in. Anyone would be lucky to try to follow in her footsteps, and words can not begin to describe all she has helped me to achieve academically, professionally, and personally.

I would also like to thank my committee members, Drs. Aimée Surprenant and Brent Snook. Their guidance and feedback have been insightful and integral in making this a dissertation that I can be proud of.

Lastly, I would like to thank my family and friends, without whom this would not have been possible. My parents, Debera and Joe, have supported my academic ambitions throughout my entire life, and I will always appreciate their best attempts at trying to understand what I have been doing over the last four years. They have provided me with infinite practice in trying to relay my research to a general audience. I hope that I have made them proud. And finally, my friends, Krystal and Nikki, have gotten me through my entire graduate journey like no other. My greatest and happiest moments can be attributed to them, and they will forever have had a long-lasting impact on my life. Getting sorted into a room with them at Science Atlantic is one of the most incredible things to have happened to me.

Land Acknowledgements

I respectfully acknowledge the territory in which we gather as the ancestral homelands of the Beothuk, and the island of Newfoundland as the ancestral homelands of the Mi'kmaq and Beothuk. I would also like to recognize the Inuit of Nunatsiavut and NunatuKavut and the Innu of Nitassinan, and their ancestors, as the original people of Labrador. We strive for respectful relationships with all the peoples of this province as we search for collective healing and true reconciliation and honour this beautiful land together.

Table of Contents

List of Tables

List of Appendices

How Extraneous Facial Markings Affect Recognition

Faces can carry one's identity, which groups they are a member of, their current emotional state, and more (Bruce & Young, 1986). The ability of humans to accurately recognize the faces of others has vast implications for their day-to-day lives. For example, not recognizing a friend of a friend one has had a previous encounter with can lead to feelings of embarrassment from both parties, with one having created an awkward situation and the other left feeling that they are unmemorable. Survival can rely on distinguishing friend from foe; one must be able to recognize those who should be avoided to prevent unwanted conflict (Wilson & Hugenberg, 2010). Additionally, the incorrect identification of the suspect of a crime can lead to an innocent person spending years in prison for a crime they did not commit (Innocence Project, 2020).

Given the crucial social implications that stem from facial recognition, or the lack thereof, it is not surprising that humans are well-primed to attend to and remember many faces. In fact, facial recognition is so important to humans that neuroscientists have hypothesized that there is a special area located in the brain's inferior temporal lobe, the *fusiform face area* (FFA), where faces are processed. Though, it should be noted that there has been some debate as to whether this area is solely dedicated to facial processing or if other recognition processing occurs here as well (for a review, see Kanwisher & Yovel, 2006). Regardless of the function of this processing area, humans are expert face recognizers with the ability to recognize thousands of different faces in their lifetime (Jenkins et al., 2018), an ability that has been recognized as impressive by many for decades (Galton, 1883; Young et al., 1987). As noted by Galton so long ago, faces are generally similar to one another and thus require the ability to notice minute differences between them, making successful discrimination among so many faces an extraordinary feat. Specifically, there are but a few different features that make up a face, including those labelled as *internal* to the face (e.g., eyes, nose, and mouth) and those labelled as *external* (e.g., chin, ears, eyebrows, and hair; Althoff & Cohen, 1999; Heisz & Shore, 2008). While these features can differ in size, colour, and location, it is still impressive to be able to differentiate among the varied combinations of each feature that make up each unique face.

The current study will further explore the limits of facial recognition abilities. Specifically, the effect of unique, though often encountered, facial markings (i.e., moles, scars, and tattoos) on recognition will be examined. Although these markings are not typical of faces, as they are certainly not seen on the majority of faces one encounters, it is not entirely uncommon to see them either. What is known about the impact of such extraneous markings on facial processing and recognition is limited and thus warrants further exploration. Chapter 1 of this dissertation presents an overview of facial recognition. This overview describes the foundations of facial processing, various theories of how faces are recognized, what features lead to better recognition of some faces over others, and how facial recognition is used outside of the laboratory. In Chapters 2 – 5, direct tests of the effect of these extraneous markings on recognition are presented. Included are different memory tests comparing altered and unaltered faces, a test comparing altered and unaltered faces to faces with *distinct* features, and a test comparing untattooed faces to ones that are tattooed. Finally, in Chapter 6, a general discussion will bring together each finding and consider how they apply to the current facial recognition literature.

The Foundations of Facial Processing

When viewing a face, there are prominent features that stand out. As mentioned previously, these features could be internal to the viewing region, such as the eyes, or external to the region, such as the hair (Althoff & Cohen, 1999; Heisz & Shore, 2008). Despite being so few in number, each feature carries a wealth of information that is needed for later identification of the face they are present on; some of this information is referred to as *local* (Leder & Bruce, 1998) or *isolated* (Diamond & Carey, 1986), while other information is referred to as *relational* (Diamond & Carey, 1986; Leder & Bruce, 1998) or *configural* (Bartlett et al., 2003). For the purposes of this paper, the terms local and isolated will be used interchangeably with one another, as might relational and configural. As explained by Diamond and Carey (1986) and Leder and Bruce (1998), local information is that which is isolated to that feature itself. This is what can be described about that feature on its own, without having to reference any other feature. Examples of this type of information include eye colour, hair style, nose shape, and eyebrow thickness. Relational information, on the other hand, describes one feature in relation to one another or how the features are configured on the face. For example, relational information can include how far apart the eyes are from one another or how far below the mouth is from the nose.

Despite it being common and easier to use local information when describing a person, configurational information is what is most important in the processing and subsequent recognition of a face (e.g., Bartlett et al., 2003; Diamond & Carey, 1986; Piepers & Robbins, 2012). When describing configural information, Diamond and Carey differentiated between first- and second-order configurations. First-order configuration is the layout of the face we are used to seeing: Eyes above nose above mouth; when processing a face, we must ensure that the features are in the order they should appear. Second-order configuration is the exact metric distances between features, and their orientation (i.e., a 0-degree orientation on the face versus slanted in either direction). Diamond and Carey further explained that second-order configuration facilitates facial recognition, allowing for the unconscious encoding of these metrics to help later differentiate one face from the next.

Important to note regarding the use of configural information in facial recognition, however, is that the weight put on such information is debated. For example, Sandford and Bindemann (2020) pointed out that, in many circumstances, configural information is not stable. Simply by changing one's viewpoint of a face or the facial expression, the exact metric distances between features, at least to the viewer's eye, can change. If this information is used to differentiate faces, recognition should subsequently be hindered by such changes, though it rarely is. To further explore the use of configural information, Sandford and Bindemann conducted experiments using stimuli that included well-known (North American) celebrities that should have been familiar to the Canadian participants, as well as less-known (British and Australian) celebrities that should not have been familiar to the participants; note that the impact of familiarity on facial recognition will be discussed in more detail later in this introduction. In Experiment 2 of their study, the configuration of the eyes was altered on a copy of each face. Participants were then shown either the original, unaltered face or the altered version and had to determine whether there was a change in configuration. Participants were also asked to identify the person shown on the screen. They found that, for familiar faces, participants performed well at differentiating between changed and unchanged faces, though they could still accurately identify the celebrity when the face was changed. These results highlighted that while participants were sensitive to changes in configuration, indicating that they are encoding that information, alterations to that configuration did not prevent identification. If recognition relied primarily on the encoded configuration of a face, altering the configuration should have impacted identification.

More in line with the notion that configuration is not predominately involved in facial recognition, some have proposed that local and configural information are weighted more equally. For example, Rhodes (1988) posited that there are first- and second-order *features* instead of configurations. Essentially, information about features on their own (i.e., local information) is first-order, while relational information, like the distance between features, is second-order. Additionally, Rhodes also promotes the idea of higher-order features, those that combine first and second order information to provide additional information, such as one’s age. The critical thing to note when comparing how different sets of features should be weighted concerning recognition is that it is widely accepted that faces are processed holistically (e.g., Hayward et al., 2016; Maurer et al., 2002; Rossion, 2008), supporting the idea that each feature set is weighted equally.

Holistic versus Feature-Based Processing

With most forms of recognition, the target that one is trying to encode and later recognize is processed in a part- or feature-based manner (Biederman, 1987). Take a clock, for example. The circular frame, numbers, and hands would all be encoded as individual parts that are later combined to form the desired object. Faces, however, are not processed in the same manner. Instead of independently encoding the different sets of features, whether internal or external, local or relational, everything is encoded as one, through *holistic processing* (Tanaka & Farah, 1993). Given that this processing does rely on the configural layout of the face, this processing has also been referred to as *configural processing*, not to be confused with the configural information discussed previously (McKone, 2008; McKone & Yovel, 2009; Rossion, 2008). The terms holistic and configural processing are often used interchangeably (McKone, 2008), though some support a distinction between the two (Maurer et al., 2002).

Significant evidence that supports the fact that faces are processed holistically, and the importance of configuration to that processing, comes from the *inversion effect* (Valentine & Bruce, 1986a; Yin, 1969). The inversion effect is the finding that the recognition of faces decreases in accuracy significantly when they are presented upside-down instead of right-side-up whereas the recognition of other objects does not. In Experiment 1 of his study, Yin presented participants with pictures of faces, houses, airplanes, and stick figures in motion. At test, each studied picture was presented alongside a new category match, and participants had to indicate which of the two was studied. Importantly, for half of these study and test trials, the pictures were inverted. For the upright pictures, there were significantly fewer errors in identifying the studied faces compared to each of the other objects. For the inverted pictures, however, the face advantage had gone away. While there was also an inversion effect for two of the ordinary objects, the houses and stick figures, where errors did increase for the inverted pictures, there were more errors and a larger magnitude of the effect for the inverted faces. This overall match between encoding and retrieval would be critical for items that are holistically processed, like faces, whereas it would be less important for items that use parts-based processing, like objects. It should be noted, however, that there has been some debate over the strength of this effect (see Richler & Gauthier, 2014 for a meta analysis and review on holistic face processing), with some researchers positing that inversion only delays holistic processing (Richler et al., 2011). Others have also argued regarding whether the effect is actually attributed to other factors, such as differences in individual features (Civile et al., 2014; McKone & Yovel, 2009).

Further, and arguably better (Hayward et al., 2016), support for holistic over feature-based processing comes from the *composite effect*, in which it has been found that it is more difficult to recognize a face if either its top or bottom half has been aligned perfectly with the top/bottom half of a new face rather than if the two different halves are presented misaligned (Piepers & Robbins, 2012; Young et al., 1987). In Experiment 1 of their study, Young et al. split pictures of the faces of five notable figures in half horizontally. They then created new composites using each face's top and bottom halves; some of these composites were aligned (properly forming a new face), and some were misaligned (the top halves were placed above the bottom halves, but offset slightly).

Participants were then presented with these composites and were asked to identify which figure was present in one of the specified halves. Interestingly, the task was found to be difficult when the faces were aligned; participants were significantly faster in their identification of the specified half when the two halves were misaligned. When the two halves were joined to make one new face, it became difficult to tease apart one half from the other, supporting the notion that, when presented as one unified item, humans automatically process it as such (Hayward et al., 2016; Maurer et al., 2002). In addition to this experiment, Young et al. inverted half of the faces in Experiment 2 of the study. In support of the inversion effect, the composite effect disappeared for these inverted faces, providing greater support for the tie between configuration and holistic processing. Lastly, in Experiment 3, Young et al. found a composite effect in unfamiliar faces. Instead of using well-known figures, participants were first trained to associate randomized surnames with new, unfamiliar faces. At test, composites were again made in a fashion similar to Experiment 1 and a composite effect was found once more. Support for holistic processing, specifically through the composite effect, has also been found across ages, going as young as three months old (Turati et al., 2010).

One final example of support for holistic processing comes from the *part-whole task* (Tanaka & Farah, 1993). In their study, Tanaka and Farah had participants study face-name pairings with faces whose internal features were either intact (configured normally) or scrambled (the nose could be placed where the right eye could be while the right eye was in place of the mouth, for example). At test, for half of the studied faces, participants were asked to recognize individual parts; for example, they were presented with two isolated noses and asked which belonged to person X. For the other half of the faces, however, participants were presented with two versions of the same full-face with the only difference between the two being a change in a single feature (e.g., the noses were different between the two); participants then had to identify which feature belonged to the original face. For faces that were initially presented intact, identification of the correct feature was significantly better when the full face was presented. For faces initially presented as scrambled, identification of the correct feature was significantly better when the feature was presented in isolation. Together, this provides more credence to the notion that, when presented with a face in its normal configuration, each part is encoded in conjunction with the others and is therefore recognized better as such.

One final note on holistic versus feature-based processing is the question as to whether feature-based processing occurs at all and how important it is if it does occur. A robust finding, especially after the rise of eye-tracking technology, is that some features are prioritized over others when viewing a face (Rollins et al., 2019). For example, both adults (Althoff & Cohen, 1999; Heisz & Shore, 2008) and infants (Oakes & Ellis, 2013) will often spend more time directing their gaze toward the internal features of the face rather than the external ones. Additionally, as mentioned previously, when describing someone, most people will do so on a part-by-part basis.

Given the preference in eye movement and descriptions, is there evidence that some feature-based processing occurs? In the study described above by Tanaka and Farah (1993), though identification of individual features was significantly better when those features were studied and tested on intact faces (with a 73% hit rate), there was still high identification of those features that were studied on intact faces and tested in isolation (63%). So, some individually identifiable information would appear to be encoded through the feature. Piepers and Robbins (2012), in their review of holistic facial perception, support the idea of a *holistic/part-based model* in which both holistic and feature-based processing occur simultaneously. Though, given the findings outlined above (e.g., Yin, 1969; Young et al., 1987), and other findings in which facial recognition can occur when faces are encoded only in the periphery, limiting access to feature-based encoding (McKone, 2004), feature-based processing does appear to only be a potential additive to holistic processing.

The Process of Facial Processing

While features and configurations explain which parts of the face are used in facial processing, they do not actually explain the process of encoding a face into memory or retrieving one from it. Among those who have attempted to describe the process of facial recognition are Goldstein and Chance (1980), who promoted the idea of *schema theory*. Schemata are patterns of information that allow us to organize our knowledge (Rumelhart, 1980; Vernon, 1955). The more experiences we have with certain stimuli, the more our schema for that type of stimulus develops and broadens. This schema then controls how we interact in further experiences with those stimuli, dictating what parts are attended to. These interactions then become quick and automatic, thanks to the schema. Additionally, the relationship between the schema and the stimulus is bi-directional; further interactions with the stimulus can lead to adjustments to the schema (Goldstein & Chance, 1980; Vernon, 1955).

Applying schemata to faces, Goldstein and Chance (1980) explained that, from a young age, we develop a face schema. Our interactions with others summate, leading us to efficiently and expertly encode and later recognize an extraordinary number of faces in our lifetime. This schema controls how we attend to faces and what parts are attended to in order to maximize encoding and compare faces we are attempting to recognize to those already stored in memory. Support for this view of facial processing comes from the age differences found within facial recognition. For example, as highlighted above, there are age differences in the viewing preferences of infants (Oakes & Ellis, 2013), children (Meaux et al., 2014), and adults (Althoff & Cohen, 1999). Furthermore, age differences have been found in biases toward the recognition of faces belonging to one's race, which will be discussed in more detail later in this paper (Goldstein & Chance, 1980). According to Goldstein and Chance, these age differences support the idea of a face schema in that the processing develops over time, as there are more interactions with faces, and then the processing becomes very rigid. In the case of specific biases, or even the inversion effect, the majority of one's experiences are invariable (e.g., primarily interacting with those of the same race and seeing upright faces). Due to a lack of diversity, the schema cannot easily adjust when processing these different faces.

In contrast to schema theory, some have instead posited that facial processing is more akin to the *prototype theory* of categorization (Rosch, 1973; Valentine, 1991; Valentine & Bruce, 1986b). A prototype is the prime example or mental image one holds in their mind for a given category. This prototype is a merged image of every encounter we have had with members of the given category, and current recognition of a category member is achieved by comparing that member to the prototype. If it is believed that the stimulus in front of us is a close enough match to the prototype, we will surmise that it is, in fact, the item we believe it to be (Hampton, 1993). This concept differs from schema theory primarily in that a prototype is a member of a category that is used as a comparison point, whereas a schema is the pattern of thought that guides how we interact with a member of a category. Applying prototypes to facial recognition, Valentine and Bruce (1986b) explain that the summation of one's experiences leads to the development of a face prototype. When later attempting to recognize a face, the face is compared to the prototype that is stored. This idea is supported by the differential treatment of typical versus atypical faces, which will be discussed in more detail later in this paper (Valentine, 1991). Ultimately, typical faces are classified as faces more quickly than atypical ones; if faces are compared to a singular prototype, which is the average of every face one has encountered, then the comparison should be quicker.

Multi-Dimensional Space Framework

Building on previous theories, the now more widely accepted explanation of facial processing comes from Valentine (1991), who proposed the *multi-dimensional space (MDS) framework*. Through this framework, Valentine posits that faces are encoded onto a multi-dimensional plane or space, and later recognition involves searching for the stored representation of the target stimuli in that space; some people refer to the multi-dimensional space as *face-space* (e.g., Valentine et al., 2016). To understand this framework, it is easiest to envision a coordinate plane with four quadrants; note that, given that this is a multi-dimensional framework, there are believed to be many more dimensions than just two. The origin, or centre of the axis, is the amalgamation of the average of every facial feature one possesses. This average can include size, shape, configuration, and colour. For example, say that the y-axis of this coordinate plane represents eye size, and the x-axis of this plane represents the distance between the eyes. A face with eyes of average size and distance apart would be stored at the centre of the axis. A face with large eyes but an average distance apart would be stored north of the centre, while a face with eyes that are average size but far apart would be stored to the right of the centre. On the other hand, a face with small eyes that are close together would be stored in the third quadrant, in the bottom-left of the plane. When a face is encountered and encoded, it is stored in this large multi-dimensional space after every facial feature has been taken into account.

The most important aspect of MDS, given how faces are stored, is the central point in the face-space (Valentine, 1991). This origin is usually the reference point for both storing and retrieving faces. One of the key assumptions Valentine makes regarding face-space is that this is the point of central tendency, with a higher density of stored faces around this region. This assumption is justified in that a face is more likely to be average than not. Those faces which deviate from average would be stored further away from the central point, in any given direction based on which feature(s) differed; the further away from the centre one moves, the less densely populated the stored representations should be.

Support for the assumption of central tendency is two-fold. First, as mentioned above, average faces are classified as faces more quickly than those which deviate from average or are considered *distinct* (see below for a more detailed exploration of facial distinctiveness). Note that, for the purposes of this paper, the terms *average* and *typical* will be used interchangeably, as will *atypical* and *distinct* with one another. Additionally, for the purposes of this paper, discussions involving *distinct* faces will be under the assumption that these faces are those that deviate from average in face-space, as described by Valentine (1991; this also relates to *secondary distinctiveness,* as described by Schmidt, 1991). In Experiment 4 of his study, Valentine presented participants with average and distinct face stimuli. Stimuli were also presented either intact or jumbled (e.g., nose above mouth above eyes) and either upright or inverted. The participants had to determine whether the face was a proper face (intact) or jumbled regardless of its orientation. Classification for the intact faces was indeed quicker for the average ones, though no differences were found between the average and distinct jumbled faces. It should be pointed out, however, that Valentine noted that this would not be expected of jumbled faces, regardless, as face-space does not account for such faces. This is support for central tendency as there are many more comparison points for typical faces in face-space; with an abundance of readily available stored representations, one should be quicker at confirming that the image in front of them is a face compared to having to spend more time searching a less densely populated area of face-space. Other support for the central tendency comes from findings that there is larger error when identifying typical than atypical faces. In Experiment 2 of his study, Valentine presented participants with upright average and distinct faces; participants later had to identify which faces were studied in an old-new recognition test. A mirror effect was found; not only were hits (correctly identifying a studied item as old) higher for the distinct upright faces, but false alarms (incorrectly identifying new items as old) were also lower. This difference in error supports the assumption of central tendency as there is greater competition among stimuli in a densely populated area. If more stimuli are present when trying to remember a face, the chances of confusing one face for the next are also greater. If stimuli are more spread out, that competition is lower and more accurate identifications can be made.

Valentine (1991) also notes two possible coding models in reference to the central tendency with MDS. In one, the *norm-based coding model*, this central point is a single prototype. As explained above, a prototype is the mental image of the average of all experiences with a given stimulus (Rosch, 1973). If the central point is where the average of all features meets, it will make sense for it to be represented by a face prototype. After the prototype is established, according to Valentine, newly encountered faces are stored in face-space in relation to that prototype as vectors. The more the new face differs from the prototype, the further away in face-space it is stored. Again, as more faces would be similar to average, the further away from the prototype a vector is stored, the less populated the space is. The other coding model mentioned by Valentine is an *exemplar-based coding model*. An exemplar is an individually stored representation or example of a category member (Medin & Schaffer, 1978). Unlike prototype theory, it is believed by some that every encounter with a stimulus is stored rather than averaged across one another. Within the MDS framework, according to Valentine, using an exemplar-based coding approach, the centre of the face-space is not a singular prototype but rather where the majority of faces are stored. The stored faces, or exemplars of faces, that are similar to one another are stored in close proximity to one another, so if the majority of faces are average or similar to one another, the central tendency naturally forms with that cluster at the centre. Valentine does not give preference for one coding model over another.

While the above explains how faces are stored in face-space, it does not explain how stored faces are later retrieved. This retrieval process also depends on which coding model is preferred, though they both act similarly to one another (Valentine, 1991). When giving preference to the norm-based model, retrieval is conducted by starting at the prototype. Whether one is trying to freely retrieve a face from memory or trying to decide whether a shown target face matches one in memory, search begins at the prototype in the centre and expands to search for the target from there. When giving preference to exemplar-based coding, retrieval begins with the exemplars that closely resemble the target. In either case, comparisons are made to the faces within face-space until a sufficient match is found. Any error that is encountered when retrieving a face from face-space comes from either incorrect storage of the stimuli in the space (e.g., encoding/viewing conditions were poor and did not accurately represent the face) or from the density of faces around the target to be retrieved.

In addition to the assumption of central tendency, Valentine (1991) made two other assumptions regarding face-space that should be mentioned. He asserted that there is no set match for what each dimension within each person's face-space represents and that these spaces are developed centring on each person's own race. To touch on the first point, no two people have an identical face-space. Therefore, no one diagram can be created with dimensions that fully represent how each face is stored. Additionally, which exact dimensions a person uses can quite literally be anything they use to discriminate between different faces (Valentine & Endo, 1992). Each person's experiences shape how their space is created, which leads to Valentine's second point: spaces are designed around one's own race. It has been found that different facial features are important when encoding and retrieving faces of different races (e.g., Ellis et al., 1975; Shepherd & Deregowski, 1981). As our experiences with people of other races tend to be lesser in numbers than with those of our own race (Valentine et al., 1995), preference is given to these faces in face-space, with the dimensions of our own face-space being more tailored to the features prominent in those of our race (Valentine, 1991). This favouring of own versus other-race faces has been supported by the *cross-race effect* (Malpass & Kravitz, 1969), which will be described in more detail below.

Advantages and Disadvantages in Facial Recognition Accuracy

Common within the facial recognition literature is the finding that not all faces are treated equally, as pointed out several times in this introduction thus far. Many types of faces are favoured and hold advantages over others, and the theories and frameworks above support this claim. This section will discuss the advantages of familiarity, distinctiveness, and own-biases in facial recognition, with the latter two tied to the goals of this study.

Familiarity

Faces of those known to us are far more privileged than of those who are not; not only is recognition of familiar faces highly accurate, it is also fast (Bruce, 1986; Ellis, 1981). Accurate recognition of familiar faces also persists when testing conditions are poor. For example, Burton et al. (1999) presented participants with grainy video clips of lecturing staff at a university; some of the participants were familiar with the staff in the clips (students of the staff), and some were unfamiliar with the staff (students with no prior interaction with the staff or seasoned police officers in the community). At test, participants were presented with high-quality headshots of an equal number of seen and unseen staff and were asked to identify which were previously seen in the videos. Despite the poor quality footage shown at study, the participants familiar with the staff in the videos were better able to differentiate between the seen and unseen stimuli at test, even when compared to the seasoned police officers who, in turn, performed no better than the unfamiliar students.

Unfamiliar faces, in contrast to familiar ones, are far less privileged (for a review, see Burton & Jenkins, 2011). Unlike with familiar faces, the difference between the encoding and recognition of unfamiliar faces is more severely impacted by the expression one has (e.g., Mian & Mondloch, 2012; Patterson & Baddeley, 1977), the lighting of the stimuli (e.g., Hill & Bruce, 1996), and even the viewpoint (e.g., Longmore et al., 2008; O'Toole et al., 1998). In fact, the recognition of unfamiliar faces can be poor, even in simple matching tasks. For example, Bruce et al. (1999) presented participants with an unfamiliar target face stimulus alongside an array of ten potential matches; the target match was present in the array in half of the trials and absent in the other half. In some cases, targets held the same neutral expression as their array counterpart, in some they held the same expression but were at a 30-degree angle to the camera instead of head-on, and in the rest they held a smile instead; all potential matches in the array held a neutral expression, head-on to the camera. Despite the simple match required between target and array, participants were only correct, at best, 70% of the time (with accuracy dropping to 66% and 60% when the target was smiling and at a 30-degree angle, respectively). Similarly, participants only rejected the array when the target was absent, at best, 70% of the time. Again, even for a seemingly simple task, the recognition of unfamiliar faces does not hold the privilege that it does for familiar faces, making such identifications more error-prone and less flexible.

Distinctiveness

Just as familiar faces are recognized better than unfamiliar ones, so are distinct or atypical faces generally better remembered than average or typical ones (e.g., Bartlett et al., 1984; Cohen & Carr, 1975; Going & Read, 1974; Winograd, 1981). As described above, distinct faces deviate from average across different facial feature dimensions (Valentine, 1991; Valentine et al., 2016). In addition to greater hit rates (correctly identifying studied faces as old), distinct faces hold other advantages as well; not only are false alarms (incorrectly identifying new faces as old) lower for distinct faces (e.g., Bartlett et al., 1984; Valentine, 1991; Winograd, 1981), but so too are response times (Valentine & Bruce, 1986c), and overall discrimination is higher (Valentine, 1991). Interestingly, Valentine has found that distinct faces are also advantaged when inverted. The advantage of distinctiveness has also been attributed to other findings, such as the finding that less attractive and less likeable faces are recognized better than their attractive and likeable counterparts (with less conventional faces, therefore, being more distinct; Light et al., 1981; Vokey & Read, 1992). Distinctiveness has also been attributed to the *caricature effect*, in which caricatures of face stimuli with exaggerated features (e.g., receded hairline, thick eyebrows, or large doe eyes) are better recognized than unaltered faces; this has been found with both line drawings (Rhodes et al., 1987) and photographic caricatures (Benson & Perrett, 1991).

One early explanation for the advantage of distinct faces was posited by Light et al. (1979), who claimed that atypical faces benefited from deeper processing than their typical counterparts and a lack of other similarly stored faces in memory. Indeed, items that are deeply encoded on a more meaningful level are better remembered than those which are only shallowly encoded on a surface or physical level (Craik & Lockhart, 1972); this has also been evidenced by the benefit Light et al. found to the recognition of atypical faces that were rated for likeableness compared to those just viewed without any rating task. Additionally, distinct faces are stored in more isolation than their typical counterparts, as posited by the MDS framework, which will be examined in more detail briefly (Valentine, 1991; Valentine et al., 2016).

Comparability to familiarity has also been posited to explain the distinctiveness advantage (Bartlett et al., 1984). In Experiment 1 of their study, Bartlett et al. presented participants with typical and atypical faces. One group was presented with a study list followed by a recognition test (unfamiliar condition); one group was given a familiarization task (giving a verbal description of the face or a friendliness rating; pre-familiar condition) before the study list; and one group was given a familiarization task after the study list but before the recognition test (post-familiar condition). For the unfamiliar condition, the advantage for atypical faces was replicated in both hits and false alarms; this advantage was also found in the post-familiar condition, though it was reduced. However, the advantage in false alarms was eliminated for the pre-familiar condition. Bartlett et al. used this finding to support their *familiarity hypothesis*, in which it was claimed that the difference in the distinctiveness advantage is attributed to a difference in perceived familiarity between the two types of stimuli and that increasing the familiarity would differentially impact both. By increasing the familiarity with the typical stimuli, the difference in false alarms was minimized or eliminated altogether.

It should be noted that the familiarity hypothesis has been heavily disputed. For example, Valentine and Bruce (1986c) highlighted that, despite their claim for a differential impact between typical and atypical stimuli with increased familiarity, Bartlett et al. (1984) showed no statistical difference in hit rates; the familiarity hypothesis was supported mainly through false alarm rates. Valentine and Bruce also directly compared familiarity and distinctiveness in their study. They had participants make familiarity judgements on both familiar (staff in their department) and unfamiliar faces that ranged in distinctiveness. While significant negative correlations were found between distinctiveness and response time, and familiarity and response time, the direct correlation between distinctiveness and familiarity was non-significant. This independence between familiarity and distinctiveness has also been supported by others (e.g., Vokey & Read, 1988).

A now more widely accepted explanation of the distinctiveness advantage comes from the MDS framework (Valentine, 1991; Valentine et al., 2016). As mentioned above, the clustering of stored faces within face-space varies in density depending on how much the face deviates from the average. Faces that are average or typical are densely populated in one area of face-space (the centre), but faces that are atypical or distinct are spread out across face-space and are more isolated. Given the magnitude of dimensions represented within face-space, two faces that are similar but deviate on just one dimension (e.g., one has eyes that are close together, but the other has eyes that are far apart) can end up stored quite a distance away from another. The advantages in hits, false alarms, and response times are supported through the framework because of the isolation: Without a large number of competing stored representations, one can easily find a face in face-space, does not have to worry about confusing the face with another similar one, and can quickly make a judgement given the lack of choices.

The MDS framework can also explain the advantages of distinctiveness when faces are inverted (Valentine, 1991; Valentine et al., 2016). As mentioned, the recognition of inverted faces is poor (Yin, 1969). Through the MDS framework, Valentine explains that, given the rigidity at which faces are encoded into face-space, inversion is simply a disruption that increases the number of encoding errors. For typical faces, that leads to more confusion between similar members in a densely populated area; for distinct faces, those errors are less of a detriment as the faces are still encoded relatively isolated from one another, even if they are not stored exactly as they should. Lastly, Valentine’s framework can be used to explain the independent effects of both distinctiveness and familiarity: a familiar face is simply one that has been strongly encoded into face-space; it is independent of the density of faces and the resulting effect of distinctiveness.

Own-Biases

The final advantage that will be discussed is the one for faces that belong to one's own group (e.g., Bäckman, 1991; Cross et al., 1971; Malpass & Kravitz, 1969). As will be discussed shortly, the definition of one's own group can be taken liberally. These advantages, however, are referred to as *own-biases*. The most heavily researched own-bias relates to one's own race, with research dating back over a century (Feingold, 1914). In general, recognition is better for faces sharing one's own race compared to faces of other races; this finding has been dubbed the *cross race effect*, *own race effect,* or *own race bias* (for a recent meta-analysis, see Lee & Penrod, 2022). The own race bias has not only been found in experimental settings but has also been found outside of the lab (in realistic police lineups; for a review, see Meissner & Brigham, 2001). Interest in the own race bias is not surprising as the presence of it outside of the lab can lead to dire consequences, such as false accusations of crimes (Brigham & Malpass, 1985; Kassin et al., 2001) and wrongful convictions (e.g., the case of Ronald Cotton; Innocence Project, n.d.).

The advantages for faces of one's own race vary. They can include higher hit rates and lower false alarm rates (e.g., Chance & Goldstein, 1996; Meissner & Brigham, 2001), a reduced inversion effect (Valentine & Bruce, 1986a), higher confidence in decisions (Corenblum & Meissner, 2006), and better use of both holistic and parts-based feature encoding (e.g., Hayward et al., 2008; Michel et al., 2006; Tajfel, 1970). Same-race faces have also been found to be classified as faces quicker than other-race faces. In Experiment 5 of his study, Valentine (1991) used a similar method to a prior experiment and presented intact and jumbled Caucasian and African-American face stimuli to Caucasian participants. The participants' classification of the faces as faces was quicker for the Caucasian faces. Interestingly, though, other-race faces have sometimes been found to be classified by *race* more quickly than same-race faces. For example, in Experiment 4 of their study, Valentine and Endo (1992) presented Caucasian and Japanese participants with pictures of Caucasian and Japanese face stimuli. They asked the participants to classify which race the face belonged to. While there were no differences in classification speed for the Japanese participants, the Caucasian participants classified the Japanese faces as the correct race quicker than the Caucasian faces.

As can be observed through findings in studies like Valentine and Endo (1992), there are some differences in the own race bias between different groups. Overall, however, the finding is quite robust, with different races showing an advantage for own-race faces over other-race faces (e.g., Bothwell et al., 1989; Meissner & Brigham, 2001; Shapiro & Penrod, 1986). However, it should be noted that the magnitude of the bias does appear to be higher in Caucasian participants than in participants of other races (e.g., Chiroro & Valentine, 1995; Valentine & Endo, 1992). Regardless, the bias also persists across ages, with individuals of all ages exhibiting the own race bias (Pezdek et al., 2003). For example, Corenblum and Meissner (2006) used three groups of school-aged Euro-Canadian students (grades 2-4, 5-6, and 7-8) and one group of young adults (university students) in Experiment 2 of their study. These students were shown pictures of Black-Canadian, Indigenous-Canadian, and Euro-Canadian adults and later performed an old-new recognition task. Discriminability and confidence were higher for the same-race faces across all age groups.

Theories of Own-Biases. Given the long history of the study of the own race bias, it is not surprising that numerous theories behind the cause of the bias exist. Of note, however, it is widely accepted that inherent differences in the faces themselves do not cause it. As noted above, different facial features are thought to be used in the recognition of people of different races. For example, Ellis et al. (1975) found that, using both Caucasian and African-American participants and face stimuli, Caucasian faces were consistently described in terms of the eyes and hair, while African-American faces were consistently described in terms of the nose and mouth. Despite the apparent difference in prominent features, this being the cause of the own race bias has been disputed given the presence of the bias across different racial groups; if some faces were more challenging to recognize than others based on their inherent metrics, then this difficulty should be consistently present (Brigham & Malpass, 1985). Also of note is that it is also widely accepted that the bias does not occur due to prejudices, though some have argued in support of prejudice playing some role. For example, Galper (1973) found a reduced own race bias in Caucasian students enrolled in a Black Studies college course compared to Caucasian students who were not enrolled and surmised that overall racial attitudes might have caused the reduction. Many have since argued against this as the cause (e.g., Brigham & Barkowitz, 1978; Carroo, 1987; Ferguson et al., 2001; Meissner & Brigham, 2001).

A more accepted, though not currently leading, explanation for the own race bias is the number of experiences one has with their own versus other-race faces. This is known as the *contact-hypothesis* (Valentine et al., 1995) or, similarly, the *perceptual expertise hypothesis* (Meissner & Brigham, 2001); note the similarities between these hypotheses and the schema theory of facial recognition discussed earlier (Goldstein & Chance, 1980). Generally, one would have more experiences with same-race faces as families tend to share a racial identity; friend and community groups also tend to share similar racial identities (Valentine et al., 1995). This increased contact can then lead to more expertise in the facial recognition of those race faces, whether it is because of better differentiation among relevant facial features of the same-race face (e.g., MacLin & Malpass, 2001) or worse holistic processing of the other-race face (e.g., Rhodes et al., 1989). These hypotheses can also explain the differential magnitude of the own race bias found across races; a significant representation of Caucasian faces in the media leads to greater exposure of Caucasian faces among those of other-races (Valentine et al., 1995).

Evidence to support the perceptual expertise hypothesis comes from both in-lab and real-world examples of reduced or eliminated biases. Inside the laboratory, training on other-race faces, simply by allowing participants more time and practice discriminating between other-race faces, has been successful in reducing the own race bias when otherwise present (e.g., Elliott et al., 1973; Goldstein & Chance, 1985; Malpass et al., 1973). Do note that some have been unsuccessful in finding this training advantage (e.g., Lavrakas et al., 1976; Malpass & Kravitz, 1969; Ng & Lindsay, 1994). Outside of the laboratory, a reduced own race bias has been found in multicultural populations (Bar-Haim et al., 2006), integrated (versus segregated) neighbourhoods (Cross et al., 1971), and integrated schools (Feinman & Entwisle, 1976). Interestingly, further support for the perceptual expertise hypothesis was provided by Sangrigoli et al. (2005), who found a reversed own race bias in adults of Korean descent who were adopted and raised by Caucasian-European parents.

The MDS framework has also been used to explain the own race bias. Essentially, given how different facial features are attended to differently depending on the race of the face (Ellis et al., 1975; Shepherd & Deregowski, 1981), the encoding of such a face does not correctly match on to a person's face-space (Valentine, 1991; Valentine et al., 1995; Valentine et al., 2016). As Valentine and colleagues explain, other-race faces act similar to distinct faces in that they are encoded away from the centre of face-space due to not matching the average feature dimensions typically used in the space. However, they differ from distinct faces in that they still end up clustered around one another in face-space due to the difference in how much weight is given to the different feature dimensions. This off-centre clustering can then lead to decreased hit rates and increased false-alarm rates. Support for the MDS account of the own race bias comes from Hills et al. (2013), who used eye-tracking to record where participants looked when viewing faces; Caucasian and Black-British participants and face stimuli were used. Regardless of the race of the face stimuli, when freely observing the faces, each group naturally spent longer looking at the features prominent in their own-race faces (e.g., the Caucasian participants spent longer viewing the eyes while the Black-British participants spent longer viewing the nose). In other words, each group focused on the features that better matched the dimensions within their own face-space, regardless of the actual race of the face shown.

The final and most currently accepted theory of why the own race bias occurs is the c*ategorization-individuation model* (CIM; Hugenberg et al., 2010). The CIM originated from similar predecessors, such as the *cognitive-disregard model* (Rodin, 1987) and the *feature-selection model* (Levin, 1996, 2000). These two older models suggest that people are selective in whether they are processing features useful in the later recognition of individuals or disregarding such processing. Through the CIM, Hugenberg et al. break the selection down into the categorization or individuation of faces. Categorization is surface-level, unmotivated encoding that stops at basic-level categories (e.g., race, age, or gender). At the same time, individuation is deep, extensive, motivated encoding that incorporates features and dimensions that are useful in later recognition. The critical aspect of the CIM is that the categorization of any face that is encountered is quick and automatic (Ito & Urland, 2003), so it is, therefore, *motivation* that drives the individuation of faces (e.g., Hugenberg et al., 2007; Hugenberg et al., 2010; Macrae & Bodenhausen, 2000). The cause of the own race bias, then, is that people are more motivated to remember faces of those seen as *in-group* members as opposed to *out-group* members as those are the ones they would benefit from remembering in future encounters (e.g., more likely to be a member of a friend group; Hugenberg et al., 2010). In-group members can be anyone with whom we share an identity, such as those of the same race, age, gender, nationality, or even job or school program. Given this, it can be predicted that the own race bias can be reversed purely through a change in motivation and that an own-bias can occur with more than just race (which it does and will be discussed below).

Indeed, there is plenty of support in favour of individuation and motivation driving the own race bias. For example, Hugenberg et al. (2007) presented Caucasian participants with both Caucasian and African-American face stimuli and warned some of them about the own race bias. Specifically, they told the participants about the bias and explained that they should pay attention to what differentiated the faces. This simple instruction, without the need for intensive training, was enough to eliminate the bias in recognition accuracy (though some have failed to replicate this finding using the same instructions; see Tullis et al., 2014). Similarly, Pauker et al. (2009) presented Caucasian and African-American participants with Caucasian, African-American, and racially ambiguous faces to study. In Experiment 1, the racially ambiguous faces were remembered only as well as other-race faces. However, in Experiment 2, participants were explicitly told to pay attention to how they categorized biracial faces; recognition of the ambiguous faces significantly increased in line with recognition of own-race faces (for similar findings using racially ambiguous Hispanic and African American participants and faces, see Hourihan et al., 2013). Other support for motivation comes from Wright et al. (2003), who found a reversed own race bias in Black South Africans and highlighted the intense power dynamics happening in the country at the time and how it would have been more motivating for the participants to remember the faces of those who looked similar to those who held power over them. Lastly, Shriver and Hugenberg (2010) presented Caucasian participants with pictures of Caucasian and African-

American face stimuli that were assigned either high-power occupation titles (e.g., doctors or chief executive officers) or low-power occupation titles (e.g., mechanics or plumbers). It was hypothesized that participants would be more motivated to remember the faces of those with higher-power positions, regardless of race, and that is what was found.

Other Own-Biases. As mentioned above, given that the basis of the own race bias is thought to be due to in-group versus out-group motivation, the CIM also supports the presence of other types of biases (Hugenberg et al., 2010). Indeed, other biases, such as the *own age bias* (for reviews, see Anastasi & Rhodes, 2005; Wiese et al., 2013) and the *own gender bias* (sometimes referred to as the own sex bias; for a review, see Cross et al., 1971; Herlitz & Loven, 2013; Sporer, 2001) have gained attention in recent years. Additionally, better recognition has been found for faces labelled as belonging to one's same university or personality type (Bernstein et al., 2007), sexual orientation (Rule et al., 2007), and religious beliefs (Rule et al., 2010).

Important to note regarding these other own-biases is that they do not appear to be as robust as the own race bias. For example, some studies have found that both young and older adults recognize faces of similar age groups better than the opposing age group (e.g., Anastasi & Rhodes, 2005; Bäckman, 1991). Similarly, some studies have found that children also recognize faces belonging to their own age group better (e.g., Anastasi & Rhodes, 2005; Crookes & McKone, 2009). However, other studies have found the bias present in young adults only and not older adults (e.g., Bartlett & Leslie, 1986; Fulton & Bartlett, 1991). Additionally, the own gender bias has also been found to be more consistent in women viewing female versus male faces than men viewing female versus male faces (e.g., Herlitz & Loven, 2013; Lewin & Herlitz, 2002; Lovén et al., 2011; Lovén et al., 2012).

Regarding explaining the own age bias and own gender bias, some have posited perceptual expertise as the cause. For example, most early caregivers are women (Rennels & Davis, 2008); this could explain the smaller or negligible effect of the bias in men as they would have increased contact with the opposite gender (Herlitz & Loven, 2013). Similarly, a significant amount of time is spent with those in similar age groups (Wiese et al., 2012) and, in some cases where large amounts of time is spent with those of other age groups (e.g., teachers of young children), a diminished own age bias has been found (Harrison & Hole, 2009). However, Harrison and Hole noted that if perceptual expertise was the cause, the own age bias should not exist in younger adults when recognizing children's faces, as they should have ample experiences with younger faces from their youth.

As with the own race bias, the CIM is found to be a more fitting explanation for these biases as well (Hugenberg et al., 2013). For example, Hills et al. (2018) used a sample of gay and straight adults in Experiment 4 of their study; each group contained both men and women. Participants were presented with male and female faces to study and were later given an old-new recognition test. While the own gender bias was present for the group of gay participants, no such bias was found in the group of straight participants. This finding would support the CIM of own-biases as gay individuals would be more motivated to remember faces of same-sex individuals while straight individuals would be more motivated to remember the faces of opposite-sex individuals in line with their sexual orientations.

Facial Recognition Applied to Naturalistic Contexts and the Role of Extraneous Markings

Thus far, most of what has been covered regarding facial recognition has focused on in-lab and theoretical findings and applications of the literature. This final section will describe research and findings more applicable to facial recognition usage within naturalistic contexts.

Eyewitnesses

One of the most commonly studied applications of facial recognition that is relevant outside of the laboratory is eyewitness memory (for a review on the role of eyewitnesses in the legal system, see Wells & Olson, 2003). Often, witnesses to a crime are crucial in the later identification of those suspected of committing said crime. Not only are eyewitnesses useful in creating the description of suspects to aid in the police search for them, but eyewitnesses are also used in the recognition match to confirm that the person that has been arrested is indeed the one who committed the crime. Additionally, eyewitnesses can sometimes be *the* leading evidence used to convict suspects (Smith et al., 2004). Altogether, this highlights the massive weight put on eyewitness identifications, and, as such, it would be essential to ensure that this recognition is error-proof, as errors can lead to life-changing consequences.

Unfortunately, it has been found that eyewitness identifications are not error-proof, and many people are wrongfully convicted as a result. For example, the Innocence

Project (2020), an organization which seeks to overturn and prevent wrongful convictions, found that several hundred of the exonerations that have taken place in the United States have involved eyewitness misidentification, making it one of the leading contributors of wrongful convictions in the country. These exonerations have been both DNA-based, with misidentification involved in around 70% of those cases, and non-DNA based.

The reasons why misidentification occurs vary; it can occur at the encoding stage when the eyewitnesses are first viewing the suspect or at the retrieval stage when they are making the actual identification. At the encoding stage, many factors can impact how well the face of the suspect is encoded, some of which has been discussed already. For example, the general diminished recognition of unfamiliar faces compared to familiar ones can play a part in later misidentification (Burton & Jenkins, 2011). The encoding of cross-race faces (the own race bias) can also be a cause of concern; cross-race misidentification has been found to have occurred in numerous exonerations in the United States (Scheck et al., 2003), and Canada and the United Kingdom (Smith et al., 2004). Additional factors to be cautious of include bad lighting (Hill & Bruce, 1996; Wells & Olson, 2003), disguises that cover the eyes (e.g., sunglasses; Hockley et al., 1999) or hair (Cutler et al., 1987), the addition of facial hair (Read et al., 1990), and the removal of make-up (Ellis et al., 1978).

Attention is another factor at the time of encoding that can impede later identification. Generally, some eyewitnesses might not be aware that they should be paying attention to what is happening around them, and some might not be directing their attention to where they should be, even if they are aware. For example, Leippe (1978) staged the theft of a package directly in front of participants; before or after the theft, some participants were told the item contained a high-value item, while others were told it contained a low-value item. Later identification was highest amongst those told the package contained a high-value item before the theft despite all participants being aware of the package and being present during the theft. Interestingly, many of the participants in the other conditions were not even aware of the theft until after it had occurred. Problematic attentiveness has also been highlighted while examining the *weapon-focus effect* (for a review, see Steblay, 1992). The presence of a weapon has been found to draw one's attention away from a suspect's face and toward the weapon itself, leading to poor identification of the suspect, a finding that has been supported through eye-tracking (Loftus et al., 1987).

In addition to misidentification errors occurring at the time of encoding, they can also occur at and around the time of retrieval. For example, eyewitnesses can be influenced by talking to co-witnesses and the descriptions given by the co-witnesses in their own identification process (Clark & Wells, 2008). Additionally, the instructions given to the eyewitness by law enforcement during the identification process can also lead or re-direct them to a particular member in a lineup (Wells & Olson, 2003).

How suspect lineups are constructed, whether simultaneous or sequential, can also greatly impact misidentification at retrieval (Wells, 2008). Simultaneous lineups are ones in which multiple individuals are presented to the eyewitness at one time, and the eyewitness is asked to choose which, if any, the suspect is. In contrast, sequential lineups present individuals to the eyewitness one at a time, and the eyewitness must accept or reject each one individually as the suspect. These lineups can be Target-Present (TP; the suspect is among those shown to the eyewitness) or Target-Absent (TA; the suspect is not among those shown); the target is the person who is suspected of the crime while the other members of the lineup are referred to as foils. In theory, the best identification, regardless of lineup type, would be a correct selection of the target in TP lineups and correct rejection of the foils in TA lineups. While false identification is lower in TA sequential versus simultaneous lineups, choice accuracy is higher in TP simultaneous versus sequential lineups (Steblay et al., 2001). Additionally, children and older adults have also been found to have higher false identifications in TA lineups, while accuracy is stable across age groups in TP lineups (Pozzulo & Lindsay, 1998).

Another concern regarding lineup construction is how well each lineup member resembles the description of the suspect and one another. Specifically, it has been found that the person who *most* closely resembles the description of the suspect in simultaneous lineups in particular, regardless of whether they are the actual suspect or even a close match, is likely to be chosen because they are considered to be the "best fit" (Doob & Kirshenbaum, 1973; Wells et al., 1993; Wells et al., 1998; Wells, 1984). Additionally, it has been found that the person who stands out in the lineup, or is more unlike the others, will also likely be identified as the suspect, even if they are a foil who is innocent (Clark, 2012; Fitzgerald et al., 2013; Wells, 1984). While foils can be erroneously chosen by having facial hair that others do not or a tattoo that others do not (Badham et al., 2013), they can even be chosen just because they had a different facial expression at the time of identification (Flowe et al., 2014).

The reason for choosing the lineup member that stands out the most can be explained through the *diagnostic feature detection model* (Wixted & Mickes, 2014). Using this model, researchers posit that eyewitnesses give weight to specific features when trying to match the individuals in a lineup to the description of the suspect. This can be helpful when there are shared key or diagnostic features between the lineup member and the suspect (e.g., both have full beards while the rest of the foils only have moustaches), but it can also lead to neglecting the fact that there are other features which are not shared between the two (e.g., two different hair-styles; Colloff et al., 2016).

Given the effect of features that stand out, it is important to increase awareness about avoiding biased lineups (Wells et al., 1998). In practice, there have been law enforcement agencies in both the United States (Wogalter et al., 2004) and the United Kingdom (Zarkadi et al., 2009) that have tried to minimize the effect of *distinct markings*. These markings can commonly appear on a person, specifically on their face, and are not the typical features that make up one's facial configuration. For example, these could be tattoos, moles, scars, bruises, and piercings as opposed to the eyes, nose, and mouth. To avoid confusion with the discussion of distinct facial features (e.g., large doe eyes), these distinct markings will be referred to as *extraneous markings* for the remainder of this paper.

Research in recent years has explored how best to approach the presence of extraneous markings on suspects; to make a lineup fair, it would not be appropriate for the suspect to have features that stand out among the rest. To address this, some have compared the replication of markings across all lineup members to the digital removal of the marking on the target. Zarkadi et al. (2009), for example, presented participants with pictures of faces to study; some of these faces were unaltered and had no extraneous markings, while others had a marking, such as a bruise, mole, piercing, moustache, scar, or tattoo. At test, participants were presented with different arrays of six faces (a simultaneous lineup; in Experiment 1, there were only TP lineups, but in Experiment 2, there were TP and TA lineups). These arrays either replicated the marking on the target across all six faces or digitally removed it from just the target. In Experiment 1, not only was correct identification higher when the markings were replicated instead of removed, but the false alarms to the foils were also lower. In Experiment 2, participants again had better accuracy in the TP lineups, though they were equally likely to misidentify foils (and more likely to misidentify than reject the full lineup) regardless of replication or removal.

Badham et al. (2013), who replicated Zarkadi et al.'s findings (2009), explained that replicating such markings in a TP lineup is better as it creates a fair lineup and provides a better study-test context match for the target face. Indeed, according to those who support the *encoding specificity paradigm*, the recognition and recall of information are better when the context at test matches that in which the information was studied (Tulving & Thomson, 1973). Colloff et al. (2016) also use the diagnostic feature detection model to support such findings, with replication preventing the erroneous weighting of certain diagnostic features at retrieval.

Biometrics

The final topic regarding facial recognition that will be touched on briefly is biometric facial recognition. *Biometrics* are the anatomical or behavioural characteristics that can be used to identify a person, including fingerprints and faces (Jain & Kumar, 2012). Though biometrics can be used in the human-generated identification of people (e.g., fingerprint experts who compare samples manually), they have been instrumental in the growing field of technology-generated identification as well. For example, there are numerous facial recognition programs that have been developed to assist in facial recognition (Jain & Park, 2009). These programs can detect biometric characteristics, including details about one's eyes, nose, mouth, and skin texture, as well as the overall configuration of the face, to compare with pre-existing faces in their databases to make matches. Similar to human facial recognition, some of these programs operate using holistic or global encoding of the faces they encounter (e.g., Belhumeur et al., 1997), while others use feature-based or local encoding (e.g., Penev & Atick, 1996; Wiskott et al., 1997). Also similar to human facial recognition, the programs' ability to accurately detect face matches can be impaired by external factors such as viewpoint, lighting, facial expression, and natural aging of the face (Lee et al., 2012).

Recent developments in technology have been able to expand technology-generated identification to include *soft biometrics*. Soft biometrics are characteristics which, on their own, cannot lead to the exact identification of a person but can assist with the correct identification of them; these can include gender, race, and extraneous facial markings (Arigbabu et al., 2015; Jain & Kumar, 2012). Speaking of extraneous facial markings specifically, facial recognition programs can detect changes in skin texture, shapes, and colours in order to isolate and identify such markings (Becerra-Riera et al., 2018). Due to this, the creative possibilities of different special programs to detect extraneous markings are endless; for example, Lee et al. (2008) created a program solely to identify tattoos and scars, while Choudhury and Mehata (2012) created a program to identify markings covered up by cosmetics. Such programs are also used by law enforcement to help catalogue and later identify such markings (e.g., Datta et al., 2008; Spaun, 2007).

The Current Study

The goal of the current study was to further explore the effect that extraneous markings, such as moles, scars, and tattoos, have on facial recognition. As described above, there has been recent work in both the social-cognitive and computer science and engineering literature involving these types of markings. Specifically, the social-cognitive literature has highlighted the effect that the presence of stand-out features has on eyewitness identification (e.g., Clark, 2012; Wells, 1984). Given this, there have been attempts at minimizing the impact of such features through testing feature replication and feature removal in lineup identifications (e.g., Badham et al., 2013; Zarkadi et al., 2009). In the computer science and engineering literature, the need for technology-generated identification of such features has been highlighted, and programs have been designed to detect these features (e.g., Choudhury & Mehata, 2012; Lee et al., 2008). From a cognitive standpoint, however, the memorial benefit, or hindrance, of such markings on facial recognition is unknown. The extent to which studied faces with these markings are remembered compared to studied faces without these markings, outside of the context of eyewitness lineups and the replication/removal of these markings, is important to explore. Knowing how memorable faces with extraneous markings are can help determine whether the identification of individuals based on these markings can be trusted or if the identification should be treated with caution.

Also important to explore is whether recognition of faces with extraneous markings is accounted for by theories such as the MDS framework (Valentine, 1991; Valentine et al., 2016). As mentioned above, extraneous markings, in the social-cognitive literature, are referred to as distinctive markings, and such markings are thought to make a face stand out, warranting them to be controlled for in eyewitness identification (e.g., Badham et al., 2013). Within the MDS framework, some features can be distinct in comparison to others, and such features do make faces more memorable (e.g., Bartlett et al., 1984; Valentine, 1991; Valentine et al., 2016). Also within the MDS framework, however, the features that are described are typically limited to ones such as the eyes, nose, and mouth (Valentine, 1991; Valentine et al., 2016). The extent to which extraneous markings, such as moles, scars, and tattoos, are represented within face-space is unknown. The dimensions used for the encoding and retrieval of faces in face-space are user-dependent, so depending on the frequency with which extraneous markings are encountered, individuals could very well be able to seamlessly and accurately integrate faces with extraneous markings into face-space.

Across four experiments, the effect of extraneous facial markings on recognition was tested. In Experiment 1, participants studied faces that were either unaltered (with no extraneous markings present), altered with a single mole, or altered with a single scar. They then completed an old-new recognition test with an equal number of new faces in each condition. These two alterations were used as scars and moles can frequently, and somewhat naturally, appear on faces. In Experiment 2, the study procedure was the same as Experiment 1, but two-alternative-forced-choice was used at test. At test, each studied face was presented alongside the same base-face in either of the other two conditions. For example, at study, participants saw face X with a scar present and, at test, they saw face X with a scar alongside face X, unaltered or face X with a mole, and had to choose which was studied. This change in test was made in order to explore whether participants were able to discriminate between the exact face they studied and a slightly changed copy of the studied face.

In Experiment 3, faces with distinct features (e.g., large eyes or eyes that were far apart) were used in place of faces with scars in order to directly compare faces with extraneous markings to faces that are distinct according to the MDS framework. Lastly, in Experiments 4a and b, faces with and without tattoos were compared, with 4b testing recognition using both upright and inverted faces. Tattoos, compared to both moles and scars, are more variable; while each can vary in size and placement on the face, one should be able to discriminate between tattoos better. For example, differentiating between a face with a tattoo of a star and one with a tattoo of an anchor should be different from differentiating between two faces with similar but differently placed moles. Additionally, despite increasing in popularity over the last several decades (Gitnux, 2023; Heywood et al., 2012; Jackson, 2013), many still hold negative beliefs surrounding tattoos, rating people with tattoos more negatively on certain characteristics such as trustworthiness, honesty, and intelligence (e.g., Broussard & Harton, 2018). As such, a polarizing feature might lead to differences that might not occur when using moles and scars, such as the in-group/out-group differences that would be suggested by Hugenberg et al.'s (2010) categorization-individuation model of own-biases, depending on participant views on tattoos.

Overall, it was initially hypothesized that faces with extraneous markings would be remembered better than unaltered faces, similar to traditionally distinct ones. Based on the ideas supported within the social-cognitive literature (e.g., Clark, 2012; Wells, 1984), faces with extraneous markings should stand out and be memorable in comparison to unaltered faces. Applying the MDS framework (Valentine, 1991; Valentine et al., 2016) to faces with extraneous features, the frequency at which moles and scars are encountered in our day-to-day encounters with others should be sufficient to allow for the more precise encoding of these faces in face-space. Additionally, despite the potential that facial tattoos are so polarizing (e.g., Broussard & Harton, 2018) to some participants that out-group biases are formed (e.g., Bäckman, 1991; Cross et al., 1971; Malpass & Kravitz, 1969), leading to reduced recognition of faces with tattoos compared to those without, it was hypothesized that the tattoos themselves would be variable and unique enough that they should benefit facial recognition by allowing for better discriminability between faces.

Chapter 2: Experiment 1 – Old-New Recognition

In Experiment 1, the recognition of unaltered faces was compared to that of faces with extraneous markings, specifically moles and scars. These two markings were explicitly chosen as they are more naturally appearing and could be encountered frequently enough to allow for the development of dimensions in face-space to accommodate them. Participants were presented with faces belonging to the three conditions at study and were later tested using an old-new recognition test. If the presence of extraneous markings is equivalent to distinctive facial features, such as large eyes, making faces with those markings stand out, then there should be an advantage of such markings: Memory should be better for the two extraneous markings conditions than for the unaltered condition.

Method

Participants

Seventy-nine undergraduate students from Memorial University of Newfoundland participated through the Psychology Research Experience Pool (PREP) in exchange for course credit. Data collection was conducted entirely online using the E-Prime Go software (Psychology Software Tools, 2020). Thirty-three participants were tested but excluded due to computer issues (e.g., computer incompatibility with the E-Prime Go software) and failure to return their data files, leaving a final sample size of 46 participants.

Of the 46 participants, 30 identified as female (65.22%), twelve identified as male (26.09%), one identified as non-binary (2.17%), and three chose not to respond. The average age of participants was 21.00 years old (*SD* = 3.52). Twenty-nine participants identified as Caucasian (63.04%), three as South Asian (6.52%), three as East Asian (6.52%), two as Black (4.35%), one as Indigenous (2.17%), and six chose not to respond. Regarding household income, 13 participants reported an average income of less than $20,000 (28.26%), three between $20,000 and $35,000 (6.52%), four between $35,000 and $50,000 (8.70%), five between $50,000 and $75,000 (10.87%), five between $75,000 and $100,000 (10.87%), and five over $100,000 (10.87%); eleven participants chose not to respond. Ethics clearance was received by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) on March 5, 2020[1] (approval number: 20201578-SC; see Appendix A).

Materials

One hundred and twenty different faces from the Chicago Face Database were used (Ma et al., 2015); half were male faces, and half were female faces. All faces were of Caucasian individuals and were chosen based on a limited number of existing facial markings (e.g., existing moles, scars, freckles, and facial hair.). Each face was altered using Adobe Photoshop (Adobe, 2019) so that three versions of each could be used (360 total stimuli): an unaltered face, a face with a single mole, and a face with a single scar (see Appendix B for example stimuli). The moles and scars varied in size and placement on the face so that there were an equal number on the left and right sides and upper and lower halves of each face. The opacity of each alteration on the face varied so that the alteration matched the skin tone of the face. The experiment was programmed through

[1] Ethics clearance is active until March 31, 2024.

the E-Prime 3.0 software (Psychology Software Tools, 2016) and converted to a downloadable E-Prime Go file.

Optional post-experiment and demographic questionnaires were used (see Appendix C and D, respectively). The post-experiment questionnaire probed whether the participant noticed the extraneous marking on some faces and whether they believed such a marking would make a face memorable. The demographic questionnaire asked for basic background information such as gender, age, and ethnicity, as well as whether the participant had any tattoos and their opinions on them. For Experiment 1, each questionnaire was provided as a password-protected Word document. Jamovi (Jamovi Project, 2021) was used for all statistical analyses.

Procedure

Following providing consent through a separate Qualtrics (2022) survey, participants were provided with the E-Prime Go file to download and the password-protected questionnaires; they were given instructions on how to download the file to run the study and told that they would receive the password to the questionnaire document following the study. In the study phase, participants were told to remember faces for an upcoming memory test. Sixty faces were presented at random, one at a time, at the centre of the computer screen for 1500 milliseconds on a grey background; the stimuli took up roughly ¾ of the screen. A screen with a fixation cross in the middle of it was presented following each face stimulus for 500ms. Half of the stimuli were male faces, and half were female faces. Of the 30 male and 30 female faces, one-third (10 of each) were unaltered, one-third had a mole, and one-third had a scar; all faces were unique, with each individual face presented only in one condition for a given participant, and assignment of face to condition was random. The study phase was followed by a three-minute retention task where participants had to differentiate between even and odd numbers.

At test, all 60 studied faces and 60 new faces were presented one by one on grey backgrounds; similar to the studied faces, an equal number of male and female, and unaltered condition, mole condition, and scar condition faces were presented. Participants were asked to indicate whether the face was studied previously or new by pressing the ‘z’ or ‘m’ keys on the keyboard, respectively. Following the experiment, participants were given the password to the questionnaires so they could complete them and then given instructions on returning all files to the experimenter.

Results

Of the 44 participants who responded to the post-experiment questionnaire, all had noticed something unique about some of the faces they studied. Estimates of what percentage of faces had a unique marking ranged from 5-95% (*M* = 62.44%; *SD* = 22.92%). When asked to clarify what exactly was unique about each face, 40 participants had mentioned the presence of scars (90.91%) and 41 had mentioned the presence of moles (93.18%). When asked whether such a marking had made the faces more memorable, 24 participants responded “yes” (54.55%) and 20 responded “no” (45.45%). Of those who responded “yes,” 19 participants mentioned that such a marking stands out and makes a face more distinguishable. Of those who responded “no,” 13 participants mentioned that such a marking was distracting and prevented them from remembering the rest of the face and that the large number of faces shown with the markings made those faces less distinguishable.

From the old-new recognition test, hit and false alarm rates, the signal detection theory (Green & Swets, 1966) measure of sensitivity *d'*[2], the response bias measure *c*, and response times for hits and correct rejections were calculated (see Table 1 for descriptive statistics). Alpha levels of .05 were used for all analyses. A G*Power repeated measures Analysis of Variance (ANOVA) sensitivity analysis (Faul et al., 2009), using this number of participants, power set to .95, and an alpha level of .05, determined that the minimum detectable effect size was .24.

The four main measures were analyzed using separate one-way repeated measures ANOVAs (Condition: Unaltered × Mole × Scar). For the hit rates, the main effect of condition was not significant, $F(2,90) = 1.43$, $MSE = 0.02$, $p = .25$, $\eta^2p = .03$. For false alarm rates, the main effect of condition was significant, $F(2,90) = 5.57$, $MSE = 0.01$, $p = .005$, $\eta^2p = .110$. Follow up uncorrected post-hoc comparisons found no differences in false alarms between the mole and scar conditions, $t(45) = 0.40$, $p = .69$, Cohen's $d = 0.05$, but fewer false alarms to the unaltered condition compared to both the mole, $t(45) = 3.04$, $p = .004$, Cohen's $d = 0.41$, and scar, $t(45) = 2.74$, $p = .01$, Cohen's $d = 0.34$, conditions.

For the sensitivity measure *d'*, the main effect of condition was significant, $F(2,90) = 6.91$, $MSE = 0.24$, $p = .002$, $\eta^2p = .133$. Follow up uncorrected post-hoc

[2] For all *d*' analyses, scores were corrected for perfect hits and no false alarms by subtracting or adding 0.5 from the count of relevant trials, respectively (e.g., 19.5/20 replaced scores of 20/20).

comparisons found no differences in discriminability between the mole and scar conditions, *t*(45) = 1.20, *p* = .24, Cohen's *d* = 0.19, but better discriminability for the unaltered condition compared to both the mole condition, *t*(45) = 3.56, *p* < .001, Cohen's *d* = 0.54, and the scar condition, *t*(45) = 2.40, *p* = .02, Cohen's *d* = 0.36. Lastly, for the response bias measure *c*, the main effect of condition was not significant, *F*(2,90) = 0.15, *MSE* = 0.08, *p* = .86, η^2p = .003.

Table 1

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), Response Bias Measure (c) and Response Times (RT) for Hits and Correct Rejections (CR) by Condition for Experiment 1

Condition	Hit	FA	*d*'	*c*	RT (Hits)	RT (CR)
Unaltered	.63 (.03)	.21 (.02)	1.27 (.11)	.25 (.06)	1384 (102)	1558 (106)
Mole	.58 (.03)	.27 (.02)	0.91 (.09)	.23 (.06)	1520 (134)	1620 (95)
Scar	.60 (.03)	.26 (.02)	1.02 (.09)	.23 (.06)	1469 (92)	1638 (112)

Note: Standard Errors are presented in parentheses below their corresponding means. Response times are in milliseconds.

Response times were analyzed using a 3 (Condition: Unaltered × Mole × Scar) × 2 (Decision Type: Hits × Correct Rejections) repeated measures ANOVA. The main effect of condition was significant, $F(2,90) = 4.02$, $MSE = 64492$, $p = .02$, $\eta^2p = .08$. Follow up uncorrected post-hoc comparisons found no differences in RT's between the mole ($M = 1570$, $SEM = 106$) and scar ($M = 1553$, $SEM = 94$) conditions, $t(45) = 0.42$, $p = .67$, Cohen's $d = 0.02$, but quicker responses for the unaltered ($M = 1471$, $SEM = 97$) condition compared to both the mole condition, $t(45) = 2.50$, $p = .02$, Cohen's $d = 0.13$, and the scar condition, $t(45) = 2.47$, $p = .02$, Cohen's $d = 0.12$. The main effect of decision type was also significant, $F(1,45) = 5.01$, $MSE = 301300$, $p = .03$, $\eta^2p = .10$, with decisions being faster for Hits ($M = 1458$, $SEM = 106$) than Correct Rejections ($M = 1605$, $SEM = 98$). The interaction was not significant, $F(2,90) = 0.41$, $MSE = 94617$, $p = .67$, $\eta^2p = .01$.

In addition to the analyses conducted above, exploratory gender- and race-based analyses were conducted; these can be found in Appendix E and F, respectively. As mentioned previously, both gender (Cross et al., 1971) and race (Malpass & Kravitz, 1969) biases have been found in facial recognition so, given our access to both types of participant information, this has been explored further. However, given that these analyses were secondary to the main goals of the current research, no attempt was made to recruit equal numbers of participants in each group.

Discussion

In Experiment 1, I presented participants with a series of faces that were either altered to include a mole or a scar or were unaltered. I then tested recognition of these faces in an old-new recognition test. I hypothesized that extraneous markings would function as distinct features, leading to more accurate and quicker recognition of faces that include them. In general, my hypothesis was not supported. Not only was recognition for the stimuli with the extraneous markings not more accurate than recognition of the stimuli without, in some cases, it was significantly worse. No differences were found among the conditions in terms of hit rates, but participants were more likely to false alarm to both scar and mole stimuli compared to the unaltered stimuli. Taking both hit and false alarm rates into account, overall performance (discriminability) was better for the unaltered stimuli. Additionally, while there were no differences in response bias between conditions, participants were quicker in their correct identification and rejection of the unaltered stimuli as well.

Initially, my hypothesis was founded on the assumption that encounters with extraneous markings would be so frequent that the recognition of such features could be built into our face-space. As described in detail previously, it is believed that the process of facial recognition occurs within multi-dimensional space (Valentine, 1991; Valentine et al., 2016). Each facial feature that is used to recognize a face becomes stored as a dimension within our own face-space, and faces are encoded into the space by determining how much the features of the face we are encoding differ from the central point of each dimension. Each person's face-space is unique, with the exact dimensions used within face-space differing depending on what is most helpful to the person in differentiating between faces. If a feature or a marking is encountered often enough, it is reasonable to assume that it is possible for a dimension relating to that marking to develop.

Additionally, given the uniqueness of extraneous markings, being something that does not appear on every face (though likely occurring often enough to be incorporated into face-space), it was also assumed that such a marking would act as a *distinct* feature. As described previously, faces with distinct features, such as large eyes or wide mouths, are remembered better than typical faces as they are stored in more isolation in face-space (Valentine, 1991; Valentine et al., 2016). When retrieving a distinct face, memory is better as fewer faces are in close proximity to the target to compete with. When factoring in a dimension within face-space that is dedicated to an extraneous marking, given that not every face has such a marking, the target face should have the advantage of similarly being stored in isolation away from other similar faces.

Given the surprising results, however, I suggest that my assumptions were incorrect. It is possible that our encounters, or at least my participants' encounters, with those with moles and scars, are not frequent enough to warrant a dedicated dimension for that feature in face-space. It is also possible that face-space is strictly limited to internal facial features, such as the eyes, nose, and mouth, with no allowance for other features to be represented within the space at all (i.e., the features are not encoded at all). More likely, however, is that the differentiation between extraneous markings is just not *important* enough to be incorporated into face-space. Valentine et al. (1995) have shown this importance before, particularly through the own-race bias in which face-spaces become tailored to those who are more socially beneficial (important) to remember.

So, it is likely that neither the actual frequency of encounters with extraneous markings nor whether face-space allows for dimensions outside of internal facial features is what led to the detriment in recognition of altered faces found in this experiment. It is likely the lack of individual importance of such markings to later recognition that has prevented the development of the appropriate dimensions from forming. This would then lead to the clustering of altered faces in face-space away from the central point, similar to what is found with other-race faces. This clustering would lead to more false alarms, as was found, as well as lead to slower response times as differentiation would take more time. Despite not observing a difference in hit rates, as is found with other-race faces, I did find worse discriminability overall, further supporting the idea that these faces are being stored in a similar manner to other-race faces.

Expanding outside of the MDS framework, there are several other possibilities to discuss. For one, given that differences were found in false alarm rates and not hit rates, participants could have set different decision criteria at test which was less strict toward altered faces. Participants were more accurate and quick in their decisions for the unaltered faces, which differs in their decisions for the altered faces; though it should be noted that there were no differences in response bias found. Alternatively, extraneous markings might not factor into facial processing at all. As described previously, faces are processed in a holistic manner that is not reliant on individual features (Tanaka & Farah, 1993). Though part-based processing can occur for faces, it is not necessary and is not the key factor in facial recognition abilities. Given this and the rigidity with which holistic processing occurs, it could be possible that the presence of these markings was ignored and was not factored into recognition, leading to the lack of benefit that had been expected. This is not likely the case, however. From the post-experiment questionnaire, I confirmed that participants had noticed the presence of the markings, with almost all mentioning a mole and a scar as unique features in the stimuli they were shown. Participants were also roughly correct, on average, in their estimates of what percentage of faces had a unique feature. So, the features were at least noticed. Moreover, given that the presence of the markings was a detriment to recognition, with significantly better discriminability for the unaltered faces, it appears that the features had some sort of effect on the altered stimuli.

The post-experiment questionnaire responses also suggest another possibility: The extraneous markings were distracting and prevented facial and holistic processing from occurring. Interestingly, participants were split on whether the presence of the unique markings they noticed made the faces more memorable. While almost half of the participants said that faces were more memorable due to the markings because they made them stand out, almost half said the opposite, with some claiming that the markings were distracting and prevented the participant from encoding the rest of this face. The responses of the former here echo the sentiment within the related social-cognitive literature described previously that extraneous markings stand out and must be controlled for when constructing police line-ups (e.g., Zarkadi et al., 2009). The responses of the latter, however, reflect the findings of this experiment. Participants only had a short period of time to encode each face during the study phase. If they were distracted by the markings, they might have been prevented from sufficiently encoding the rest of the face, hindering later recognition.

Distraction interfering with facial processing is not a new concept. As mentioned previously, the mere presence of a weapon while witnessing a crime can negatively impact later identification (Steblay, 1992). The weapon focus effect demonstrates that weapons can draw the attention of eyewitnesses away from the face of the person holding them. Due to this distraction, recognition of the target (suspect) can be poor. The presence of extraneous markings could be acting in a similar manner, drawing the participants' gaze and focus away from the face just as effectively.

Similarly, the inherent nature of the markings themselves could have further hindered recognition of the altered faces. While moles and scars can vary in size, position on the face, and colour, they are not as variable in perceptual details as are other facial features such as eyes. For example, if given two different moles to study, it would be more difficult to differentiate between the two than, say, two different sets of eyes. As mentioned above, while biometrics can help differentiate one person from the next, soft biometrics, such as extraneous markings, are more descriptive and less helpful in the actual identification of an individual (Jain & Kumar, 2012). When adding the lack of differentiating information to the limited amount of time to encode the rest of each face at study, the participants would mostly have to rely on their memory for the marking itself. This reliance would then hinder memory performance for the altered stimuli as a mole alone would not help in telling an old face from a new one so long as that new face had a similar marking that had been seen on other faces. This is supported by the increased false alarms that were observed for the altered faces.

One final comment on the findings of Experiment 1 regards the number of altered stimuli that were used compared to unaltered stimuli. One participant in the post-experiment questionnaire noted the overwhelming number of faces that had extraneous markings. With two-thirds of the stimuli being altered, though the two-thirds were split between two different types of alterations, the impact of the markings could have been muddled by the volume of the markings. When added to the lack of variability among the markings, this could also explain the difference in false alarms that were observed.

Regardless of whether the results of this experiment are explained in terms of the MDS framework or not, several conclusions can be made. First, it appears that the processing of faces is disrupted when other extraneous markings are present. This disruption could be caused by the incompatibility of the markings with face-space, leading to clustering, or it could be caused by a general distraction by the markings, taking away from further facial encoding. Second, it also appears that faces with extraneous markings are not akin to distinct faces. Whereas distinct faces hold an advantage in facial recognition, I only found a detriment to recognition of altered faces compared to unaltered ones.

Chapter 3: Experiment 2 – Two Alternative Forced Choice

In Experiment 2, I wished to further explore the first conclusion made in Experiment 1, specifically that extraneous markings distracted participants from encoding the entirety of the altered faces at study. If participants spent their time looking toward the marking, and not the rest of the face, this would explain the overall difference in discriminability between the altered and unaltered faces as well as why false alarms were higher for the altered faces. In this experiment, I presented participants with a two-alternative-forced-choice test, presenting them with two versions of the same face, as a direct test of whether the entirety of the altered faces was being encoded. It was hypothesized that, based on the results of Experiment 1, participants would have more difficulties in differentiating between different versions of the same face when the target (studied) face was altered because they would not have encoded the whole face well enough to connect the exact marking to each face. Aside from how the recognition of faces was tested, two other major changes were made between Experiments 1 and 2: study duration was increased as test difficulty increased, and confidence ratings were added.

Method

Participants

Fifty-seven undergraduate students from Memorial University of Newfoundland participated through the Psychology Research Experience Pool (PREP) in exchange for course credit. Data collection was conducted entirely online using the E-Prime Go software (Psychology Software Tools, 2020). Twenty-five participants were tested but excluded due to computer issues (e.g., computer incompatibility with the E-Prime Go software) and corrupted data files, leaving a final sample size of 32 participants.

Of the 32 participants, twelve identified as female (37.50%), eight identified as male (25.00%), four identified as non-binary (12.50%), and eight chose not to respond. The average age of participants was 21.28 years old (*SD* = 3.51). Seventeen participants identified as Caucasian (53.13%), three as Black (9.38%), two as Indigenous (6.25%), one as Middle Eastern (3.13%) and nine chose not to respond. Regarding household income, five participants reported an average income of less than $20,000 (15.63%), five between $20,000 and $35,000 (15.63%), one between $35,000 and $50,000 (3.13%), three between $50,000 and $75,000 (9.38%), three between $75,000 and $100,000 (9.38%), and five over $100,000 (15.63%); ten participants chose not to respond.

Materials

The materials used in Experiment 2 were similar to those used in Experiment 1, with a few differences. First, only 48 face stimuli were used at study. Similar to Experiment 1, half of the face stimuli were of male faces and half of female faces. Also similar to Experiment 1, of these faces, there were an equal number of faces in the unaltered, mole, and scar conditions (n = 16). Unlike in Experiment 1, one alternative version of each studied face was additionally used at test; no new faces were used. Each studied face was presented alongside the same face in one of the other two conditions. For example, if the participants saw the unaltered version of face X at study, they would have seen the same face, alongside either the mole or scar version of face X, at test;

which of the two alternative versions was presented was randomized with an equal number of both being presented as lures.

Other differences in materials between Experiment 1 and 2 were that both the E-Prime Go file and Questionnaires were available online. The E-Prime Go file was an URL link instead of a downloadable file that ran the experiment in the participant's computer browser, and the Questionnaires were hosted online via Qualtrics (2022) and were linked to the participant following the experiment.

Procedure

Following providing consent through a separate Qualtrics survey, participants were provided with a link to access the experiment online through E-Prime Go. Similar to Experiment 1, participants were told to remember faces for an upcoming memory test in the study phase. Forty-eight faces were presented, one at a time, at the centre of the computer screen for 3000 milliseconds[3] on a grey background; the stimuli took up roughly ¾ of the screen. A screen with a fixation cross in the middle of it was presented following each face stimulus for 500ms. Half of the stimuli were male faces, and half were female faces. Of the 24 male and female faces, one-third (16) were unaltered, one-third had a mole, and one-third had a scar; all faces were unique, with none being presented in another condition, and assignment of face to condition being random. The study phase was followed by a three-minute retention task where participants had to differentiate between even and odd numbers.

[3] Presentation time was increased from 1500ms in Experiment 1 to 3000ms in Experiment 2 after the pilot testing of 19 participants yielded low overall hit rates (*M* = .53) and high false alarm rates (*M* = .47).

Participants then completed a two-alternative-forced-choice (2AFC) test. All 48 studied (target) faces were presented, one by one, alongside an alternate (distractor) version of the same face, half (8) in each of the other two conditions, assigned at random. Participants were asked to select whether the face on the left side of the screen (by pressing the “z” key) or the right side of the screen (by pressing the “m” key) was the face they had studied. The side of the screen the target face appeared on was counterbalanced. Following each selection, participants were also asked to provide their confidence in their choice by typing a number between 1-6 and pressing the “Enter” key; it was indicated by instructions on the screen that the scale ranged from 1 (not confident) to 6 (confident). Following the experiment, data were automatically uploaded to the E-Prime Go online server, and a link to the Qualtrics survey containing the questionnaires was automatically opened.

Results

Of the 32 participants who responded to the post-experiment questionnaire, all had noticed something unique about some of the faces they studied. Estimates of what percentage of faces had a unique marking ranged from 25-100% (M = 69.77%; SD = 19.71%). When asked to clarify what exactly was unique about each face, 22 participants had mentioned the presence of scars (68.75%) and 23 had mentioned the presence of moles (71.88%). When asked whether such a marking had made the faces more memorable, 19 participants responded “yes” (79.17%) and 5 responded “no” (20.83%). Of those who responded “yes,” 9 participants mentioned that such a marking stands out and makes a face more distinguishable. Of those who responded “no,” 4 participants mentioned that such a marking was distracting and prevented them from remembering the rest of the face and that the large number of faces shown with the markings made those faces less distinguishable.

From the 2AFC test, overall hit and false alarm rates, and scaled measures of sensitivity (*d'*) and response bias (*c*) were calculated[4]; overall response times and confidence ratings for hits were also calculated (see Table 2 for descriptive statistics). Additionally, hit and false alarm rates, and scaled measures of *d'* and *c* were calculated for each target based on the specific distractor condition used (see Table 3 for descriptive statistics). Alpha levels of .05 were used for all analyses. A G*Power post hoc, repeated measures Analysis of Variance (ANOVA) analysis (Faul et al., 2009), using this number of participants, power set to .95, and an alpha level of .05, determined that the minimum detectable effect size was .29. Similar to Experiment 1, exploratory gender- and race-based analyses can be found in Appendix E and F, respectively, though, again, no attempt was made to recruit equal numbers of participants in each group given that these were secondary to the main goals.

For the overall analyses, all six measures were analyzed using separate one-way repeated measures ANOVAs (Condition: Unaltered × Mole × Scar). For the hit rates, the main effect of condition was significant, $F(2,62) = 3.06$, $MSE = 0.04$, $p = .05$, $\eta^2p = .09$. Follow up uncorrected post-hoc comparisons found no differences in hits between the

[4] For 2AFC, only scores one side of the participants' responses are used in calculating hits and false alarms (Macmillan & Creelman, 2005); we chose the responses to the left stimuli (when the participant pressed the "z" key). For these corrected measures, an accuracy of 1 would be considered a hit and an accuracy of 0 would be considered a false alarm. Additionally, the formulas for the *d*' and *c* calculations were those used by Macmillan and Creelman.

mole and scar conditions, $t(31) = 0.08$, $p = .94$, Cohen's $d = 0.02$, higher hits to the unaltered condition compared to the scar condition, $t(31) = 2.15$, $p = .04$, Cohen's $d = 0.57$, and marginally higher hits to the unaltered condition compared to the mole condition, $t(31) = 1.98$, $p = .06$, Cohen's $d = 0.58$.

Table 2

Mean Hit and False Alarm (FA) Proportions, Scaled Measures of Sensitivity (d') and Response Bias (c), and Response Time (RT) and Confidence for Hits by Condition for Experiment 2

Condition	Hits	FA	*d*'	*c*	RT (Hits)	Confidence
Unaltered	.66 (.03)	.34 (.04)	0.68 (.13)	.004 (.06)	4646 (432)	3.88 (0.19)
Mole	.55 (.03)	.40 (.04)	0.32 (.13)	.09 (.05)	4573 (385)	3.65 (0.19)
Scar	.55 (.03)	.48 (.03)	0.14 (.11)	-.05 (.05)	4764 (330)	3.71 (0.20)

Note: Standard Errors are presented in parentheses below their corresponding means.

For the false alarm rates, the main effect of condition was significant, $F(2,62) = 4.10$, $MSE = 0.04$, $p = .02$, $\eta^2p = .12$. Follow up uncorrected post-hoc comparisons found no differences in false alarms when the target was unaltered or had a mole, $t(31) = 1.00$, $p = .33$, Cohen's $d = 0.27$, fewer false alarms when the target was unaltered compared to when it had a scar, *t*(31) = 3.05, *p* = .005, Cohen's *d* = 0.73, and marginally fewer false alarms when the target had a mole compared to when it had a scar, *t*(31) = 1.88, *p* = .07, Cohen's *d* = 0.45.

For the scaled sensitivity measure *d'*, the main effect of condition was significant, *F*(2,62) = 4.79, *MSE* = 0.51, *p* = .01, η^2p = .13. Follow up uncorrected post-hoc comparisons found no differences in discriminability between the mole and scar conditions, *t*(31) = 1.12, *p* = .27, Cohen's *d* = 0.27, better discriminability for the unaltered condition compared to the scar condition, *t*(31) = 3.20, *p* = .003, Cohen's *d* = 0.81, and marginally better discriminability for the unaltered condition compared to the mole condition, *t*(31) = 1.80, *p* = .08, Cohen's *d* = 0.51. Additionally, for the corrected response bias measure *c*, the main effect of condition was not significant, *F*(2,62) = 1.64, *MSE* = 0.10, *p* = .20, η^2p = .05.

Lastly, there was no main effect of condition for the response times for hits, *F*(2,62) = 0.08, *MSE* = 3.64 x 10^6, *p* = .92, η^2p = .003. For the confidence ratings, however, there was a marginally significant main effect of condition, *F*(2,62) = 2.99, *MSE* = 0.15, *p* = .06, η^2p = .09. Follow up uncorrected post-hoc comparisons found no differences in confidence between the mole and scar conditions, *t*(31) = 0.59, *p* = .56, Cohen's *d* = 0.05, higher confidence for the unaltered condition compared to the mole condition, *t*(31) = 2.39, *p* = .02, Cohen's *d* = 0.21, and marginally higher confidence for the unaltered condition compared to the scar condition, *t*(31) = 1.80, *p* = .08, Cohen's *d* = 0.16.

Comparisons for each measure based on the specific distractor each target was presented alongside were analyzed using separate paired samples *t*-tests. For the hit rates, no differences were found between the unaltered targets regardless of whether the distractor was belonged to the mole or scar conditions, $t(31) = 1.09$, $p = .29$, Cohen's $d = .19$, nor were there differences found between the mole targets based on distractor type, $t(31) = 0.48$, $p = .64$, Cohen's $d = 0.08$. There were, however, marginally greater hits to the scar targets presented alongside distractors in the mole condition compared to the unaltered condition, $t(31) = 1.83$, $p = .08$, Cohen's $d = 0.32$. For the false alarm rates, no differences were found between the unaltered targets based the distractor type, $t(31) = 0.225$, $p = .82$, Cohen's $d = .04$, the mole targets based on distractor type, $t(31) = 1.34$, $p = .18$, Cohen's $d = 0.24$, nor the scar targets based on distractor type, $t(31) = 1.37$, $p = .18$, Cohen's $d = 0.24$.

Table 3

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Distractor in Experiment 2

Target	Distractor	Hit	FA	*d*'	*c*
Unaltered	Mole	.66	.37	0.61	-.05
		(.04)	(.04)	(.13)	(.06)
	Scar	.62	.36	0.53	.04
		(.03)	(.04)	(.13)	(.06)
Mole	Unaltered	.54	.38	0.32	.14
		(.05)	(.04)	(.15)	(.07)
	Scar	.56	.44	0.24	.001
		(.04)	(.04)	(.12)	(.06)
Scar	Unaltered	.52	.50	0.04	-.03
		(.04)	(.04)	(.13)	(.07)
	Mole	.62	.41	0.42	-.05
		(.04)	(.04)	(.11)	(.07)

Note: Standard Errors are presented in parentheses below their corresponding means.

For the sensitivity measure *d'*, no differences were found between the unaltered targets regardless of the distractor, $t(31) = 0.46$, $p = .65$, Cohen's $d = .08$, nor were there differences found between the mole targets regardless of distractor, $t(31) = 0.49$, $p = .63$, Cohen's $d = 0.09$. There was, however, better discriminability for the scar targets presented alongside distractors in the mole condition compared to the unaltered condition, $t(31) = 2.26$, $p = .03$, Cohen's $d = 0.40$. For response bias, no differences were found between the unaltered targets regardless of the distractor, $t(31) = 1.07$, $p = .30$, Cohen's $d = .19$, the mole targets regardless of distractor, $t(31) = 1.36$, $p = .18$, Cohen's $d = 0.24$, nor the scar targets regardless of distractor, $t(31) = 0.15$, $p = .88$, Cohen's $d = 0.03$.

Discussion

In Experiment 2, I presented participants with a series of faces that were either altered to include a mole or a scar, or were unaltered. I then tested recognition of these faces in a 2AFC test. I hypothesized that discriminating between two different versions of the same face would be more difficult when the target was from one of the altered conditions. In general, my hypothesis was supported. Performance was better overall when the target stimuli were unaltered. Specifically, for the unaltered targets, hits were higher, false alarms were lower, and discriminability was better. Confidence was also higher for the unaltered targets, with no difference in response times or response biases between conditions. Additionally, when comparing performance within each condition based on which distractor was used, there were no differences between distractors for either the unaltered or mole conditions. There was, interestingly, better performance for scar targets when the mole condition was the distractor rather than when the unaltered condition was used.

These results provide further support for the conclusions made in Experiment 1. The presence of facial markings does disrupt facial processing, hindering subsequent recognition performance for the altered faces. Specifically, it appears to be that this disruption is caused by the distraction these markings cause. Despite having twice as long to study the faces presented to them, participants still performed worse when the studied face had an extraneous marking than when unaltered. If the participants were adequately encoding the entirety of the face at study, they should have been able to correctly differentiate between the conditions at test, knowing whether the specific face presented to them originally had a mole, a scar, or neither. If the participants were being distracted by the mole or scar, then differentiation at test would be more difficult because they would not have made the connection between the face itself and which marking was present. For the studied faces with no markings, however, the whole face would have been encoded, and participants would have been able to correctly remember at test that person X, for example, did not have any other markings.

Connecting back to what was discussed in Experiment 1, it suggests that the hindered performance for the altered stimuli was not predominantly due to the lack of variability between markings. Compared to the first experiment, participants in this experiment did not have to differentiate between old and new faces. So, they would not have been hindered by whether they remembered the exact mole or scar from one face to the next or how different the marking was from one face to the next. The main task in this experiment was to figure out whether person X had a scar, a mole, or neither, which resulted in poorer performance when the target was in one of the altered conditions. Regardless of whether they remembered the exact marking, so long as they remembered which type of marking was on which face, their performance would not have suffered.

That is not to say that the lack of variability between markings did not affect recognition performance in the first experiment, but the results from both experiments thus far are consistent with the idea that extraneous markings distract from and disrupted holistic facial processing at encoding.

While distraction causing disruption in processing does seem to fit with my results thus far, it does not mean that my results cannot be explained through an MDS framework (Valentine, 1991; Valentine et al., 2016). In Experiment 1, it was discussed that if faces with extraneous markings act similarly to other-race faces within our face-space, the faces would be stored in a cluster, hindering attempts at retrieving faces from that area of face-space. This, too, can be applied to the results of this experiment. Studied unaltered faces would have been stored within face-space according to the correct dimensions present within the space. At retrieval, there would not be confusion as to whether the face itself was the target (unaltered) or whether it contained one of the markings because of how clearly the face was stored. For the altered targets, however, increased noise around the target face in face-space could lead to difficulties in determining the exact condition to which each face belonged. Even if only one version of each face was stored, the chance of an error occurring at encoding and retrieval increases because of the reduced encoding accuracy of each caused by disrupted holistic processing. That being said, it would also be expected that, due to difficulties such clustering would cause, there should also be a difference in response times for each condition, as observed in Experiment 1. No such difference was found in this study, though this could be due to the difference in the type of recognition tests used between the two studies.

One last issue to discuss regarding Experiment 2 is the differences found in performance based on distractor type. Specifically, while performance for the unaltered and mole conditions did not differ depending on distractor type, performance for the scar condition did. Interestingly, performance was better when the distractor was the other alteration, the mole, than when there was no alteration. These differences cannot be caused by a general preference to choose the unaltered stimuli, regardless of whether that face was the target or not, because the same differences were not found when the target stimuli had moles. One possibility is that there was a preference in choosing the unaltered faces, but only when the target contained a scar.

Chapter 4: Experiment 3 – Distinct Faces

In Experiment 3, I wished to further explore the second conclusion made in Experiment 1, specifically that faces with extraneous markings are not processed in the same way that distinct faces are, leading to reduced, rather than improved memory performance. In Experiment 3a, I had participants rate altered and unaltered faces based on distinctiveness to obtain *distinct* stimuli that could be used in Experiment 3b. In Experiment 3b, I presented participants with altered, unaltered, and distinct faces and tested them using an old-new recognition test. It was hypothesized that, based on the results of Experiment 1, recognition performance would be best for the distinct condition, then the unaltered condition, then the mole condition.

Experiment 3a

Method

Participants. Fifty participants were recruited through the online data collection website, Prolific (2022) in exchange for the minimum per-hour pay rate. No inclusion/exclusion criteria were used, and no demographic information was obtained. No data were excluded.

Materials. In addition to the 60 faces of each sex used in Experiment 1, 12 other faces of each sex were used. For each of these 144 (72 male and 72 female) faces, norming data from the Chicago Face Database supplementary Norming Data and Codebook file (Ma et al., 2015) was obtained. Specifically, this file provided pre-rated data on a large number of characteristics of each face, including how unusual the face looked (i.e., how much would each face stand out in a crowd rated on a scale of 1 [not likely] -7 [extremely likely]); this is a common method of rating facial distinctiveness (e.g., Valentine & Bruce, 1986b).

Given the low mean ratings of distinctiveness for both male (*M* = 2.39, *SD* = 0.54, Range = 1.3 – 4.29) and female (*M* = 2.27, *SD* = 0.46, Range = 1.5 – 3.88) unaltered faces in the database, the 36 highest pre-rated distinctive faces of each sex were altered using the Photoshop Liquify filter (Adobe, 2019) to increase facial distinctiveness. For each face, one of four transformations was made: distance between eyes increased, distance between eyes decreased, mouth widened, or mouth narrowed. For the change to eye distance, the Eye Size and Eye Distance within Photoshop Liquify were set to either 100 (for increased distance) or -100 (for decreased distance). For the change to mouth width, the Upper and Lower Lip and Mouth Width and Height within Photoshop Liquify were set to either 100 (for widened mouth) or -100 (for narrowed mouth).

Each altered and unaltered face was used in a distinctiveness ratings task conducted through Qualtrics (2022). The rating instructions were based on other similar distinctiveness tasks, namely Valentine and Bruce (1986b), and were as follows:

> We would like you to imagine that you must meet each person at a railway station and rate each face for how easy it would be to spot them in a crowd. A face that is very distinctive (or unusual) and so would be relatively easy to spot in a crowd should be rated 7, a typical face that is difficult to identify in a crowd should be rated 1.

Ratings were conducted on a 7-pt Likert scale ranging from 1 (Extremely Difficult) to 7 (Extremely Easy) with lower ratings indicating a more typical face that does not stand out.

Procedure. Upon signing up for the study through Prolific, participants were able to open a link to a Qualtrics survey which contained the consent form and the rating task. Following consent, participants were informed that they would be shown faces, one by one, and would have to rate them based on how easy they would be to spot in a crowd. Faces were presented at the top left of the screen on a white background, taking up roughly 1/4th of the screen, and the rating scale was presented immediately below the picture, taking up roughly ½ the screen. Half of the faces shown were of men and half were of women, and half of the faces were unaltered, while half were altered with a distinct facial feature. While the presentation order of the faces was random, individual faces were not randomly assigned to condition. Following completion of the rating task, participants were thanked for their time.

Results

The distinctiveness ratings for each face were averaged across participants to obtain an average distinctiveness score (see Appendix G). The 16 highest rated faces for both males and females were used as the stimuli in the distinct condition in Experiment 3b. These faces were all stimuli altered to be more facially distinct. The 16 lowest rated faces for both males and females were used as the stimuli in the unaltered condition in Experiment 3b. These faces were all stimuli that were unaltered. Paired samples *t*-tests determined that distinctiveness ratings did differ between the male distinct and unaltered faces, $t(15) = 28.50$, $p < .001$, Cohen's $d = 7.11$, and the female distinct and unaltered faces, $t(15) = 21.90$, $p < .001$, Cohen's $d = 5.47$. Additionally, the 16 middle rated faces for both males and females were used to help select which mole altered stimuli[5] would be used in the mole condition in Experiment 3b.

Experiment 3b

Method

Participants. Thirty-eight undergraduate students from Memorial University of Newfoundland participated through the Psychology Research Experience Pool (PREP) in exchange for course credit. Data collection was conducted both in-person ($n = 13$) and online ($n = 25$)[6] using the PsychoPy (Peirce et al., 2019) online extension, Pavlovia (Open Science Tools, 2022). Four online participants were tested but excluded due to corrupted data files, leaving a final sample size of 34 participants.

Of the 34 participants, 24 identified as female (70.59%), three identified as male (8.82%), three identified as non-binary (8.82%), and four chose not to respond. The average age of participants was 20.23 years old ($SD = 1.91$). Twenty-three participants identified as Caucasian (67.65%), three as Black (8.82%), one as Indigenous (2.94%), one as South Asian (2.94%) and six chose not to respond. Regarding household income, four participants reported an average income of less than $20,000 (11.76%), two between $20,000 and $35,000 (5.88%), one between $35,000 and $50,000 (2.94%), two between

[5] Mole stimuli were chosen over the scar stimuli for the altered condition in Experiment 3b given the high false alarm rates and low discriminability of the scar stimuli in Experiment 2. Also note that the mole stimuli themselves were not rated on distinctiveness.

[6] Given that there were no meaningful interactions when comparing data from in-person and online participants (see below), all data in the in-text reported analyses were collapsed over testing condition.

$50,000 and $75,000 (5.88%), five between $75,000 and $100,000 (14.71%), and 14 over $100,000 (41.18%); six participants chose not to respond.

Materials. The study materials were similar to those used in Experiment 1, with a few exceptions. First, instead of 120 faces, only 96 were used. Second, instead of faces altered with scars, some faces were altered to be *distinct*; the process of creating and selecting these stimuli is described in Experiment 3a (see Appendix B for example stimuli). Experiment 3b was also programmed using PsychoPy (Peirce et al., 2019) instead of E-Prime. For the online participants, the experiment was conducted through Pavlovia. Additionally, all post-experiment and demographic questionnaires were conducted through Qualtrics.

Procedure. Following consent through a separate Qualtrics survey, the online participants were provided with a link to access the experiment online through Pavlovia; the in-person participants were brought into the laboratory with the experiment running on the computer screen before them. Similar to Experiment 1, participants were told to remember faces for an upcoming memory test in the study phase. Forty-eight faces were presented, one at a time, at the centre of the computer screen for 1500 milliseconds on a grey background; the stimuli took up roughly ¾ of the screen. A screen with a fixation cross in the middle of it was presented following each face stimulus for 500ms. Half of the stimuli were male faces, and half were female faces. Of the 48 faces (24 male and 24 female), one-third (16) were unaltered, one-third had a mole, and one-third were altered to have distinct facial features; all faces were unique, with none being presented in another condition. Though the presentation order was random, assignment to condition was not random. The study phase was followed by a three-minute retention task where participants had to differentiate between even and odd numbers.

At test, all 48 studied faces and 48 new faces were presented one by one on grey backgrounds; similar to the studied faces, an equal number of male and female, and unaltered condition, mole condition, and distinct condition faces were presented. Participants were asked to indicate whether the face was studied previously or new by pressing the ‘z’ or ‘m’ keys on the keyboard, respectively. Following each selection, participants were also asked to provide their confidence in their choice by typing a number between 1-6 and pressing the “Enter” key; it was indicated by instructions on the screen that the scale ranged from 1 (not confident) to 6 (confident). Following the experiment for the online participants, data were automatically uploaded to Pavlovia, and a link to the Qualtrics survey containing the questionnaires was automatically opened. Following the experiment for the in-person participants, data were automatically uploaded to the local server, and a link to the Qualtrics survey containing the questionnaires was automatically opened.

Results

Of the 30 participants who responded to the post-experiment questionnaire, all had noticed something unique about some of the faces they studied. Estimates of what percentage of faces had a unique marking ranged from 15-90% (M = 71.81%; SD = 17.70%). When asked to clarify what exactly was unique about each face, 21 participants had mentioned the presence of moles (70.00%) and 29 had mentioned differences in eyes/mouth sizes (96.67%). When asked whether such a marking had made the faces more memorable, 25 participants responded "yes" (83.33%) and 5 responded "no" (16.67%). Of those who responded "yes," 14 participants mentioned generally that the different unique features and markings stand out and make a face more distinguishable while 6 participants specifically mentioned that the exaggerated facial features (i.e., distinct stimuli) made the face more distinguishable. Of those who responded "no," 4 participants mentioned that a common marking such as the mole was distracting and prevented them from remembering the rest of the face, though they did mention that the exaggerated facial features were not as distracting.

From the old-new recognition test, overall hit and false alarm rates, a measure of sensitivity *d'*, and the response bias measure *c* were calculated (see Table 4 for descriptive statistics). Additionally, overall response times and confidence ratings for hits and correct rejections were also calculated (see Table 5 for descriptive statistics). Alpha levels of .05 were used for all analyses. A G*Power post hoc, repeated measures Analysis of Variance (ANOVA) analysis (Faul et al., 2009), using this number of participants, power set to .95, and an alpha level of .05, determined that the minimum detectable effect size was .28. Exploratory gender- and race-based analyses can be found in Appendix E and F, respectively, though sample sizes were again not controlled for.

The four main measures were analyzed using separate one-way repeated measures ANOVAs (Condition: Unaltered × Mole × Distinct). For the hit rates, the main effect of condition was significant, $F(2,66) = 13.80$, $MSE = 0.02$, $p < .001$, $\eta^2p = .30$. Follow up uncorrected post-hoc comparisons found no differences in hits between the mole and unaltered conditions, $t(33) = 1.55$, $p = .13$, Cohen's $d = 0.27$, but more hits to the distinct condition compared to both the mole, *t*(33) = 4.87, *p* < .001, Cohen's *d* = 0.84, and unaltered, *t*(33) = 3.56, *p* = .001, Cohen's *d* = 0.61, conditions.[7]

For false alarm rates, the main effect of condition was significant, *F*(2,66) = 4.33, *MSE* = 0.02, *p* = .02, η^2p = .12. Follow up uncorrected post-hoc comparisons found no differences in false alarms between the mole and unaltered conditions, *t*(33) = 0.49, *p* = .63, Cohen's *d* = 0.08, but fewer false alarms to the distinct condition compared to both the mole, *t*(33) = 2.51, *p* = .02, Cohen's *d* = 0.43, and unaltered, *t*(33) = 2.77, *p* = .01, Cohen's *d* = 0.48, conditions.[8]

For the sensitivity measure *d'*, the main effect of condition was significant, *F*(2,66) = 21.5, *MSE* = 0.25, *p* < .001, η^2p = .40. Follow up uncorrected post-hoc comparisons found no differences in discriminability between the mole and unaltered conditions, *t*(33) = 0.35, *p* = .73, Cohen's *d* = 0.06, but better discriminability for the distinct condition compared to both the mole, *t*(33) = 5.63, *p* < .001, Cohen's *d* = 0.97, and unaltered, *t*(33) = 5.60, *p* < .001, Cohen's *d* = 0.96, conditions.[9] Lastly, for the

[7] A 3 (Condition: Unaltered × Mole × Distinct) x 2 (Testing Condition: In-Person × Online) mixed measures ANOVA was also conducted to test whether testing condition had an effect on hit rates; there was no effect of testing condition, *F*(1,32) = 0.98, *MSE* = 0.04, *p* = .33, η^2p = .03, and no interaction, *F*(2,64) = 1.04, *MSE* = 0.02, *p* = .36, η^2p = .03.

[8] A 3 (Condition: Unaltered × Mole × Distinct) x 2 (Testing Condition: In-Person × Online) mixed measures ANOVA was also conducted to test whether testing condition had an effect on false alarm rates; FAs were higher for participants tested online (*M* = .30, *SEM* = .02) compared to in-person (*M* = .22, *SEM* = .03), *F*(1,32) = 4.86, *MSE* = 0.04, *p* = .04, η^2p = .13, but there was no interaction, *F*(2,64) = 1.02, *MSE* = 0.02, *p* = .37, η^2p = .03.

[9] A 3 (Condition: Unaltered × Mole × Distinct) x 2 (Testing Condition: In-Person × Online) mixed measures ANOVA was also conducted to test whether testing condition had an effect on *d*'; there was no effect of testing condition, *F*(1,32) = 1.88, *MSE* = 0.55, *p* = .18, η^2p = .06, but the interaction was significant, *F*(2,64) = 3.26, *MSE* = 0.23, *p* = .05, η^2p = .09. Follow up Tukey post-hoc comparisons showed no meaningful differences.

response bias measure *c*, the main effect of condition was not significant, $F(2,66) = 1.36$, $MSE = 0.12$, $p = .27$, $\eta^2p = .004$.[10]

Table 4

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), Response Bias Measure (c) by Condition for Experiment 3

Condition	Hits	FA	*d'*	*c*
Unaltered	.60	.30	0.88	.16
	(.02)	(.03)	(.09)	(.07)
Mole	.56	.29	0.83	.25
	(.03)	(.03)	(.11)	(.08)
Distinct	.72	.21	1.54	.11
	(.03)	(.02)	(.11)	(.06)

Note: Standard Errors are presented in parentheses below their corresponding means.

Response times and confidence ratings were analyzed using separate 3 (Condition: Unaltered × Mole × Scar) × 2 (Decision Type: Hits × Correct Rejections) repeated measures ANOVA. For the response times, the main effect of Condition was significant, $F(2,66) = 6.97$, $MSE = 783528$, $p = .002$, $\eta^2p = .17$. Follow up uncorrected post-hoc comparisons found no differences in RTs between the mole ($M = 2677$, $SEM = 193$) and unaltered ($M = 2876$, $SEM = 263$) conditions, $t(33) = 1.11$, $p = .28$, Cohen's d =

[10] A 3 (Condition: Unaltered × Mole × Distinct) x 2 (Testing Condition: In-Person × Online) mixed measures ANOVA was also conducted to test whether testing condition had an effect on *c*; there was a higher (more conservative) response bias for the in-person participants ($M = .29$, $SEM = .07$) compared to the online participants ($M = .09$, $SEM = .06$), $F(1,32) = 4.36$, $MSE = 0.24$, $p = .05$, $\eta^2p = .12$, but there was no interaction, $F(2,64) = 0.66$, $MSE = 0.12$, $p = .52$, $\eta^2p = .02$.

0.13, but quicker responses for the distinct (*M* = 2317, *SEM* = 153) condition compared to both the mole condition, $t(33) = 3.10$, $p = .004$, Cohen's $d = 0.32$, and the unaltered condition, $t(33) = 3.67$, $p < .001$, Cohen's $d = 0.42$. The main effect of decision type was not significant, $F(1,33) = 0.41$, $MSE = 665530$, $p = .52$, $\eta^2 p = .01$, nor was the interaction, $F(2,66) = 0.30$, $MSE = 520199$, $p = .74$, $\eta^2 p = .01$.

For the confidence ratings, the main effect of condition was significant, $F(2,66) = 17.13$, $MSE = 0.21$, $p < .001$, $\eta^2 p = .34$. Follow up uncorrected post-hoc comparisons found higher confidence for the mole (*M* = 4.19, *SEM* = .12) than the unaltered (*M* = 4.04, *SEM* = .12) condition, $t(33) = 2.27$, $p = .03$, Cohen's $d = 0.18$, as well as higher confidence for the distinct (*M* = 4.50, *SEM* = .12) condition compared to both the mole condition, $t(33) = 3.42$, $p = .002$, Cohen's $d = 0.36$, and the unaltered condition, $t(33) = 5.58$, $p < .001$, Cohen's $d = 0.54$. The main effect of decision type was also significant, $F(1,33) = 18.51$, $MSE = .62$, $p < .001$, $\eta^2 p = .01$, with higher confidence for the hits (*M* = 4.48, *SEM* = .11) than the correct rejections (*M* = 4.00, *SEM* = .14). The interaction was non-significant, $F(2,66) = 0.26$, $MSE = .22$, $p = .77$, $\eta^2 p = .01$.

Table 5

Response Times and Confidence Ratings for Hits and Correct Rejections by Condition for Experiment 3

	Response Times (ms)		Confidence	
	Hits	Correct Rejection	Hits	Correct Rejection
Unaltered	2967	2786	4.25	3.84
	(278)	(293)	(.13)	(.15)
Mole	2674	2680	4.43	3.94
	(198)	(239)	(.13)	(.14)
Distinct	2340	2295	4.75	4.24
	(146)	(170)	(.14)	(.15)

Note: Standard Errors are presented in parentheses below their corresponding means.

Discussion

In Experiment 3, I presented participants with a series of faces that were either altered to include a mole, altered to be distinct (i.e., with exaggerated facial features) or were unaltered. I then tested recognition of these faces in an old-new recognition test. I hypothesized that recognition would be best for the distinct stimuli, then the unaltered stimuli, then the mole stimuli. My hypothesis was partially supported. Performance was better overall for the distinct stimuli but similar for the unaltered and mole stimuli. Specifically, for the distinct stimuli, there were more hits, fewer false alarms, better discriminability, quicker response times, and higher confidence compared to the other conditions. The only differences found between the unaltered and mole conditions were in the confidence ratings, in which participants were more confident in their decisions regarding the mole stimuli over the unaltered stimuli.

Our hypothesis was supported regarding performance being best for the distinct stimuli. As has been described previously, memory is better for these types of faces as they hold an advantage in face-space (Valentine, 1991; Valentine et al., 2016). The distinctiveness advantage is quite robust (e.g., Bartlett et al., 1984; Cohen & Carr, 1975; Going & Read, 1974; Winograd, 1981), and this finding supports the effectiveness of my manipulation of those stimuli. Not supported, however, was my prediction that there would also be an advantage for the unaltered stimuli over the altered (mole) ones. This prediction was based on the findings of Experiments 1 and 2, in which the advantage was found.

One plausible reason for the lack of benefit to the unaltered stimuli over mole stimuli in this experiment was discussed in Experiment 1. Specifically, one participant in the first experiment discussed the overwhelming number of altered faces that were presented. With two-thirds of the faces shown in that experiment having a marking, the distinctiveness that could have been caused by the markings would have been reduced and possibly created a paradoxical distinctiveness of the unaltered faces. This could not only prevent a benefit to memory for the altered faces but could also create a detriment to memory for the faces. In this experiment, compared to the first two, there were an equal number of unaltered and mole faces. If the sheer volume of faces with extraneous markings overshadowed the individual impact of the markings in Experiments 1 and 2, then this would explain the lack of difference found in the current experiment. It should be noted, however, that there still was no benefit to the markings being present in this experiment; there just was no detriment.

A different explanation could be that the presence of the distinct stimuli has mitigated the distraction that the extraneous markings cause. In the post-experiment questionnaire, most of the participants stated that the uniqueness of some of the faces made them more memorable; of these participants, most specifically mentioned the distinct/exaggerated features and not the mole as causing the memorability. Similarly, of the few who said that the faces were not more memorable due to their uniqueness, they specifically mentioned that the mole was distracting and did not mention the exaggerated features. If participants noticed the exaggerated features and found them to be helpful in remembering the faces, they could have adjusted their study strategy to look for these beneficial features on each new study trial, regardless of which trial it was. This would, unknowingly, ensure that some internal facial features were encoded and lead to better performance at test because the moles would not have distracted from encoding to the same extent as they would have without a change in study strategy. It should be noted that it is known that the participants noticed that some of the faces had moles on them, as moles were explicitly mentioned in the post-experiment questionnaire when asked about what made the faces unique. Nevertheless, it is likely that these markings were noticed after part of the face has already been encoded.

Another possibility based on the MDS framework (Valentine, 1991; Valentine et al., 2016) could be that my choice of stimuli resulted in the clustering of both unaltered and mole stimuli in face-space. As described in Experiment 3a, for the unaltered stimuli, I chose those that were rated as the most average in comparison to the distinct stimuli to differentiate the two. As described previously, typical faces with average dimensions of facial features are clustered around the central point in face-space (Valentine, 1991; Valentine et al., 2016). This clustering, like with other-race faces, results in lower hits and higher false alarms, a mirror effect, for typical faces. In the previous experiments, despite the Chicago Face Database norming data typicality ratings indicating the stimuli were relatively typical faces, the stimuli used still had a larger range of typicality ratings than in the current study. By choosing the stimuli in the manner I did in this experiment, I may have equated the density of clustering in face-space between the unaltered and mole conditions, resulting in a similar performance for the two conditions.

Regardless of the above, it does appear that my conclusions from Experiment 1 hold. The presence of extraneous markings likely disrupts facial processing, either by distracting from encoding the face or by creating a cluster of faces in face-space that is hard to sort through. However, when participants are presented with faces with memorable internal features, the detrimental effect of the markings should be mitigated due to the different viewing patterns of the participant. Additionally, extraneous markings do not provide the same benefit as distinct faces do, with different effects on facial processing being caused by each.

Chapter 5: Experiment 4 – Face Tattoos

In Experiment 4, faces with and without tattoos were used to address concerns over the lack of variability between the extraneous markings used in the previous experiments. As mentioned, while the moles and scars varied slightly between faces, the lack of variability could have hindered memory for faces with those markings. The usage of tattoos should mitigate the effects of variability as they can differ to a greater extent than can other types of markings. Additionally, these types of markings are not naturally occurring (unlike moles and scars) and could provide different insights into memory for different types of extraneous markings. In Experiment 4a, participants studied faces with and without tattoos and were tested using an old-new recognition test. Despite the possibility that participants could hold strong or negative opinions about tattoos, creating an out-group bias for the tattooed stimuli and negatively impacting recognition, it was hypothesized that memory would actually be more accurate for the faces with tattoos as the benefit of the variability of the tattoos would outweigh biases and increase recognition for those faces. Following the findings of Experiment 4a, Experiment 4b replicated the study phase of 4a, but tested faces as either upright or inverted in an old-new recognition test to explore whether the presence of the extraneous markings was causing a different type of encoding to occur. Participants were also asked questions regarding their opinions on tattoos and their own tattoos to monitor any biases present.

Experiment 4a

Method

Participants. Thirty-two undergraduate students from Memorial University of Newfoundland participated through the Psychology Research Experience Pool (PREP) in exchange for course credit. Data collection was conducted in-person and no participants' data had to be excluded.

Of the 32 participants, 24 identified as female (75.00%), seven identified as male (21.88%), and one chose not to respond. The average age of participants was 21.10 years old (*SD* = 2.21). Twenty participants identified as Caucasian (62.50%), four as Black (12.50%), two as East Asian (6.25%), two as Middle Eastern (6.25%), one as South Asian (3.13%), and three chose not to respond. Regarding household income, eight participants reported an average income of less than $20,000 (25.00%), two between $20,000 and $35,000 (6.25%), two between $35,000 and $50,000 (6.25%), two between $50,000 and $75,000 (6.25%), six between $75,000 and $100,000 (18.75%), and ten over $100,000 (31.25%); two participants chose not to respond.

Materials. Experiment 4a used a new set of stimuli. To ensure the realism of the tattoos, images of Caucasian individuals with face tattoos were taken from Google Images and modified using Adobe Photoshop (Adobe, 2019; see Appendix H for sources). Images were sought using the search terms "faces with tattoos," "male faces with tattoos," "men with facial tattoos," "female faces with tattoos," and "women with facial tattoos." Faces were chosen based on being unfamiliar (i.e., non-celebrity) Caucasian individuals with a limited number of tattoos (covering less than roughly 25%

of the face). Tattoos varied in number, size, and placement on the face; while some tattoos were of the same object (e.g., more than one tattoo of an anchor), no two tattoos were identical. A comparable number of faces faced forward with both eyes visible and faced slightly to the side with at least one eye visible. A comparable number of faces were unpierced and had facial piercings. Additionally, for the male faces, a comparable number was clean-shaven and bearded, as well as were wearing a hat and no hat.

Two versions of 96 faces were created: tattooed and untattooed. Half of the faces were of males (48), and half were of females. For each Tattooed stimulus, the image was converted to black-and-white, and the background was removed, leaving only the head (see Appendix B for example stimuli); any face facing to one side was flipped so that it was looking to the left from the viewer's perspective. For the untattooed stimuli, the tattooed equivalent was taken and further altered to remove any visible tattoo and piercing. The experiment was programmed using PsychoPy (Peirce et al., 2019), and the post-experiment and demographic questionnaires were conducted through Qualtrics (Qualtrics, 2022).

Procedure. Following consent through a separate Qualtrics survey, the participants were brought into the laboratory with the experiment on the computer screen. Like Experiment 1, participants were told to remember faces for an upcoming memory test in the study phase. Forty-eight faces were presented, one at a time, at the centre of the computer screen for 1000 milliseconds[11] on a grey background; the stimuli took up

[11] Based on a pilot Honours student project which presented faces for 3000ms each and resulted in high hit rates (*M* = .80 for tattooed faces; *M* = .81 for untattooed faces), study duration was reduced to 1000ms for this experiment.

roughly ¾ of the screen. A screen with a fixation cross in the middle of it was presented following each face stimulus for 500ms. Half of the stimuli were male faces, and half were female faces. Of the 24 male and 24 female faces, half (12) of each had tattoos, and half did not; all faces were unique, with none being presented in another condition. Assignment to condition was randomized. The study phase was followed by a three-minute retention task where participants had to differentiate between even and odd numbers.

At test, all 48 studied faces and 48 new faces were presented one by one on grey backgrounds; similar to the studied faces, an equal number of male and female and tattooed and untattooed condition faces were presented. Participants were asked to indicate whether the face was studied previously or was new by pressing the 'z' or 'm' keys on the keyboard, respectively. Following each selection, participants were also asked to provide their confidence in their choice by typing a number between 1-6 and pressing the "Enter" key; it was indicated by instructions on the screen that the scale ranged from 1 (not confident) to 6 (confident). Following the experiment, data were automatically uploaded to the local server, and a link to the Qualtrics survey containing the questionnaires was automatically opened.

Results

Of the 31 participants who responded to the post-experiment questionnaire, all had noticed something unique about some of the faces they studied. Estimates of what percentage of faces had a unique marking ranged from 5-87.50% (M = 54.05%; SD = 24.19%). When asked to clarify what exactly was unique about each face, 29 participants had mentioned the presence of tattoos (93.55%). When asked whether such a marking had made the faces more memorable, 23 participants responded “yes” (74.19%) and 7 responded “no” (22.58%); 1 did not respond to the question. Of those who responded “yes,” 16 participants mentioned generally that the tattoos stood out and made the faces more distinguishable. Of those who responded “no,” 5 participants mentioned that they either remembered the tattoos more than the faces or that the tattoos were not memorable if they were small or had a basic design (e.g., a simple heart or star). For responses to questions regarding tattoos, see Appendix I.

From the old-new recognition test, overall hit and false alarm rates, a measure of sensitivity *d’*, and the response bias measure *c* were calculated (see Table 6 for descriptive statistics). Overall response times and confidence ratings for hits and correct rejections were also calculated (see Table 7 for descriptive statistics). Alpha levels of .05 were used for all analyses. A G*Power post hoc, paired samples difference of means analysis (Faul et al., 2009), using this number of participants, power set to .95, and an alpha level of .05, determined that the minimum detectable effect size was .66. Exploratory gender- and race-based analyses can be found in Appendix E and F, respectively, though sample sizes were again not controlled for. Additionally, comparisons based on the presence of participants’ tattoos can be found in Appendix J; sample sizes were not controlled for.

The four main measures were analyzed using separate paired-samples *t*-tests. For the hit rates, there were no differences between the tattooed and untattooed conditions, $t(31) = 0.05$, $p = .96$, Cohen’s $d = 0.01$. For the false alarm rates, there were no differences between the tattooed and untattooed conditions, *t*(31) = 1.35, *p* = .19, Cohen's *d* = 0.24. For the sensitivity measure *d'*, there were no differences between the tattooed and untattooed conditions, *t*(31) = 1.43, *p* = .16, Cohen's *d* = 0.25. Lastly, for the response bias measure *c*, there were no differences between the tattooed and untattooed conditions, *t*(31) = 1.23, *p* = .23, Cohen's *d* = 0.22.

Table 6

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Condition for Experiment 4a

Condition	Hits	FA	*d'*	*c*
Tattooed	.63	.27	1.08	.20
	(.02)	(.03)	(.11)	(.07)
Untattooed	.63	.31	0.91	.12
	(.02)	(.03)	(.09)	(.06)

Note: Standard Errors are presented in parentheses below their corresponding means.

Response times and confidence ratings were analyzed using separate 2 (Condition: Tattooed × Untattooed) × 2 (Decision Type: Hits × Correct Rejections) repeated measures ANOVA. For the response times, the main effect of condition was significant, *F*(1,31) = 6.62, *MSE* = 521696, *p* = .02, η^2p = .18, with quicker responses to the untattooed stimuli (*M* = 2957, *SEM* = 188) than the tattooed stimuli (*M* = 3285, *SEM* = 267). The main effect of decision type was non-significant, *F*(1,31) = 2.86, *MSE* =

566900, p = .10, η^2p = .09; the interaction was also non-significant, F(1,31) = 0.06, MSE = 316839, p = .80, η^2p = .002.

For confidence, the main effect of condition was non-significant, F(1,31) = 0.29, MSE = .17, p = .59, η^2p = .01. The main effect of decision type was significant, F(1,31) = 34.87, MSE = .45, p < .001, η^2p = .53, with higher confidence to hits (M = 4.69, SEM = .10) than correct rejections (M = 4.00, SEM = .13). The interaction was non-significant, F(1,31) = 0.34, MSE = .12, p = .56, η^2p = .01.

Table 7

Response Times and Confidence Ratings for Hits and Correct Rejections by Condition for Experiment 4a

	Response Times (ms)		Confidence	
Condition	Hits	Correct Rejection	Hits	Correct Rejection
Tattooed	3160	3410	4.73	4.00
	(270)	(298)	(.11)	(.12)
Untattooed	2857	3057	4.66	4.00
	(191)	(206)	(.12)	(.14)

Note: Standard Errors are presented in parentheses below their corresponding means.

Discussion

In Experiment 4a, I presented participants with a series of faces that were either tattooed or untattooed. I then tested the recognition of these faces in an old-new recognition test. I hypothesized that recognition would be best for the tattooed stimuli.

My hypothesis was not supported. There were no significant differences in the recognition of faces regardless of condition for the four primary measures, response bias, and confidence. The only differences found were in response times, with decisions being made more quickly for non-tattooed faces than tattooed ones.

When predicting the results of this experiment, two outcomes were considered. First, given that tattoos could be seen as polarizing and are often associated with negative characteristics, such as untrustworthiness (Broussard & Harton, 2018) and deviance (as evidenced by my post-experiment questionnaires across all experiments, see Appendix I), performance could have been worse for the tattooed stimuli due to out-group biases being formed. If the participants did not identify with the faces with tattoos and saw them as “other” because of their opinions on who does and does not get face tattoos, this could have been detrimental to their recognition, akin to other out-group biases (e.g., Anastasi & Rhodes, 2005; Hills et al., 2018; Malpass & Kravitz, 1969). Alternatively, in comparison to the lack of variability between moles and scars, which could have been a detriment to performance in earlier experiments, the variability among individual tattoos could have instead benefitted memory for tattooed faces. If any amount of the face was encoded on each study trial, combined with the differences among the tattoos between trials, recognition performance might actually be improved. In my predictions, I favoured the latter of the two possibilities.

Unfortunately, my results supported neither of the two considered outcomes. Performance was not better for the untattooed nor tattooed faces; it was relatively comparable. In regard to the possibility of out-group biases being formed, this was not the case, as evidenced by the similar performance between conditions. The supplementary analyses also evidenced this by comparing participants with and without tattoos (see Appendix J). As no differences were found, even for participants with and without visible tattoos, no differences in encoding based on out-group biases were likely present in this experiment.

In regard to my hypothesis that tattoos would be different enough from another such that the variability would benefit memory beyond that of untattooed faces, they might not have led to increased recognition performance at test, but they at least did not hinder performance either, unlike the extraneous markings in Experiments 1 and 2. Through each experiment thus far, it has been discussed that extraneous markings likely distract participants from encoding the faces the markings are on at study. When left with just the markings – the moles and scars – to assist in differentiation at test, they are too similar to one another to be beneficial, leading to the detriments I have observed. In this experiment, if participants were distracted by the tattoos, they, again, would not have encoded enough of each face at study to help in later recognition. However, participants would at least have been able to use the differences among the tattoos to assist in their recognition of the tattooed faces instead. These differences would be more comparable to the differences found between different internal facial features than what the differences between moles would be, for example.

Something important to note, however, is the comparable pattern between this experiment and Experiment 3, in which it was also found that there was no difference in performance between the altered and unaltered stimuli. In that experiment, it was discussed that one cause for the lack of detriment to the altered faces could have been the differences in the number of each stimulus presented. Namely, while two-thirds of the stimuli used in the first two experiments were altered to include an extraneous marking, there were an equal number of faces with an extraneous marking and faces with no marking or alteration used in Experiment 3. If the sheer volume of markings present in the first two experiments dulled the effect of the markings, that would explain the equitable performance in the third. Similarly, this experiment also used an equal number of faces with an extraneous marking and faces with no marking; the comparable difference in performance would also support the volume of markings affecting the findings. However, this would not explain why there was still no benefit found for the faces with the markings if the number of stimuli in each condition was critical.

Alternatively, it was also discussed in Experiment 3 that it was more likely that the lack of difference in performance between unaltered and mole conditions could have been due to the presence of the faces with distinct internal facial features in the study list. If participants believed the features to be helpful in later recognition, attention would have shifted toward the internal features first at study, preventing the distraction of the similar moles. In a similar vein, as mentioned above, while there was nothing preventing participants' attention from being taken away from the markings in this experiment, the detriment of that distraction could have been negated instead by the differences within the markings.

Based on the findings throughout each experiment thus far, it is also important to consider at this point what else might be happening when the participants' attention is taken away from the other facial features at study. Specifically, while I have discussed that the presence of extraneous markings could be “distracting,” I have also discussed that this distraction is dependent on the type of marking that is present and what other types of stimuli are present during study. In the current experiment, while no differences were found in memory for the faces with and without the markings, there were differences in the response times to each type of stimulus, with quicker responses to the untattooed faces. This could support that extraneous markings not only take attention away from the face itself when present but also that these markings are processed differently in some way. Specifically, instead of the extraneous markings being incorporated into facial recognition and facial processing, as has been the assumption thus far, participants might instead be shifting to the use of object recognition when the markings are present, neglecting the encoding of the rest of the face.

As mentioned, objects and faces are processed differently, with objects being processed in a feature- or part-based manner (Biederman, 1987) and faces being processed holistically (Tanaka & Farah, 1993). Given that the tattoos in this experiment were of objects, such as anchors, hearts, and stars, it is possible that they were not processed in relation to the faces themselves (as *facial tattoos*) but rather processed in an entirely different manner (as *objects*). The similar performance between the two conditions, then, could have just been due to a similar benefit of either type of processing. Additionally, applying this possibility to the previous experiments, it is also possible that the moles and scars used previously were also encoded as objects instead of being encoded as a feature of the face. Following along with previous discussions, recognition performance can be promoted or hindered, depending on the degree of differentiation among similar objects. Lastly, as discussed in Experiment 3, the presence of the distinct internal facial features in that experiment could have discouraged shifting to object recognition, which might have otherwise occurred.

To touch briefly on an alternative explanation for the findings here, in previous discussions on the MDS framework (Valentine, 1991; Valentine et al., 2016), it was suggested that faces with extraneous markings were clustered in face-space, resulting in worse performance for those faces. That is not entirely an implausible explanation here, either. While the tattooed stimuli could have been stored separately in face-space, given the incompatibility of the dimensions of the faces with each participant's face-space, they still could have been different enough from one another and memorable enough to allow for proper and error-less encoding of those faces. This separation of the cluster in the space could have prevented better performance for those faces, but the distinction between faces could have led to sufficient enough differentiation that performance was not hindered. Adding to this, the differences in response times could have been due to the longer processing time required to reach the distant cluster within the space.

One final note is regarding the study duration, which was shorter than in previous experiments. Given the quick presentation time, only 1000ms compared to 1500ms (Experiments 1 and 3) and 3000ms (Experiment 2), it is possible that participants did not have enough time to encode the entire configuration of the faces at study; this could even have forced them to rely on object recognition instead processing the tattoos in relation to the faces. However, despite the quick presentation time, overall recognition of the faces was comparable to that of the previous experiments. Additionally, participants had twice as long to study each face in Experiment 2, compared to Experiments 1 and 3, which did not increase performance for altered faces. Given this, the study duration in this experiment is not a concern.

Taken together, the results of Experiment 4a seem to be consistent with my previous findings but also add insight into the actual processes happening when extraneous markings are present. When present, extraneous markings disrupt facial processing, not only in distracting away from encoding the faces themselves but also in creating a shift in the type of processing that occurs. This disruption and shift in processing are not necessarily as bad as previously found, however, as their benefit or detriment is influenced by other factors, such as the variability among markings.

Experiment 4b

Based on the new insights that were discussed in Experiment 4a, Experiment 4b explored whether extraneous markings are processed as objects using inversion. Specifically, if the facial tattoos used in the previous experiment are processed as objects, as opposed to as features of the faces themselves, then the recognition of inverted tattooed and untattooed faces should differ. As discussed previously, the inversion of a face is more detrimental to the recognition of it compared to the inversion of an object (Yin, 1969; see also: Civile et al., 2014; McKone & Yovel, 2009; Richler & Gauthier, 2014). Given this, if participants are primarily encoding the tattoos on tattooed trials, and they are shifting their processing to be that of the processing of an object, then recognition should be less impaired when the studied tattooed faces are later inverted at test than when a studied untattooed face is later inverted at test.

Method

Participants. Fifty-four undergraduate students from Memorial University of Newfoundland participated through the Psychology Research Experience Pool (PREP) in exchange for course credit. Data collection was conducted online using the PsychoPy (Peirce et al., 2019) online extension, Pavlovia (Open Science Tools, 2022), and six participants' data had to be excluded due to not following instructions, leaving a final sample size of 48 participants.

Of the 48 participants, 32 identified as female (66.67%), five identified as male (10.42%), one as non-binary (2.08%), and ten chose not to respond. The average age of participants was 21.26 years old (*SD* = 3.96). Twenty-eight participants identified as

Caucasian (58.33%), five as South Asian (10.42%), three as Middle Eastern (6.25%), one as East Asian (2.08%), one as African (2.08%), and ten chose not to respond. Regarding household income, nine participants reported an average income of less than $20,000 (18.75%), four between $20,000 and $35,000 (8.33%), four between $35,000 and $50,000 (8.33%), four between $50,000 and $75,000 (68.33%), four between $75,000 and $100,000 (8.33%), and twelve over $100,000 (25.00%); eleven participants chose not to respond.

Materials. Materials were identical to Experiment 4a with the exception that identical but inverted versions of each tattooed and untattooed face were created by rotating the original image 180°. The experiment was programmed using PsychoPy (Peirce et al., 2019), and the post-experiment and demographic questionnaires were conducted through Qualtrics (Qualtrics, 2022).

Procedure. Following consent through a separate Qualtrics survey, participants were provided with a link to access the experiment online through Pavlovia. Like Experiment 1, participants were told to remember faces for an upcoming memory test in the study phase. Forty-eight upright faces were presented, one at a time, at the centre of the computer screen for 1000 milliseconds on a grey background; the stimuli took up roughly ¾ of the screen. A screen with a fixation cross in the middle of it was presented following each face stimulus for 500ms. Half of the stimuli were male faces, and half were female faces. Of the 24 male and 24 female faces, half (12) of each had tattoos, and half did not; all faces were unique, with none being presented in another condition. Assignment to condition was randomized. The study phase was followed by a three-minute retention task where participants had to differentiate between even and odd numbers.

At test, 24 of the studied faces (12 male and 12 female faces; 12 tattooed and 12 untattooed faces) were presented upright, in their original form, and 24 were presented as inverted (a 180° rotation). Additionally, 48 new faces (24 upright and 24 inverted; 24 male and 24 female faces; 24 tattooed and 24 untattooed faces) were also shown. All faces were presented one by one on grey backgrounds. Participants were asked to indicate whether the face was studied previously, regardless of whether it was upright or inverted, or was new by pressing the 'z' or 'm' keys on the keyboard, respectively. Following each selection, participants were also asked to provide their confidence in their choice by typing a number between 1-6 and pressing the "Enter" key; it was indicated by instructions on the screen that the scale ranged from 1 (not confident) to 6 (confident). Following the experiment, data were automatically uploaded to Pavlovia, and a link to the Qualtrics survey containing the questionnaires was automatically opened.

Results

Of the 38 participants who responded to the post-experiment questionnaire, all had noticed something unique about some of the faces they studied. Estimates of what percentage of faces had a unique marking ranged from 13-90% (M = 60.08%; SD = 20.08%). When asked to clarify what exactly was unique about each face, 36 participants had mentioned the presence of tattoos (94.74%). When asked whether such a marking had made the faces more memorable, 33 participants responded "yes" (86.84%) and two responded "no" (5.26%); three did not respond to the question. Of those who responded "yes," 16 participants mentioned generally that the tattoos stood out and made the faces more distinguishable. Of those who responded "no," participants mentioned that the tattoos were memorable themselves, but they did not make the face memorable. For responses to questions regarding tattoos, see Appendix I.

From the old-new recognition test, overall hit and false alarm rates, a measure of sensitivity *d'*, and the response bias measure *c* were calculated (see Table 8 for descriptive statistics). Overall response times and confidence ratings for hits and correct rejections were also calculated (see Table 9 for descriptive statistics). Alpha levels of .05 were used for all analyses. A G*Power post hoc, paired samples difference of means analysis (Faul et al., 2009), using this number of participants, power set to .95, and an alpha level of .05, determined that the minimum detectable effect size was .55. Exploratory gender- and race-based analyses can be found in Appendix E and F, respectively, though sample sizes were again not controlled for. Additionally, comparisons based on the presence of participants' tattoos can be found in Appendix J; sample sizes were not controlled for.

The four main measures were analyzed using separate 2 (Condition: Tattooed × Untattooed) × 2 (Orientation: Upright × Inverted) repeated measures ANOVAs. For the hit rates, the main effect of condition was significant, with greater hits for the tattooed stimuli (*M* = .61; *SEM* = .02) than the untattooed stimuli (*M* = .55; *SEM* = .02), $F(1,47) = 7.78$, *MSE* = 0.03, *p* = .01, $\eta^2 p = .14$. The main effect of orientation was also significant, with greater hits when the stimuli were upright (*M* = .61; *SEM* = .02) than when they were inverted (*M* = .55; *SEM* = .12), $F(1,47) = 10.70$, *MSE* = 0.01, *p* = .002, $\eta^2 p = .19$.

The interaction was also significant, $F(1,47) = 4.96$, $MSE = 0.01$, $p = .03$, $\eta^2p = .10$. Follow up Tukey post-hoc comparisons found no differences in hit rates between the upright tattooed and upright untattooed stimuli, $t(47) = 1.02$, $p = .74$, Cohen's $d = 0.17$, but greater hits to the inverted tattooed stimuli compared to the inverted untattooed stimuli, $t(47) = 3.45$, $p = .01$, Cohen's $d = 0.57$. There were also no differences in hit rates between the upright and inverted tattooed stimuli, $t(47) = 0.85$, $p = .83$, Cohen's $d = 0.13$, but greater hit rates to the upright untattooed stimuli compared to the inverted untattooed stimuli, $t(47) = 3.81$, $p = .002$, Cohen's $d = 0.51$. That is, inversion negatively affected detection of studied untattooed faces, but had no effect on detection of studied tattooed faces.

For the false alarm rates, the main effect of condition was non-significant, $F(1,47) = 0.29$, $MSE = 0.03$, $p = .60$, $\eta^2p = .01$. The main effect of orientation was significant, with greater false alarms when the stimuli were inverted ($M = .38$; $SEM = .02$) than when they were upright ($M = .33$; $SEM = .02$), $F(1,47) = 5.67$, $MSE = 0.02$, $p = .02$, $\eta^2p = .11$. The interaction was non-significant, $F(1,47) = 1.33$, $MSE = 0.01$, $p = .26$, $\eta^2p = .03$.

For the sensitivity measure *d'*, the main effect of condition was significant, with greater sensitivity for the tattooed stimuli ($M = 0.75$; $SEM = .07$) than the untattooed stimuli ($M = 0.53$; $SEM = .07$), $F(1,47) = 7.38$, $MSE = 0.34$, $p = .01$, $\eta^2p = .14$. The main effect of orientation was also significant, with greater sensitivity when the stimuli were upright ($M = 0.78$; $SEM = .07$) than when they were inverted ($M = 0.50$; $SEM = .06$), $F(1,47) = 18.72$, $MSE = 0.21$, $p < .001$, $\eta^2p = .29$. The interaction was non-significant, $F(1,47) = 0.20$, $MSE = 0.26$, $p = .66$, $\eta^2p = .004$.

Lastly, for the response bias measure *c*, the main effect of condition was non-significant, $F(1,47) = 1.61$, $MSE = 0.14$, $p = .21$, $\eta^2 p = .03$. The main effect of orientation was also non-significant, $F(1,47) = 0.16$, $MSE = 0.09$, $p = .69$, $\eta^2 p = .003$. The interaction was significant, $F(1,47) = 6.90$, $MSE = 0.05$, $p = .01$, $\eta^2 p = .13$. Follow up Tukey post-hoc comparisons, however, found no significant differences. The interaction lay in small numeric differences that reached significance when considering uncorrected *p*-values; the only notable difference was a more conservative response bias for the inverted untattooed stimuli compared to the inverted tattooed stimuli, $t(47) = 2.28$, $p = .03$, Cohen's $d = .35$.

Table 8

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Condition for Experiment 4b

Condition	Orientation	Hits	FA	*d'*	*c*
Tattooed	Upright	.62	.31	0.88	.10
		(.02)	(.02)	(0.09)	(.05)
	Inverted	.60	.38	0.63	.03
		(.02)	(.03)	(0.08)	(.05)
Untattooed	Upright	.59	.35	0.68	.09
		(.02)	(.02)	(0.08)	(.05)
	Inverted	.50	.37	0.37	.19
		(.03)	(.03)	(0.09)	(.07)

Note: Standard Errors are presented in parentheses below their corresponding means.

Response times and confidence ratings were analyzed using separate 2 (Condition: Tattooed × Untattooed) × 2 (Orientation: Upright × Inverted) × 2 (Decision Type: Hits × Correct Rejections) repeated measures ANOVAs. For the response times, the main effect of condition was non-significant, $F(1,47) = 0.50$, $MSE = 1.19 \times 10^7$, $p = .48$, $\eta^2p = .01$. The main effect of orientation was significant, with quicker responses to the upright stimuli ($M = 2229$, $SEM = 128$) than the inverted stimuli ($M = 2578$, $SEM = 168$), $F(1,47) = 9.13$, $MSE = 1.28 \times 10^6$, $p = .004$, $\eta^2p = .16$. The main effect of decision type was non-significant, $F(1,47) = 0.01$, $MSE = 1.16 \times 10^6$, $p = .92$, $\eta^2p = .00$. The Condition × Orientation interaction was significant, $F(1,47) = 6.70$, $MSE = 719174$, $p = .01$, $\eta^2p = .13$. Follow up Tukey post-hoc comparisons found longer response times to inverted untattooed stimuli than upright untattooed stimuli, $t(47) = 3.18$, $p = .01$, Cohen's $d = 0.41$, but no differences between inverted and upright tattooed stimuli $t(47) = 1.31$, $p = .56$, Cohen's $d = 0.09$. The Condition × Decision Type interaction was non-significant, $F(1,47) = 0.78$, $MSE = 1.67 \times 10^6$, $p = .38$, $\eta^2p = .02$. The Orientation × Decision Type interaction was non-significant, $F(1,47) = 0.05$, $MSE = 811232$, $p = .82$, $\eta^2p = .001$. The Condition × Orientation × Decision Type interaction was also non-significant, $F(1,47) = 0.16$, $MSE = 2.26 \times 10^6$, $p = .69$, $\eta^2p = .003$.

For confidence, the main effect of condition was non-significant, $F(1,47) = 2.44$, $MSE = .37$, $p = .13$, $\eta^2p = .05$. The main effect of orientation was significant, with higher confidence in decisions regarding the upright stimuli ($M = 4.25$, $SEM = .10$) than the inverted stimuli ($M = 3.94$, $SEM = .11$), $F(1,47) = 40.94$, $MSE = .22$, $p < .001$, $\eta^2p = .47$. The main effect of decision type was significant, with higher confidence in decisions made to hits (*M* = 4.51, *SEM* = .01) than correct rejections (*M* = 3.67, *SEM* = .12), *F*(1,47) = 68.05, *MSE* = .77, *p* < .001, η^2p = .65. The Condition × Orientation interaction was non-significant, *F*(1,47) = 0.66, *MSE* = .29, *p* = .42, η^2p = .01. The Condition × Decision Type interaction was non-significant, *F*(1,47) = 0.47, *MSE* = .31, *p* = .50, η^2p = .01. The Orientation × Decision Type interaction was non-significant, *F*(1,47) = 2.85, *MSE* = .28, *p* = .10, η^2p = .06. The Condition × Orientation × Decision Type interaction was also non-significant, *F*(1,47) = 1.10, *MSE* = .33, *p* = .30, η^2p = .02.

Table 9

Response Times and Confidence Ratings for Hits and Correct Rejections by Condition for Experiment 4b

		Response Times (ms)		Confidence	
Condition	Orientation	Hits	Correct Rejection	Hits	Correct Rejection
Tattooed	Upright	2423	2337	4.73	3.82
		(231)	(189)	(.11)	(.14)
	Inverted	2590	2421	4.44	3.58
		(249)	(156)	(.11)	(.15)
Untattooed	Upright	2066	2088	4.70	3.74
		(109)	(165)	(.12)	(.13)
	Inverted	2556	2745	4.19	3.54
		(209)	(289)	(.15)	(.13)

Note: Standard Errors are presented in parentheses below their corresponding means.

Discussion

In Experiment 4b, I presented participants with a series of upright faces that were either tattooed or untattooed. I then tested the recognition of these faces in an old-new recognition test in which half of the faces were upright, and half were inverted. I hypothesized that if the distraction of the extraneous markings caused a shift to processing the markings as an object instead of a feature of the face it was on, there would be a difference in the impact of inversion on recognition, with untattooed faces being more greatly impacted by the inversion. My hypothesis was generally supported. In regard to the differences between the tattooed and untattooed faces, I found greater hits and discriminability for the tattooed faces, though no differences in false alarms, response bias, response times, or confidence. In regard to the effects of face orientation, I found a greater number of hits, fewer false alarms, better discriminability, quicker response times, and more confidence in decisions for upright faces; no differences were found in response biases. Importantly, I found that hit rates were lower and response times were slower for inverted untattooed faces, relative to upright faces, but no such differences for tattooed faces based on orientation. I did not find interactions between condition and orientation for false alarms, discriminability, response bias, or confidence.

As discussed in Experiment 4a, while extraneous markings do appear to take attention away from encoding faces at study, more appears to be happening as well. Specifically, there appears to be a shift in the type of processing that is occurring, with participants encoding the marking as its own entity, as an object, focusing less on encoding the faces behind the markings. To support this, the current experiment used an inversion manipulation at test, as facial processing is more disrupted by inversion than is object processing (Yin, 1969). When participants were presented with pictures of faces and objects that were either upright or inverted, Yin found that faces held an advantage over objects when they were upright, with participants making fewer errors in selecting between a target and a foil. When inverted, however, errors in the recognition of both types of stimuli increased, but the increase was more substantial for faces compared to objects. One reason for this differential impact of inversion is in the manner in which each type of stimuli is processed, namely that faces are processed holistically, and objects are processed part by part. Applying this to the current experiment, a similar finding would occur if extraneous markings are indeed processed as objects. While I did not find an interaction in false alarm rates or discriminability, I did find an interaction in my hit rates and response times. Importantly, I found untattooed faces were negatively impacted by inversion, as would be expected with facial recognition, and I found that tattooed faces were not impacted by inversion, as would be more expected with object recognition.

Support for differential processing aside, an interesting finding in this experiment that should be highlighted is that, for the first time, overall performance was better for the faces with extraneous markings compared to the faces without any markings. Not only were hits higher for the tattooed faces, but discriminability was also better. This is highly noteworthy as this difference was not found in Experiment 4a, and the study phases were identical in both experiments. One reason for this discrepancy between experiments could be that this experiment used data from more participants and was statistically more powerful. Had power been higher in Experiment 4a, similar findings may have occurred.

Additionally, if the reason for this discrepancy between the experiments is power, then it is important to note that the hypothesis for Experiment 4a would have been correct, though the reason may be different than originally suggested. Initially, it was hypothesized that the memory for the tattoos would have been compounded with memory for the faces themselves, resulting in better recognition overall. Given the differences based on inversion here, however, it is more likely that recognition was based on tattoos alone.

It is also important to compare the findings of this experiment to that of Experiment 1, which was similar in power but found the reverse pattern of results. As mentioned, one such difference between the two experiments is in the number of stimuli with markings presented. Given that there were an equal number of stimuli with and without markings in this experiment, similar to Experiment 3, the detriment in performance for the altered stimuli in Experiment 1 could have been caused by the higher proportion of stimuli reducing any potential encoding benefits that may arise from them. Another key difference is the variability among markings. As noted, the similarities among markings in the first experiment could have made differentiation between like-items difficult, whereas the same could not be said for the tattoos used here. Given that Experiment 3 showed comparable performance between the altered and unaltered conditions, instead of better for the altered one, it is more likely that the differences lie in the markings themselves and not the relative proportion of stimuli with markings.

Finally, as with previous experiments, alternative explanations for my overall findings here can also align with discussions on the MDS framework (Valentine, 1991).

As discussed in Experiment 4a, if tattoos are incompatible with the existing dimensions within the participants' face-space, the tattooed stimuli would be stored in a cluster within the space that is away from the central point. Given the uniqueness of each tattoo, however, there would be somewhat of an advantage of these types of markings over other markings, such as moles, as they could be better spaced from one another. This could explain why the tattooed faces were more resistant to inversion than the untattooed faces. As described by Valentine, inversion is not as detrimental to faces within face-space that are more isolated from one another, such as ones with distinct internal features, than those that are more clustered around one another. Inversion would cause more errors when searching for faces within face-space as the search conditions do not match the encoding conditions, but those errors are influenced by the number of competing faces around the target. If this is the case, along with the differences in hit rates and discriminability between the tattooed and untattooed faces, it would suggest that extraneous markings and distinct internal facial features are more comparable within face-space than previously considered. In fact, it might even suggest that these markings, at least tattoos, are fully incorporated into face-space as a dimension instead of placed as a separate cluster within the space. What makes this unlikely, however, is that, unlike when comparing distinct faces to more average ones (Valentine & Bruce, 1986c; also see Experiment 3b), response times to recognize the upright tattooed stimuli in this experiment did not differ from the upright untattooed stimuli.

General Discussion

In this study, I sought to explore the effect that extraneous markings – markings not inherently present on a face, though not uncommon to see either – had on facial recognition. While some literature describes such features as *distinct*, impacting eyewitness identification and needing to be controlled for to ensure fair lineup identification (e.g., Zarkadi et al., 2009), the full extent of the memorability of such features was previously unknown. Additionally, a prominent framework used to explain the process of facial recognition, the multi-dimensional space framework (e.g., Valentine, 1991; Valentine et al., 2016), limits the definition of distinctiveness to include features such as the eyes, nose, and mouth, making no reference to extraneous markings. My goals with this study were to determine whether faces with extraneous markings had a memorial benefit over faces without them and whether frameworks, like MDS, that describe facial distinctiveness should or can account for extraneous markings. I initially hypothesized that the recognition of faces with markings would be similar to that of faces with distinct features: better than faces with no extraneous markings.

I examined my hypotheses across four separate experiments. In Experiment 1, I compared recognition for unaltered faces to faces with moles or scars using old-new recognition; I surprisingly found better performance for the unaltered stimuli. In Experiment 2, I explored whether participants could differentiate between the studied faces and the same faces in different conditions using two-alternative-forced-choice; I again found that performance was better for the unaltered stimuli. In Experiment 3, I directly compared recognition for the altered and unaltered faces to recognition of *distinct* faces using old-new recognition; I found better performance for the distinct stimuli and, surprisingly, no differences between the unaltered and mole stimuli. Finally, in Experiment 4a, I compared memory for faces with and without tattoos using old-new recognition, finding no differences between conditions, and in Experiment 4b, I used inversion at test and found better performance for the stimuli with the tattoos.

Addressing the first goal of this study – to determine whether extraneous markings benefit facial recognition by making faces more memorable – memorability appears to be dependent on the type of marking itself. Specifically, it was initially believed that extraneous markings disrupted facial processing, making faces less memorable. This was evidenced by the differences in performance between the altered and unaltered faces in Experiments 1 and 2. This difference in performance was not found in Experiment 3, though the reason for this will be discussed shortly. This difference was also not found in Experiment 4a, but the reverse pattern was found in Experiment 4b, in which performance was, in fact, better for the faces with the markings. While it is likely that the discrepancy in the pattern of results between Experiments 4a and 4b was due to statistical power differences, the discrepancy in results among Experiments 1, 2, and 4b is likely due to the difference in the types of markings used. While Experiments 1 and 2 presented faces with moles and scars, Experiment 4b presented faces with tattoos. Notably, the key difference between these types of markings is the uniqueness between like-markings (i.e., secondary distinctiveness; Schmidt, 1991). Although moles, for example, can vary in size, shape, colour, and position on the face, the relative perceptual variability among moles is smaller in magnitude than the perceptual variability among tattoos; the current stimuli in fact used unique tattoos on each face. This variability among like-markings is more likely to be a factor in the memorability of a face with a marking than just the presence of a marking itself.

It should be noted, however, that the implications of my findings on the memorability of such markings differ from the findings of previous studies that have used similar markings. Specifically, previous studies have explored the impact of markings, such as moles, scars, and tattoos, on fair police lineups and how best to prevent some lineup members from standing out (Badham et al., 2013; Colloff et al., 2016; Zarkadi et al., 2009). Consistently, it has been found that accuracy in the identification of suspects (or rejection of lineups in which suspects are absent) increases when all members of a lineup are more similar than when one member stands out. Accuracy increased when the marking was removed from the suspect so that no member of the lineup had a marking and increased more when the same marking was replicated across all lineup members. In comparison to my findings, these previous findings would suggest that the distinction between markings should not matter, given that the previous studies used similar markings and, if participants were only encoding the markings at study when present, later use of similar markings should lead to worse recognition.

There are notable differences between previous studies and my current one, however. First, both Badham et al. (2013) and Zarkadi et al. (2009) used far fewer faces with extraneous markings at study (with only six out of 32 studied faces possessing a marking). Additionally, while all of Colloff et al.'s (2016) studied stimuli possessed extraneous markings, the stimuli consisted of videos of only four suspects. As noted in previous discussions, it is possible that the number of altered faces used in Experiments 1 and 2 could have impacted my findings, especially considering that recognition was better for the tattooed faces in Experiment 4b, in which an equal number of tattooed and untattooed faces were present. Another notable difference, however, was that Badham et al. (2013), Colloff et al. (2016), and Zarkadi et al. (2009) used a test phase similar to that of a simultaneous lineup, whereas this study used test phases that were more similar to a sequential lineup. Whereas participants in previous studies had to make judgements in the presence of foils, allowing them to make relative comparisons among faces within lineups, participants in this study had to make judgements one at a time without consideration for other faces. So, while the memorability of faces with extraneous markings could be impacted by the total number of faces present that have markings, differences could also lie in whether the later recognition of faces is made in the presence of other faces. In other words, recognition could differ depending on how memorability is tested.

Findings in the current study do fall in line with some other studies similar to those mentioned above, however. In particular, Jones et al. (2020) had participants view a single face that either had a black eye or did not. Participants who were not initially shown the face with the black eye were then either shown the same face alongside five foil faces (target-present lineup) or only shown six foils (target-absent lineup). Participants who were initially shown the face with the black eye then either saw the same face alongside five foils with black eyes (marking replication) or the same face without the black eye (marking removal); these participants also received either a target-present or absent lineup. Interestingly, similar to my study, Jones et al. found that the presence of the black eye harmed discriminability at test. Additionally, unlike past studies (e.g., Zarkadi et al., 2009), no differences were found between the marking removal and marking replication conditions. While this study does provide further support for my findings, with the presence of the marking hindering later performance instead of improving it (Badham et al., 2013; Zarkadi et al., 2009), it should be noted that, as mentioned above, Jones et al. differed greatly in terms of methodology, using only a single face at study and using a simultaneous lineup design at test. Another key difference is in the type of marking used as well. Black eyes differ in that they are usually temporary, healing over time, while a mole, scar, or tattoo is likely permanent (for differences in changeable and invariant aspects of faces, see Haxby et al., 2000).

Addressing the second goal of this study – to determine whether a framework that accounts for distinct internal facial features, such as MDS (Valentine, 1991; Valentine et al., 2016), can account for extraneous facial markings – the effects of extraneous markings on recognition is not parsimoniously explained by the MDS framework. With the MDS framework, it has been described by Valentine that faces are believed to be stored within a multi-dimensional face-space and that each dimension used in facial recognition, like eye size or colour, is stored as a dimension within the space. Depending on how average or atypical a facial feature is in comparison to the rest, faces can either be stored in isolation, away from other faces, or in a cluster close to other faces. Faces with distinct internal facial features are advantaged compared to others because they are stored in isolation, and competition among faces at retrieval is lowered. Additionally, faces with prominent features that are not stored as dimensions within the space are clustered together and off-centre from the rest of the faces (Valentine et al., 1995).

In order to apply the MDS framework to the current findings, it would have to be assumed that the extraneous markings are not stored as dimensions within face-space and that faces with such markings are stored clustered, off-centre within the space. This would make sense, given the poor performance for the altered faces in Experiments 1 and 2, similar to the detrimental performance of other-race faces, which have diagnostic features that differ from the prominent features of same-race faces (Valentine et al., 1995). In Experiment 3, while the comparable performance for the altered and unaltered stimuli does not necessarily support the difference in how the faces are stored, the unaltered stimuli in that experiment were explicitly chosen to be more *typical* than in the other experiments, which could have led to a clustering within face-space that was similar to the off-centre clustering of the extraneous markings, equating performance. Finally, although the improved performance for the tattooed faces in Experiment 4b might support the idea that the tattoos were represented as features within face-space (as differences in discriminability in that experiment were similar to those between faces with distinct internal facial features and faces with no markings in Experiment 3), the lack of differences in response times between overall conditions in Experiment 4b are more supportive of the off-centred clustering. If tattoos were represented as dimensions within face-space, therefore giving tattooed faces the same distinctiveness advantage as distinct internal facial features, overall response times for tattooed faces should have been quicker, as they are with other distinct faces (Valentine, 1991; Valentine et al., 2016). So, while the MDS framework can be applied to explain the effect of extraneous markings on facial recognition, alternative explanations should be considered.

As described previously, the presence of own-biases can impact facial recognition. Specifically, facial recognition is typically worse for faces belonging to a group that is different from our own (e.g., Anastasi & Rhodes, 2005; Cross et al., 1971; Valentine et al., 1995). Given the fact that extraneous markings do not appear on the majority of faces, and that tattoos, in particular, can be seen negatively (Broussard & Harton, 2018) and could cause those with them to be seen as "other, " it is not implausible that own-biases were present in the current study, though it is unlikely. Primarily, had the faces with the markings been processed as "other" by the participants, given their fundamental differences, this could explain the poor performance observed for the altered faces in Experiments 1 and 2. This could not, however, explain the lack of differences found between the mole and unaltered conditions in Experiment 3. Additionally, this could also not explain the increase in performance for the tattooed faces that was observed in Experiment 4b. Though participants were not asked about whether they had moles and scars, meaning that an explicit own-bias analysis could not be made conducted for these experiments, the exploratory analyses conducted in Appendix J showed no differences between participants with and without tattoos in Experiments 4a and 4b. So, while it is possible that the presence of moles and scars caused participants to view these faces as "other" while the presence of tattoos did not, it is more likely that own-biases did not impact the overall findings in relation to the extraneous markings.

Instead of extraneous markings leading faces to be encoded as off-centred clusters within face-space or as “others,” it is more likely that the markings draw attention away from encoding the faces they are on, forcing a disruption in the holistic processing of the face and creating a shift toward encoding the markings themselves as *objects*. The encoding and recognition of these objects, then, relies on the uniqueness (or perceptual distinctiveness) of one object compared to another, which in turn affects how discriminable they are (see Schmidt, 1991). Though not uncommon to see on faces, markings such as moles and scars are absent on the majority of faces. If not expecting it, my participants’ focus could have been pulled away from the internal features of the faces and directed toward the markings, enacting a shift in the processing that occurred. If the markings were very similar to one another, as they were in Experiments 1 and 2, then performance would have been worse for the faces with these markings. If other memorable features were present, such as the distinct features in Experiment 3, and these features negated the distraction caused by the markings, then performance would be similar for the faces with and without these markings. Lastly, if the markings were very different from one another, as they were in Experiment 4b, then performance would have been better for the faces with these markings.

Experiment 4b also provides additional support for the shift in processing type in that faces with and without the markings were differentially impacted by inversion at test, at least in hit rates and response times. As discussed previously, support for the theory that faces are processed holistically instead of in a part-based manner comes from the inversion effect (Yin, 1969), with inversion being more detrimental to the recognition of stimuli relying on holistic processing (faces) than parts-based processing (objects). So, again, while it is possible to explain the effect of extraneous markings through the MDS framework or own-biases, for example, it is more likely that extraneous markings create a shift from the holistic processing of faces to the parts-based processing of the markings (*objects*) that are on the face.

Limitations and Future Directions

The current study was not without limitations. First, as mentioned, the presentation times of faces at study were short, ranging from 1000ms (Experiments 4a and 4b) to 1500ms (Experiments 1 and 3) to 3000ms (Experiment 2). Comparatively, presentation times in other facial recognition studies have ranged from 1500ms (e.g., Valentine & Bruce, 1986b) to 2000ms (e.g., Hills et al., 2013; Zarkadi et al., 2009) to 3000ms (e.g., Badham et al., 2013). Given the short presentation times in my study, participants could have been limited in the amount of information they could encode in the time they had, resulting in poorer performance for faces with extraneous markings. Had participants had more time to study each face, they may have been able to encode both the markings and the internal features of the face, compounding the benefit of encoding both, leading to better overall recognition of the altered faces in all cases. However, it should be noted that Experiment 2 used presentation times that were comparable to other studies, and performance was similar to that of Experiment 1, though the tests used in each differed as well.

Another limitation of the current study was the difference in the ratio of faces with and without extraneous markings in the study phase (i.e., primary distinctiveness;

see Schmidt, 1991). Specifically, in Experiments 1 and 2, although the ratio of mole to scar to unaltered faces was 1:1:1, the ratio of stimuli with extraneous markings to stimuli without markings was 2:1. In contrast, the latter ratio was 1:2 in Experiment 3, and 1:1 in Experiment 4. The difference in numbers could have resulted in a deficit in performance for the altered stimuli that were not present in Experiments 3 and 4a. However, overall performance across experiments was above chance, even for the altered stimuli.

Finally, one limit to the current study, which should be explored further in the future, is that eye-tracking data were not obtained in conjunction with the behavioural data. As noted previously, eye-tracking data has been used to show differences in how some faces are processed in comparison to others. For example, participants have been found to look at different internal facial features depending on the race of the face (Hills et al., 2013). The main conclusions made for this study were on the assumption that the participant's attention was taken away from the internal features of the faces in Experiments 1, 2, 4a, and 4b and that the faces themselves were not encoded as well as they otherwise would have been encoded. Additionally, it was assumed that, in Experiment 3, the presence of the distinct features on some trials negated the effect of the presence of the moles on other trials. Eye-tracking data could further support these conclusions if these markings are found to take attention away from the faces, similar to how eye-tracking data has supported the behavioural data for the weapon focus effect (Loftus et al., 1987). Eye-tracking data could also further support these conclusions if participants give preference to looking at internal facial features when distinct internal facial features are present on some trials. An experiment using eye-tracking was originally planned to accompany this study, though time constraints caused by the COVID-19 pandemic left us unable to feasibly implement it.

Final Conclusions

Across four experiments, I explored whether extraneous facial markings – moles, scars, and tattoos – made faces more memorable and whether their effect on facial recognition could be explained by a framework, such as the multi-dimensional space framework (Valentine, 1991; Valentine et al., 2016). I found that the overall presence of these markings did not help or hinder facial recognition. Instead, the extent to which the markings differed from other like-markings played a more significant role in memorability. I also found that, while it is possible to describe the role of these markings within face-space, it is better suited to describe their influence in terms of a distraction that shifts processing from that for faces (holistic) to that for objects (parts-based), which is, again, affected by the variability among markings. Together, my findings show that although holistic face processing is preferred (Tanaka & Farah, 1993), this preference is not infallible to the presence of external stimuli, such as extraneous markings. Additionally, though you can refer to these markings as *distinct* because they make members of a police lineup stand out, for example (e.g., Badham et al., 2013; Zarkadi et al., 2009), care should be taken by differentiating between extraneous markings and distinct internal facial features (e.g., Bartlett et al., 1984; Cohen & Carr, 1975; Going & Read, 1974; Winograd, 1981), as their overall effects on facial recognition may vary.

References

Adobe. (2019). *Adobe photoshop*. https://www.adobe.com/products/photoshop.html

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997-1010. https://doi.org/10.1037/0278-7393.25.4.997

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*(6), 1043-1047. https://doi.org/10.3758/BF03206441

Arigbabu, O. A., Ahmad, S. M. S., Adnan, W. A. W., & Yussof, S. (2015). Recent advances in facial soft biometrics. *The Visual Computer, 31*(5), 513-525. https://doi.org/10.1007/s00371-014-0990-x

Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition*, *19*(1), 63-71. https://doi.org/10.3758/BF03198496

Badham, S. P., Wade, K. A., Watts, H. J., Woods, N. G., & Maylor, E. A. (2013). Replicating distinctive facial features in lineups: Identification performance in young versus older adults. *Psychonomic Bulletin & Review, 20*(2), 289-295. https://doi.org/10.3758/s13423-012-0339-2

Bar-Haim, Y., Talee, Z., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*(2), 159-163. https://doi.org/10.1111/j.1467-9280.2006.01679.x

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12*(3), 219-228. https://doi.org/10.3758/BF03197669

Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for faces versus single views of faces. *Memory & Cognition*, *14*(5), 371-381. https://doi.org/10.3758/BF03197012

Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition? In *Perception of faces, objects, and scenes: Analytic and holistic processes.* (pp. 21-47). Oxford University Press.

Becerra-Riera, F., Morales-González, A., & Méndez-Vázquez, H. (2018). Facial marks for improving face recognition. *Pattern Recognition Letters, 113,* 3-9. https://doi.org/https://doi.org/10.1016/j.patrec.2017.05.005

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711-720. https://doi.org/10.1109/34.598228

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, *3*(1), 105-135. https://doi.org/10.1080/09541449108406222

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706-712. https://doi.org/10.1111/j.1467-9280.2007.01964.x

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-147. https://doi.org/10.1037/0033-295X.94.2.115

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin, 15*(Mar 89), 19-25.

Brigham, J. C., & Barkowitz, P. (1978). Do "they all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology*, *8*(4), 306-318. https://doi.org/10.1111/j.1559-1816.1978.tb00786.x

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41*(3), 139-155. https://doi.org/10.1111/j.1540-4560.1985.tb01133.x

Broussard, K. A., & Harton, H. C. (2018). Tattoo or taboo? Tattoo stigma and negative attitudes toward tattooed individuals. *The Journal of Social Psychology*, *158*(5), 521-540. https://doi.org/10.1080/00224545.2017.1373622

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, *15*(4), 387-397. https://doi.org/10.1068/p150387

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339-360. https://doi.org/10.1037/1076-898X.5.4.339

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

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. *The Oxford Handbook of Face Perception*, 287-306. https://doi.org/10.1093/oxfordhb/9780199559053.013.0015

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*(3), 243-248. https://doi.org/10.1111/1467-9280.00144

Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual and Motor Skills*, *64*(1), 319-325. https://doi.org/10.2466/pms.1987.64.1.319

Chance, J. E., & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 153-176). Lawrence Erlbaum Associates, Inc.

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 A: Human Experimental Psychology, 48*(4), 879-894. https://doi.org/10.1080/14640749508401421

Choudhury, Z. H., & Mehata, K. (2012). Robust facial marks detection method using aam and surf. *International Journal Engineering Research and Applications*, *2*(6), 708-715.

Civile, C., McLaren, R. P., & McLaren, I. P. L. (2014). The face inversion effect-parts and wholes: Individual features and their configuration. *Quarterly Journal of Experimental Psychology 67*(4), 728-746. https://doi.org/10.1080/17470218.2013.828315

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7*(3), 238-259. https://doi.org/10.1177/1745691612439584

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32*(5), 406-422. https://doi.org/10.1007/s10979-007-9115-7

Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the von restorff effect. *Bulletin of the Psychonomic Society, 6*(4), 383-384. https://doi.org/10.3758/BF03333209

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27*(9), 1227-1239. https://doi.org/10.1177/0956797616655789

Colombatto, C., & McCarthy, G. (2017). The effects of face inversion and face race on the P100 ERP. *Journal of Cognitive Neuroscience*, *29*(4), 664–676. https://doi.org/10.1162/jocn_a_01079

Corenblum, B., & Meissner, C. A. (2006). Recognition of faces of ingroup and outgroup children and adults. *Journal of Experimental Child Psychology*, *93*(3), 187-206. https://doi.org/10.1016/j.jecp.2005.09.001

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/https://doi.org/10.1016/S0022-5371(72)80001-X

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111*(2), 219-247. https://doi.org/10.1016/j.cognition.2009.02.004

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*(6), 393-396. https://doi.org/10.3758/BF03210319

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11*(3), 233-258. https://doi.org/10.1007/BF01044644

Datta, R., Joshi, D., Li, J., & Wang, J. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2), 1-60. https://doi.org/10.1145/1348246.1348248

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117. https://doi.org/10.1037/0096-3445.115.2.107

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups-partial remembering. *Journal of Police Science and Administration, 1*(3), 287-293.

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society, 2*(2), 71-73. https://doi.org/10.3758/BF03327717

Ellis, H. D. (1981). Theoretical aspects of face recognition. In G. M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 171-197). Academic Press.

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). Remembering pictures of real and 'unreal' faces: Some practical and theoretical considerations. *The British Journal of Psychology, 69*(4), 467-474. https://doi.org/10.1111/j.2044-8295.1978.tb02123.x

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of white and black faces by white and black subjects. *International Journal of Psychology, 10*(2), 119-123. https://doi.org/10.1080/00207597508247325

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *Journal of the American Institute of Criminal Law and Criminology, 5*(1), 39-51. https://doi.org/10.2307/1133283

Feinman, S., & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47*(2), 506-510. https://doi.org/10.2307/1128809

Ferguson, D. P., Rhodes, G., Lee, K., & Sriram, N. (2001). 'They all look alike to me': Prejudice and cross-race face recognition. *The British Journal of Psychology, 92*(4), 567-577. https://doi.org/10.1348/000712601162347

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19*(2), 151-164. https://doi.org/10.1037/a0030618

Flowe, H. D., Klatt, T., & Colloff, M. F. (2014). Selecting fillers on emotional appearance improves lineup identification accuracy. *Law and Human Behavior, 38*(6), 509-519. https://doi.org/10.1037/lhb0000101

Fulton, A., & Bartlett, J. C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. *Psychology and Aging, 6*(4), 623-630. https://doi.org/10.1037/0882-7974.6.4.623

Galper, R. E. (1973). “Functional race membership” and recognition of faces. *Perceptual and Motor Skills, 37*(2), 455-462. https://doi.org/10.2466/pms.1973.37.2.455

Galton, F. (1883). *Inquiries into human faculty and its development*. MacMillan Co.

Gitnux. (2023). *The most surprising tattoo popularity statistics and trends in 2023.* Retrieved June 19 from https://blog.gitnux.com/tattoo-popularity-statistics/#:~:text=The%20popularity%20of%20tattoos%20has,29%20have%20at%20least%20one.

Going, M., & Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills, 39*(1), 109-110. https://doi.org/10.2466/pms.1974.39.1.109

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *The Journal of Psychology, 105*(1), 47-59. https://doi.org/10.1080/00223980.1980.9915131

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society, 23*(3), 211-214. https://doi.org/10.3758/BF03329829

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Hampton, J. (1993). Prototype models of concept representation. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). Academic Press.

Harrison, V., & Hole, G. J. (2009). Evidence for a contact-based explanation of the own-age bias in face recognition. *Psychonomic Bulletin & Review, 16*(2), 264-269. https://doi.org/10.3758/PBR.16.2.264

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Hayward, W. G., Crookes, K., Chu, M. H., Favelle, S. K., & Rhodes, G. (2016). Holistic processing of face configurations and components. *Journal of Experimental*

Psychology: Human Perception and Performance, *42*(10), 1482-1489. https://doi.org/10.1037/xhp0000246

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017-1027. https://doi.org/10.1016/j.cognition.2007.04.002

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8*(1), 9.1-910. https://doi.org/10.1167/8.1.9

Herlitz, A., & Loven, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analysis review. *Visual Cognition*, *21*(9-10), 1306-1336. https://doi.org/10.1080/13506285.2013.823140

Heywood, W. B. A., Patrick, K. P., Smith, A. M. A. P., Simpson, J. M. P., Pitts, M. K. P., Richters, J. P., & Shelley, J. M. P. (2012). Who gets tattoos? Demographic and behavioral correlates of ever being tattooed in a representative sample of men and women. *Annals of Epidemiology*, *22*(1), 51-56. https://doi.org/10.1016/j.annepidem.2011.10.005

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986-1004. https://doi.org/10.1037/0096-1523.22.4.986

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking study. *Visual Cognition*, *21*(7), 876-898. https://doi.org/10.1080/13506285.2013.834016

Hills, P. J., Pake, J. M., Dempsey, J. R., & Lewis, M. B. (2018). Exploring the contribution of motivation and experience in the postpubescent own-gender bias in face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 44*(9), 1426-1446. https://doi.org/10.1037/xhp0000533

Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect : Recognition of faces with and without sunglasses. *Memory & Cognition*, *27*(1), 128-138. https://doi.org/10.3758/BF03201219

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. *Memory & Cognition, 41*(7), 1021-1031. https://doi.org/10.3758/s13421-013-0316-7

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology, 43*(2), 334-340. https://doi.org/10.1016/j.jesp.2006.02.010

Hugenberg, K., Wilson, J. P., See, P. E., & Young, S. G. (2013). Towards a synthetic model of own group biases in face memory. *Visual Cognition, 21*(9-10), 1392-1417. https://doi.org/10.1080/13506285.2013.821429

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*(4), 1168. https://doi.org/10.1037/a0020463

Innocence Project. (2020). *How eyewitness misidentification can send innocent people to prison*. Retrieved May 1 from https://innocenceproject.org/news/how-eyewitness-misidentification-can-send-innocent-people-to-prison/#:~:text=There%20are%2082%20DNA%20exoneration,Rickie%20Johnson%20and%20Clifford%20Jones.

Innocence Project. (n.d.). *Ronald cotton*. Retrieved November 1 from https://innocenceproject.org/cases/ronald-cotton/

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*(4), 616-626. https://doi.org/10.1037/0022-3514.85.4.616

Jackson, C. (2013). *More americans have tattoos today than seven years ago*. Retrieved May 15 from https://www.ipsos.com/en-us/news-polls/more-americans-have-tattoos-today

Jain, A. K., & Kumar, A. (2012). Biometric recognition: An overview. In E. Mordini & D. Tzovaras (Eds.), *Second generation biometrics: The ethical, legal and social context* (pp. 49-79). Springer Netherlands. https://doi.org/10.1007/978-94-007-3892-8_3

Jain, A. K., & Park, U. (2009). *Facial marks: Soft biometric for face recognition 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt. https://doi.org/10.1109/ICIP.2009.5413921

Jamovi Project. (2021). *Computer software*. https://www.jamovi.org

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society. B, Biological sciences, 285*(1888), 20181319. https://doi.org/10.1098/rspb.2018.1319

Jones, A. R., Carlson, C. A., Lockamyeir, R. F., Hemby, J. A., Carlson, M. A., & Wooten, A. R. (2020). "All i remember is the black eye": A distinctive facial feature harms eyewitness identification. *Applied Cognitive Psychology, 34*(6), 1379-1393. https://doi.org/10.1002/acp.3714

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

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *The American Psychologist, 56*(5), 405-416. https://doi.org/10.1037/0003-066X.56.5.405

Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics, 20*(6), 475-481. https://doi.org/10.3758/BF03208285

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51*(3), 449-473.

Lee, J.-E., Jain, A. K., & Jin, R. (2008). Scars, marks and tattoos (smt): Soft biometric for suspect and victim identification. *2008 Biometrics Symposium*,

Lee, J.-E., Jin, R., Jain, A. K., & Tong, W. (2012). Image retrieval in forensics: Tattoo image database application. *IEEE Multimedia*, *19*(1), 40-49. https://doi.org/10.1109/MMUL.2011.59

Lee, J., & Penrod, S. D. (2022). Three-level meta-analysis of the other-race bias in facial identification. *Applied Cognitive Psychology*, *36*(5), 1106-1130. https://doi.org/10.1002/acp.3997

Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *Journal of Applied Psychology*, *63*(3), 345-351. https://doi.org/10.1037/0021-9010.63.3.345

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364-1382. https://doi.org/10.1037/0278-7393.22.6.1364

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559-574. https://doi.org/10.1037/0096-3445.129.4.559

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—women's faces make the difference. *Brain and Cognition*, *50*(1), 121-128. https://doi.org/10.1016/S0278-2626(02)00016-7

Light, L., Kayrastuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 212-228. https://doi.org/10.1037/0278-7393.5.3.212

Light, L. L., Hollander, S., & Kayra-Stuart, F. (1981). Why attractive people are harder to remember. *Personality & Social Psychology Bulletin, 7*(2), 269-276. https://doi.org/10.1177/014616728172014

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior*, *11*(1), 55-62. https://doi.org/10.1007/BF01044839

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 77-100. https://doi.org/10.1037/0096-1523.34.1.77

Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women's own-gender bias in face recognition memory: The role of attention at encoding. *Experimental Psychology, 58*(4), 333-340. https://doi.org/10.1027/1618-3169/a000100

Lovén, J., Rehnman, J., Wiens, S., Lindholm, T., Peira, N., & Herlitz, A. (2012). Who are you looking at? The influence of face gender on visual attention and memory for own- and other-race faces. *Memory, 20*(4), 321-331. https://doi.org/10.1080/09658211.2012.658064

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135. https://doi.org/10.3758/s13428-014-0532-5

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces - the ambiguous race face effect. *Psychology, Public Policy, and Law, 7*(1), 98-118. https://doi.org/10.1037//1076-8971.7.1.98

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : A user's guide* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410611147

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*(1), 93-120. https://doi.org/10.1146/annurev.psych.51.1.93

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330-334. https://doi.org/10.1037/h0028434

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics, 14*(2), 285-292. https://doi.org/10.3758/BF03212392

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260. https://doi.org/10.1016/S1364-6613(02)01903-4

McKone, E. (2004). Isolating the special component of face recognition: Peripheral identification and a mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 181-197. https://doi.org/10.1037/0278-7393.30.1.181

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 310-327. https://doi.org/10.1037/0096-1523.34.2.310

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16*(5), 778-797. https://doi.org/10.3758/PBR.16.5.778

Meaux, E., Hernandez, N., Carteau-Martin, I., Martineau, J., Barthélémy, C., Bonnet-Brilhault, F., & Batty, M. (2014). Event-related potential and eye tracking evidence of the developmental dynamics of face processing. *The European Journal of Neuroscience, 39*(8), 1349-1362. https://doi.org/10.1111/ejn.12496

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238. https://doi.org/10.1037/0033-295x.85.3.207

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. https://doi.org/10.1037/1076-8971.7.1.3

Mian, J. F., & Mondloch, C. J. (2012). Recognizing identity in the face of change: The development of an expression-independent representation of facial identity. *Journal of Vision, 12*(7), 1-11. https://doi.org/10.1167/12.7.17

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14*(1), 55-73. https://doi.org/10.1080/13506280500158761

Ng, W.-J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25*(2), 217-232. https://doi.org/10.1177/0022022194252004

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38*(15), 2351-2363. https://doi.org/10.1016/S0042-6989(98)00042-X

Oakes, L. M., & Ellis, A. E. (2013). An eye-tracking investigation of developmental changes in infants' exploration of upright and inverted human faces. *Infancy, 18*(1), 134-148. https://doi.org/10.1111/j.1532-7078.2011.00107.x

Open Science Tools. (2022). *Pavlovia.* www.pavlovia.org

Patterson, K. E., & Baddeley, A. D. (1977). *When face recognition fails* [doi:10.1037/0278-7393.3.4.406].

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96*(4), 795-810. https://doi.org/10.1037/a0013265

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Penev, P. S., & Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Network, 7*(3), 477-500. https://doi.org/10.1088/0954-898X/7/3/002

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: More evidence for the cross-race effect. *Journal of Applied Psychology*, *88*(4), 760-763. https://doi.org/10.1037/0021-9010.88.4.760

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, *3*, 559-559. https://doi.org/10.3389/fpsyg.2012.00559

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior*, *22*(5), 549-570. https://doi.org/10.1023/A:1025739514042

Prolific. (2022). https://www.prolific.co

Psychology Software Tools, I. (2016). [e-prime 3.0]. https://support.pstnet.com/

Psychology Software Tools, I. (2020). [e-prime go]. https://support.pstnet.com/

Qualtrics. (2022). https://www.qualtrics.com

Read, J. D., Vokey, J. R., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy-confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 870-882. https://doi.org/10.1037/0278-7393.16.5.870

Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior & Development, 31*(4), 665-678. https://doi.org/10.1016/j.infbeh.2008.04.009

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*(1), 43-63. https://doi.org/10.1068/p170043

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *The British Journal of Psychology, 80*(3), 313-331. https://doi.org/10.1111/j.2044-8295.1989.tb02323.x

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*(4), 473-497. https://doi.org/https://doi.org/10.1016/0010-0285(87)90016-8

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin, 140*(5), 1281-1302. https://doi.org/10.1037/a0037004

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research 51*(3), 333-342. https://doi.org/10.1016/j.visres.2010.11.014

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition, 5*(2), 144-165. https://doi.org/10.1521/soco.1987.5.2.144

Rollins, L., Bertero, E., & Hunter, L. (2019). Facial features and recognition memory: Eye-movement study on the encoding and retrieval of unfamiliar child faces in children and adults. *Infant and Child Development*. https://doi.org/10.1002/icd.2148

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). Academic Press.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*(2), 274-289. https://doi.org/10.1016/j.actpsy.2008.02.003

Rule, N. O., Ambady, N., Adams Jr, R. B., & Macrae, C. N. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review*, *14*(4), 687-692. https://doi.org/10.3758/BF03196822

Rule, N. O., Garrett, J. V., & Ambady, N. (2010). Places and faces: Geographic environment influences the ingroup memory advantage. *Journal of Personality and Social Psychology*, *98*(3), 343-355. https://doi.org/10.1037/a0018589

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Lawrence Erlbaum.

Sandford, A., & Bindemann, M. (2020). Discrimination and recognition of faces with changed configuration. *Memory & Cognition, 48*(2), 287-298. https://doi.org/10.3758/s13421-019-01010-7

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood.

Psychological Science, 16(6), 440-444. https://doi.org/10.1111/j.0956-7976.2005.01554.x

Scheck, B., Neufeld, P. J., & Dwyer, J. (2003). *Actual innocence: When justice goes wrong and how to make it right*. Signet. https://books.google.ca/books?id=Ic2gQgAACAAJ

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*(6), 523-542. https://doi.org/10.3758/BF03197149

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139-156. https://doi.org/10.1037/0033-2909.100.2.139

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces — a comparison of the responses of africans and europeans to faces of the same and different races. *British Journal of Social Psychology, 20*(2), 125-133. https://doi.org/https://doi.org/10.1111/j.2044-8309.1981.tb00485.x

Shriver, E. R., & Hugenberg, K. (2010). Power, individuation, and the cross-race recognition deficit. *Journal of Experimental Social Psychology, 46*(5), 767-774. https://doi.org/10.1016/j.jesp.2010.03.014

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identifications. *Canadian Journal of Behavioural Science, 36*(2), 146-154. https://doi.org/10.1037/h0087225

Spaun, N. A. (2007). Forensic biometrics from images and video at the federal bureau of investigation. *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, Crystal City, VA, USA. https://doi.org/10.1109/BTAS.2007.4401932

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*(1), 36-97. https://doi.org/10.1037/1076-8971.7.1.36 (Special Theme: The Other-Race Effect and Contemporary Criminal Justice: Eyewitness Identification and Jury Decision Making)

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*(5), 459-473. https://doi.org/10.1023/A:1012888715007

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*(4), 413-424. https://doi.org/10.1007/BF02352267

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*(5), 96-103. https://doi.org/10.1038/scientificamerican1170-96

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

Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & Cognition*, *42*(6), 863-875. https://doi.org/10.3758/s13421-014-0409-y

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352-373. https://doi.org/10.1037/h0020071

Turati, C., Di Giorgio, E., Bardi, L., & Simion, F. (2010). Holistic face processing in newborns, 3-month-old infants, and adults: Evidence from the composite face effect. *Child Development*, *81*(6), 1894-1905. https://doi.org/10.1111/j.1467-8624.2010.01520.x

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

Valentine, T., & Bruce, V. (1986a). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica*, *61*(3), 259-273. https://doi.org/10.1016/0001-6918(86)90085-5

Valentine, T., & Bruce, V. (1986b). The effects of distinctiveness in recognizing and classifying faces. *Perception*, *15*(5), 525-535. https://doi.org/10.1068/p150525

Valentine, T., & Bruce, V. (1986c). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology*, *40*, 300-305. https://doi.org/10.1037/h0080101

Valentine, T., Chiroro, P., & Dixon, R. (1995). An account of the own-race bias and the contact hypothesis based on a 'face space' model of face recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (1 ed., pp. 69-94). Routledge. https://doi.org/https://doi.org/10.4324/9780203428979

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 44*(4), 671-703. https://doi.org/10.1080/14640749208401305

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, 69*(10), 1996-2019. https://doi.org/10.1080/17470218.2014.990392

Vernon, M. D. (1955). The functions of schemata in perceiving. *Psychological Review, 62*(3), 180-192. https://doi.org/10.1037/h0042425

Vokey, J. R., & Read, J. D. (1988). Typicality, familiarity and the recognition of male and female faces. *Canadian Journal of Experimental Psychology, 42*(4), 489.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20*(3), 291-302. https://doi.org/10.3758/BF03199666

Wells, G., Rydell, S., & Seelau, E. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78*(5), 835. https://doi.org/10.1037/0021-9010.78.5.835

Wells, G., Small, M., Penrod, S., Malpass, R., Fulero, S., & Brimacombe, C. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*(6), 603-647. https://doi.org/10.1023/A:1025750605807

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*(2), 89-103. https://doi.org/10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior, 32*(1), 6-10. https://doi.org/10.1007/s10979-007-9098-4

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*(1), 277-295. https://doi.org/10.1146/annurev.psych.54.101601.145028

Wiese, H., Komes, J., & Schweinberger, S. R. (2012). Daily-life contact affects the own-age bias and neural correlates of face memory in elderly participants. *Neuropsychologia*, *50*(14), 3496-3508. https://doi.org/10.1016/j.neuropsychologia.2012.09.022

Wiese, H., Komes, J., & Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-age bias in face recognition. *Visual Cognition*, *21*(9-10), 1337-1363. https://doi.org/10.1080/13506285.2013.823139

Wilson, J. P., & Hugenberg, K. (2010). When under threat, we all look the same: Distinctiveness threat induces ingroup homogeneity in face memory. *Journal of*

Experimental Social Psychology, *46*(6), 1004-1010. https://doi.org/10.1016/j.jesp.2010.07.005

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 181.

Wiskott, L., Fellous, J. M., Kuiger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 775-779. https://doi.org/10.1109/34.598235

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262-276. https://doi.org/10.1037/a0035940

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of us police on preparation and conduct of identification lineups. *Psychology, Crime, & Law*, *10*(1), 69-82. https://doi.org/10.1080/10683160410001641873

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in south africa and england. *Applied Cognitive Psychology*, *17*(3), 365-373. https://doi.org/10.1002/acp.898

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747-759. https://doi.org/10.1068/p160747

Zarkadi, T., Wade, K. A., & Stewart, N. (2009). Creating fair lineups for suspects with distinctive features. *Psychological Science*, *20*(12), 1448-1453. https://doi.org/10.1111/j.1467-9280.2009.02463.x

Appendices

Appendix A: Ethics Clearance Approval

6/16/23, 11:07 AM Memorial University of Newfoundland Mail - ICEHR Clearance # 20201578-SC – EXTENDED

MEMORIAL UNIVERSITY

Kavanagh, Victoria Anna Jean <vajk87@mun.ca>

ICEHR Clearance # 20201578-SC – EXTENDED

1 message

dgulliver@mun.ca <dgulliver@mun.ca> Thu, Mar 2, 2023 at 9:51 AM

To: "Kavanagh Victoria(Principal Investigator)" <vajk87@mun.ca>

Cc: "Hourihan Kathleen(Supervisor)" <khourihan@mun.ca>, ors@mun.ca, dgulliver@mun.ca

MEMORIAL UNIVERSITY

Interdisciplinary Committee on Ethics in Human Research (ICEHR)

ICEHR Approval #:	20201578-SC
Researcher Portal File #:	20201578
Project Title:	*How Extraneous Markings Affect Face Recognition*
Associated Funding:	20131132; 20151025; 20210306
Supervisor:	Dr. Kathleen Hourihan
Clearance expiry date:	**March 31, 2024**

Dear Ms. Victoria Kavanagh:

Thank you for your response to our request for an annual update advising that your project will continue without any changes that would affect ethical relations with human participants.

On behalf of the Chair of ICEHR, I wish to advise that the ethics clearance for this project has been extended to **March 31, 2024**. The *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (TCPS2) requires that you submit another annual update to ICEHR on your project prior to this date.

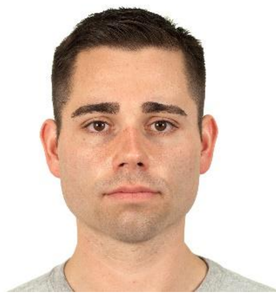
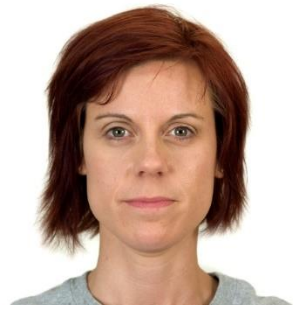
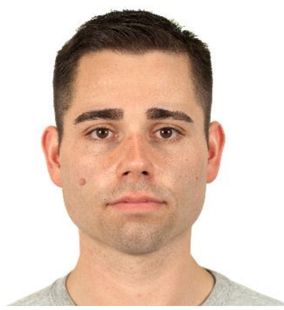
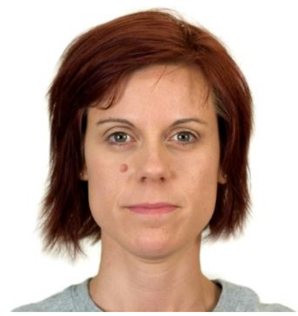
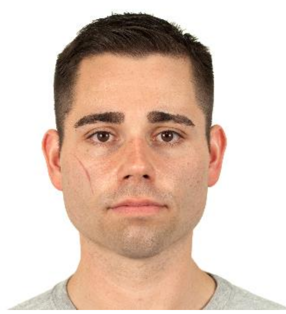
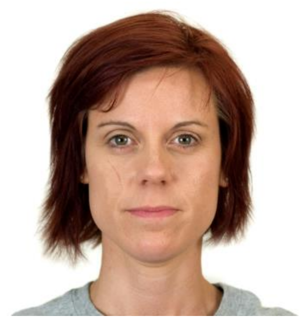
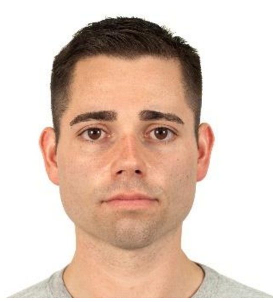
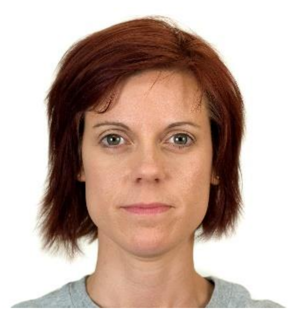
We wish you well with the continuation of your research.

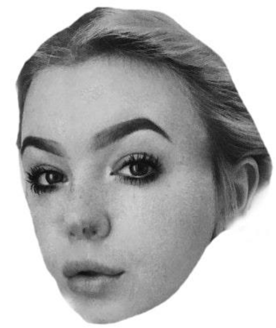
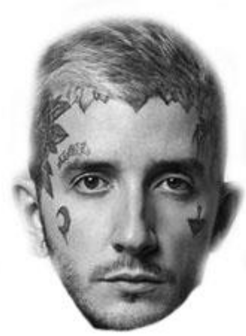
Sincerely,

DEBBY GULLIVER
Interdisciplinary Committee on Ethics in Human Research (ICEHR)
Memorial University of Newfoundland
St. John's, NL | A1C 5S7
Bruneau Centre for Research and Innovation | Room IIC 2010C
T: (709) 864-2561 |
www.mun.ca/research/ethics/humans/icehr | **https://rpresources.mun.ca/**

https://mail.google.com/mail/u/0/?ik=cb45d64411&view=pt&search=all&permthid=thread-f:1759262319157293076&simpl=msg-f:1759262319157293076 1/1

Appendix B: Sample Stimuli

	Stimuli Sex	
Condition	Male	Female
Unaltered Experiments 1-3		
Mole Experiments 1-3		
Scar Experiments 1-2		
Distinct Experiments 3		

Untattooed Experiment 4		
Tattooed Experiment 4		

Appendix C: Post-Experiment Questionnaire

Please provide a response to the questions below. You may choose to decline to provide a response to any question that you do not wish to answer. This information will remain confidential and will be stored on a password protected computer. Only averages will be reported in the dissertation and publications.

1. Did you notice something different or unique about some of the faces you studied? YES ☐/ NO ☐

2. If yes, roughly what percentage of faces seemed different/unique?

3. If yes, what did you notice that was different/unique? Describe below:

4. If yes, do you think these faces were more memorable because of the different/unique feature? If yes, please describe below: YES ☐ / NO ☐

Appendix D: Demographics Questionnaire

Please provide a response to the questions below. You may choose to decline to provide a response to any question that you do not wish to answer. This information will remain confidential and will be stored on a password protected computer. Only averages will be reported in the dissertation and publications.

1. Gender:

2. Age:

3. Ethnicity:

4. Household yearly income:
    a. Less than $20,000
    b. $20,000 to $34,999
    c. $35,000 to $49,999
    d. $50,000 to $74,999
    e. $75,000 to $99,999
    f. Over $100,000

5. Do you have any tattoos? YES ☐ / NO ☐
    a. If yes, how many?

b. If yes, are they generally visible when wearing typical clothing?

   YES ☐ / NO ☐

6. Does someone having a tattoo change your perception of them as a person?

   YES ☐ / NO ☐

   a. Do you think having a tattoo makes someone fundamentally different than someone who does not have a tattoo? YES ☐ / NO ☐

7. Thinking specifically about face and neck tattoos, does someone having a face or neck tattoo change your perception of them as a person? YES ☐ / NO ☐

   a. Again, thinking specifically about face and neck tattoos, do you think having a face or neck tattoo makes someone fundamentally different than someone who does not have a face or neck tattoo? YES ☐ / NO ☐

8. Why do you think that some people might be critical of tattoos?

Appendix E: Exploratory Gender Analyses

Experiment 1 Gender Analysis

Exploratory gender analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 10 for descriptive data). Data were excluded from four participants, the one who identified as non-binary and the three who did not provide their gender; the final analyses were conducted using the male (*n* = 12) and female (*n* = 30) identified participants. All four measures were analyzed using separate 2 (Participant Gender: Male × Female) x 2 (Sex of Stimuli: Male × Female) x 3 (Condition: Unaltered × Mole × Scar) mixed measures ANOVAs. Participant gender was the between-subjects factor while both sex of stimuli and condition were within-subject factors. Overall, effects of condition were found as in the primary analyses, though no gender or sex differences were found nor were there any interactions; so, there is nothing meaningful to report for Experiment 1.

For the hit rates, all main effects and interactions had a *p*-value of .14 or above. For false alarm rates, the main effect of condition was significant, $F(2,80) = 4.64$, $MSE = 0.02$, $p = .01$, $\eta^2p = .10$. Follow up Tukey post-hoc comparisons found no differences in false alarms between the mole ($M = .26$, $SEM = .02$) and scar ($M = .27$, $SEM = .03$) conditions, $t(40) = 0.38$, $p = .93$, but fewer false alarms to the unaltered condition ($M = .21$, $SEM = .02$) compared to both the mole, $t(40) = 2.46$, $p = .05$, and scar conditions, $t(40) = 2.92$, $p = .02$,. The other main effects and interactions had a *p*-value of .21 or above.

For the sensitivity measure *d'*, the main effect of condition was significant, $F(2,80) = 4.23$, $MSE = 0.48$, $p = .02$, $\eta^2 p = .10$. Follow up Tukey post-hoc comparisons found no differences in discriminability between the mole ($M = 0.85$, $SEM = .10$) and Scar ($M = 0.97$, $SEM = .11$) conditions, $t(40) = 1.11$, $p = .51$, nor the unaltered ($M = 1.19$, $SEM = .12$) and scar conditions, $t(40) = 1.72$, $p = .21$, but better discriminability for the unaltered condition compared to the mole condition, $t(40) = 2.86$, $p = .02$. The other main effects and interactions had a *p*-value of .18 or above. Lastly, for the response bias measure *c*, the other main effects and interactions had a *p*-value of .36 or above.

Table 10

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Gender in Experiment 1

		Male Participants				Female Participants			
Condition	Sex of Stimuli	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Unaltered	Male	.55	.21	1.07	.36	.62	.22	1.24	.26
		(.07)	(.04)	(.23)	(.13)	(.04)	(.03)	(.16)	(.09)
	Female	.59	.20	1.21	.32	.62	.22	1.25	.27
		(.06)	(.03)	(.25)	(.10)	(.04)	(.03)	(.16)	(.07)
Mole	Male	.59	.24	1.07	.28	.52	.25	0.81	.34
		(.04)	(.06)	(.14)	(.15)	(.04)	(.03)	(.11)	(.08)
	Female	.48	.29	0.59	.38	.59	.27	0.93	.21
		(.07)	(.06)	(.29)	(.13)	(.04)	(.03)	(.14)	(.07)
Scar	Male	.65	.25	1.20	.19	.56	.24	0.99	.34
		(.05)	(.05)	(.17)	(.14)	(.04)	(.03)	(.14)	(.08)
	Female	.57	.31	0.85	.22	.58	.29	0.86	.20
		(.06)	(.07)	(.19)	(.17)	(.04)	(.04)	(.15)	(.10)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 2 Gender Analysis

Exploratory gender analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 11 for descriptive data). Data were excluded from twelve participants, the four who identified as non-binary and the eight who did not provide their gender; the final analyses were conducted using the male (*n* = 8) and female (*n* = 12) identified participants. All four measures were analyzed using separate 2 (Participant Gender: Male × Female) × 2 (Sex of Stimuli: Male × Female) × 3 (Condition: Unaltered × Mole × Scar) mixed measures ANOVAs. Participant gender was the between-subjects factor while both sex of stimuli and condition were within-subject factors. Overall, no effects of condition, sex, or gender were found, though there were two three-way interactions, one for hits and one for response bias. Neither of these interactions provided any meaningful differences, however.

For the hit rates, the three-way interaction was significant, $F(2,36) = 4.83$, $MSE = 0.04$, $p = .01$, $\eta^2p = .21$, but none of the follow-up Tukey post-hoc comparisons were significant. The interaction lay in small numeric differences leading to several uncorrected *p*-values that reached significance. Most of these differences were not notable, with the most notable being that female participants had greater hits to female faces in the mole condition ($M = .64$, $SEM = .06$) than male faces in the mole condition ($M = .47$, $SEM = .07$), $t(18) = 2.36$, $p = .03$, and that male participants had greater hits to male faces in the mole condition ($M = .59$, $SEM = .08$) than female faces in the mole condition ($M = 0.41$, $SEM = .08$) $t(18) = 2.17$, $p = .04$. The other main effects and interactions had a *p*-value of .11 or above. For false alarm rates, the main effects and interactions had a *p*-value of .10 or above. For the sensitivity measure *d'*, the main effects and interactions had a *p*-value of .15 or above.

Lastly, for the response bias measure *c*, the three-way interaction was significant, $F(2,36) = 4.71$, $MSE = 0.19$, $p = .02$, $\eta^2 p = .21$, but none of the follow-up Tukey post-hoc comparisons were significant. The interaction lay in small numeric differences leading to several uncorrected *p*-values that reached significance but none of the differences were notable. The other main effects and interactions had a *p*-value of .11 or above.

Table 11

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Gender in Experiment 2

		Male Participants				Female Participants			
Condition	Sex of Stimuli	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Unaltered	Male	.64	.47	0.34	-.16	.62	.46	0.32	-.09
		(.06)	(.10)	(.28)	(.13)	(.07)	(.08)	(.28)	(.10)
	Female	.61	.31	0.63	.13	.67	.32	0.70	.02
		(.07)	(.09)	(.29)	(.10)	(.05)	(.05)	(.15)	(.10)
Mole	Male	.59	.48	0.22	-.13	.47	.42	0.11	.17
		(.09)	(.11)	(.36)	(.15)	(.06)	(.06)	(.17)	(.13)
	Female	.41	.34	0.14	.36	.64	.40	0.51	-.04
		(.08)	(.08)	(.31)	(.10)	(.06)	(.08)	(.21)	(.13)
Scar	Male	.45	.39	0.14	.21	.67	.50	0.33	-.24
		(.07)	(.07)	(.14)	(.19)	(.05)	(.05)	(.15)	(.11)
	Female	.58	.47	0.22	-.07	.56	.45	0.23	-.01
		(.09)	(.07)	(.26)	(.12)	(.09)	(.08)	(.30)	(.11)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 3 Gender Analysis

Exploratory gender analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 12 for descriptive data). Data were excluded from seven participants, the three who identified as non-binary and the four who did not provide their gender; the final analyses were conducted using the male (*n* = 3) and female (*n* = 24) identified participants. All four measures were analyzed using separate 2 (Participant gender: Male × Female) x 2 (Sex of Stimuli: Male × Female) x 3 (Condition: Unaltered × Mole × Scar) mixed measures ANOVAs. Participant gender was the between-subjects factor while both sex of stimuli and condition were within-subject factors. Overall, an effect of condition was found, though no gender or sex differences were found nor were there any interactions; so, there is nothing meaning to report for Experiment 3.

For the hit rates, the main effect of condition was significant, $F(2,50) = 4.66$, $MSE = 0.03$, $p = .01$, $\eta^2p = .16$. Follow up Tukey post-hoc comparisons found no differences in hits between the mole (*M* = 0.57, *SEM* = .05) and unaltered conditions (*M* = 0.56, *SEM* = .04), $t(25) = 0.15$, $p = .99$, but greater hits for the distinct condition (*M* = 0.71, *SEM* = .04) compared to both the mole condition, $t(25) = 2.56$, $p = .04$, and unaltered condition, $t(25) = 2.50$, $p = .05$. The other main effects and interactions had a *p*-value of .21 or above. For false alarm rates, the main effects and interactions had a *p*-value of .31 or above.

For the sensitivity measure *d'*, the main effect of condition was significant, $F(2,50) = 8.67$, $MSE = 0.41$, $p < .001$, $\eta^2p = .26$. Follow up Tukey post-hoc comparisons found no differences in discriminability between the mole (*M* = 0.83, *SEM* = .17) and unaltered conditions (*M* = 0.86, *SEM* = .12), $t(25) = 0.20$, $p = .98$, but greater discriminability for the distinct condition (*M* = 1.55, *SEM* = .17) compared to both the mole condition, $t(25) = 3.45$, $p = .01$, and the unaltered condition, $t(25) = 3.69$, $p = .003$. The other main effects and interactions had a *p*-value of .68 or above. Lastly, for the response bias measure *c*, the main effects and interactions had a *p*-value of .18 or above.

Table 12

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Gender in Experiment 3

		Male Participants				Female Participants			
Condition	Sex of Stimuli	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Unaltered	Male	.50	.17	0.99	.50	.55	.28	0.84	.28
		(.14)	(.04)	(.26)	(.27)	(.04)	(.04)	(.14)	(.10)
	Female	.58	.29	0.81	.19	.61	.34	0.82	.08
		(.04)	(.08)	(.17)	(.19)	(.04)	(.04)	(.12)	(.11)
Mole	Male	.54	.27	0.90	.34	.53	.30	0.71	.29
		(.04)	(.18)	(.46)	(.34)	(.04)	(.05)	(.12)	(.11)
	Female	.63	.33	0.83	.07	.58	.29	0.86	.21
		(.13)	(.11)	(.09)	(.33)	(.05)	(.04)	(.19)	(.09)
Distinct	Male	.67	.23	1.33	.18	.73	.24	1.51	.05
		(.11)	(.09)	(.36)	(.30)	(.03)	(.03)	(.16)	(.08)
	Female	.75	.19	1.77	.12	.70	.19	1.59	.20
		(.13)	(.10)	(.47)	(.29)	(.03)	(.03)	(.12)	(.08)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 4 Gender Analysis

Experiment 4a

Exploratory gender analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 13 for descriptive data). Data were excluded from one participant who did not provide their gender; the final analyses were conducted using the male (n = 7) and female (n = 24) identified participants. All four measures were analyzed using separate 2 (Participant gender: Male × Female) x 2 (Sex of Stimuli: Male × Female) x 2 (Condition: Tattooed × Untattooed) mixed measures ANOVAs. Participant gender was the between-subjects factor while both sex of stimuli and condition were within-subject factors. Overall, effects of sex of stimuli were found, for hits, false alarms, and *c*, with more hits, lower false alarms, and a more liberal response to female faces. There was one interaction, with uncorrected *p*-values showing a difference between male and female faces for female participants using *d*'. No other important gender or sex differences were found.

For the hit rates, main effect of sex of stimuli was significant, $F(1,29) = 11.00$, $MSE = 0.02$, $p = .002$, $\eta^2p = .28$, with greater hits to female faces ($M = 0.70$, $SEM = .03$) than male faces ($M = 0.59$, $SEM = .03$). The other main effects and interactions had a *p*-value of .11 or above. For the false alarm rates, the main effect of condition was significant, $F(1,29) = 4.31$, $MSE = 0.02$, $p = .05$, $\eta^2p = .13$, with greater false alarms to untattooed faces ($M = 0.33$, $SEM = .04$) than tattooed faces ($M = 0.25$ $SEM = .04$). The main effect sex of stimuli was also significant, $F(1,29) = 12.36$, $MSE = 0.01$, $p = .001$, $\eta^2p = .30$, with greater false alarms to female faces ($M = 0.34$, $SEM = .04$) than male faces (*M* = 0.24, *SEM* = .04). The other main effects and interactions had a *p*-value of .12 or above.

For the sensitivity measure *d'*, the main effect of condition was marginally significant, $F(1,29) = 3.93$, $MSE = 0.37$, $p = .06$, η^2p = .12, with numerically greater discriminability for tattooed faces (*M* = 1.20, *SEM* = .13) than untattooed faces (*M* = 0.94, *SEM* = .12). Additionally, the Participant Gender × Sex of Stimuli interaction was significant, $F(1,29) = 7.00$, $MSE = 0.28$, $p = .01$, η^2p = .19, but none of the follow-up Tukey post-hoc comparisons were significant. The interaction lay in small numeric differences leading to uncorrected *p*-values that reached significance; the only notable difference was that female participants had better discriminability for female faces (*M* = 1.11, *SEM* = .12) than male faces (*M* = 0.96, *SEM* = .10), $t(29) = 2.42$, $p = .02$. The other main effects and interactions had a *p*-value of .22 or above.

Lastly, for the response bias measure *c*, the main effect of sex of stimuli was significant, $F(1,29) = 17.63$, $MSE = 0.13$, $p < .001$, η^2p = .38, with a more conservative bias toward male faces (*M* = 0.28, *SEM* = .08) than female faces (*M* = -.04, *SEM* = .09). The main effect of condition was marginally significant, $F(1,29) = 3.41$, $MSE = 0.14$, $p = .08$, η^2p = .11, with a more conservative bias toward tattooed faces (*M* = 0.20, *SEM* = .09) than untattooed faces (*M* = 0.05, *SEM* = .08). The other main effects and interactions had a *p*-value of .18 or above.

Table 13

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Gender in Experiment 4a

		Male Participants				Female Participants			
Condition	Sex of Stimuli	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Tattooed	Male	.61 (.08)	.19 (.09)	1.43 (.35)	.40 (.21)	.56 (.03)	.26 (.04)	0.94 (.14)	.29 (.10)
	Female	.71 (.05)	.26 (.06)	1.32 (.25)	.06 (.13)	.67 (.03)	.31 (.04)	1.11 (.15)	.05 (.10)
Un-tattooed	Male	.67 (.06)	.25 (.07)	1.24 (.19)	.14 (.18)	.52 (.03)	.27 (.03)	0.76 (.13)	.32 (.07)
	Female	.69 (.05)	.44 (.10)	0.68 (.29)	-.20 (.21)	.70 (.04)	.33 (.04)	1.10 (.16)	-.05 (.09)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 4b

Exploratory gender analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 14 for descriptive data). Data were excluded from eleven participants, ten who did not provide their gender and one who identified as non-binary; the final analyses were conducted using the male (n = 5) and female (n = 32) identified participants. All four measures were analyzed using separate 2 (Participant gender: Male × Female) x 2 (Sex of Stimuli: Male × Female) x 2 (Orientation: Upright × Inverted) x 2 (Condition: Tattooed × Untattooed) mixed measures ANOVAs. Participant gender was the between-subjects factor while sex of stimuli, orientation, and condition were within-subject factors. Overall, no significant effects of sex of stimuli or participant gender were found.

For the hit rates, the main effect of orientation was significant, $F(1,35) = 4.93$, $MSE = 0.03$, $p = .03$, $\eta^2p = .12$, with greater hits to upright stimuli ($M = 0.60$, $SEM = .03$) than inverted stimuli ($M = 0.54$, $SEM = .04$). The main effect of condition was also significant, $F(1,35) = 4.81$, $MSE = 0.06$, $p = .04$, $\eta^2p = .12$, with greater hits to tattooed stimuli ($M = 0.62$, $SEM = .03$) than untattooed stimuli ($M = 0.53$, $SEM = .04$). Additionally, the Orientation × Condition interaction was also significant, $F(1,35) = 5.90$, $MSE = 0.03$, $p = .02$, $\eta^2p = .14$. Follow-up Tukey post-hoc comparisons showed greater hits to upright untattooed stimuli ($M = .59$, $SEM = .04$) than the inverted untattooed stimuli ($M = .46$, $SEM = .05$), $t(35) = 3.11$, $p = .02$. All other main effects and interactions had a *p*-value of .14 or above.

For the false alarm rates, no main effect or interaction reached significance; all had a *p*-value of .32 or above. For the sensitivity measure *d'*, only the main effect of orientation was significant, $F(1,35) = 4.43$, $MSE = 0.48$, $p = .04$, $\eta^2p = .11$, with greater sensitivity for upright stimuli ($M = 0.77$, $SEM = .10$) than inverted stimuli ($M = 0.52$, $SEM = .11$). All other main effects and interactions had a *p*-value of .15 or above. Lastly, for the response bias measure *c*, only the Orientation × Condition interaction was also significant, $F(1,35) = 5.05$, $MSE = 0.11$, $p = .03$, $\eta^2p = .13$. Follow up Tukey post-hoc comparisons, however, found no significant differences. The interaction lay in small numeric differences leading to uncorrected *p*-values that reached significance; the only notable difference was a more conservative response bias for the inverted untattooed stimuli compared to the inverted tattooed stimuli, $t(35) = 2.31$, $p = .03$. All other main effects and interactions had a *p*-value of .10 or above.

Table 14

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Gender in Experiment 4b

			Male Participants				Female Participants			
Condition	Sex of Stimuli	Orientation	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Tattooed	Male	Upright	.68	.33	1.03	-.06	.58	.32	0.78	.16
			(.11)	(.05)	(.27)	(.23)	(.04)	(.04)	(.15)	(.08)
		Inverted	.67	.40	0.85	-.07	.55	.34	0.61	.18
			(.11)	(.15)	(.27)	(.37)	(.04)	(.04)	(.17)	(.09)
	Female	Upright	.52	.38	0.37	.18	.66	.31	1.05	.05
			(.13)	(.10)	(.47)	(.26)	(.03)	(.03)	(.13)	(.08)
		Inverted	.63	.40	0.65	-.04	.64	.38	0.79	-.04
			(.06)	(.13)	(.19)	(.27)	(.04)	(.04)	(.16)	(.10)
Un-tattooed	Male	Upright	.60	.33	0.73	.09	.56	.32	0.71	.19
			(.09)	(.05)	(.10)	(.19)	(.04)	(.04)	(.14)	(.09)
		Inverted	.40	.30	0.22	.47	.40	.31	.26	.44
			(.05)	(.10)	(.39)	(.32)	(.05)	(.04)	(.14)	(.12)
	Female	Upright	.63	.33	0.82	.06	.58	.34	0.68	.12
			(.03)	(.07)	(.18)	(.14)	(.03)	(.03)	(.12)	(.08)
		Inverted	.52	.40	0.36	.10	.52	.38	0.44	.16
			(.12)	(.09)	(.48)	(.20)	(.03)	(.04)	(.16)	(.09)

Note: Standard Errors are presented in parentheses below their corresponding means.

Appendix F: Exploratory Race Analyses

Experiment 1 Race Analysis

Exploratory race analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 15 for descriptive data). Data were excluded from six participants, who did not provide their race; the remaining 40 participants were categorized either as belonging to the *same-race* as the face stimuli used (Caucasian; *n* = 29) or as *other-race* to the stimuli used (*n* = 11). All four measures were analyzed using separate 2 (Participant Race: Same × Other) x 3 (Condition: Unaltered × Mole × Scar) mixed measures ANOVAs. Participant race was the between-subjects factor while condition was the within-subject factors. Overall, effects of condition were found, while there was only one effect of race found. Same-race participants were found to be more conservative in their responses than other-race participants. This is to be expected as, according to the categorization-individuation model, participants are more motivated to individualize in-group members (Hugenberg et al., 2010). It would be expected, then, that participants would be more careful in their responses at test to faces that they attempted to individualize at study.

For the hit rates, the main effects and interaction had a *p*-value of .10 or above. For the false alarms, the main effect of condition was significant, $F(2,76) = 5.34$, $MSE = 0.01$, $p = .001$, $\eta^2p = .12$. Follow up Tukey post-hoc comparisons found no differences in false alarms between the mole ($M = .28$, $SEM = .03$) and scar ($M = .29$, $SEM = .03$) conditions, $t(38) = 0.85$, $p = .68$), nor between the unaltered ($M = .22$, $SEM = .02$) and mole conditions, $t(38) = 2.16$, $p = .09$, but fewer false alarms to the unaltered condition compared to the scar condition, $t(38) = 3.33$, $p = .01$. The other main effect and interaction had a *p*-value of .35 or above.

For the sensitivity measure *d'*, the main effect of condition was significant, $F(2,76) = 5.87$, $MSE = 0.22$, $p = .004$, $\eta^2p = .13$. Follow up Tukey post-hoc comparisons found no differences in discriminability between the mole ($M = 1.00$, $SEM = .11$) and scar conditions ($M = 0.91$, $SEM = .11$), $t(38) = 0.87$, $p = .66$, better discriminability for the unaltered condition ($M = 1.29$, $SEM = .14$) compared to the scar condition, $t(38) = 3.10$, $p = .01$, and marginally better discriminability for the unaltered condition compared to the mole condition, $t(38) = 2.39$, $p = .06$. The main other effect and interaction had a *p*-value of .11 or above.

Lastly, for the response bias measure *c*, the main effect of participant race was marginally significant, $F(1,38) = 3.68$, $MSE = 0.28$, $p = .06$, $\eta^2p = .09$, with a higher (more conservative) response bias for same-race ($M = .27$, $SEM = .06$) than other-race participants ($M = .06$, $SEM = .09$). The other main effect and interaction had a *p*-value of .68 or above.

Table 15

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Race in Experiment 1

	Same-Race Participants				Other-Race Participants			
Condition	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Unaltered	.62	.21	1.26	.28	.69	.24	1.32	.10
	(.04)	(.02)	(.14)	(.07)	(.05)	(.04)	(.26)	(.09)
Mole	.54	.26	0.83	.30	.69	.29	1.16	.02
	(.03)	(.03)	(.11)	(.06)	(.05)	(.06)	(.17)	(.14)
Scar	.59	.26	1.00	.24	.62	.33	0.82	.07
	(.03)	(.03)	(.12)	(.07)	(.05)	(.04)	(.17)	(.10)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 2 Race Analysis

Exploratory race analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 16 for descriptive data). Data were excluded from eight participants, who did not provide their race; the remaining participants were categorized either as belonging to the *same-race* as the face stimuli used (Caucasian; $n = 17$) or as *other-race* to the stimuli used ($n = 7$). All four measures were analyzed using separate 2 (Participant Race: Same × Other) x 3 (Condition: Unaltered × Mole × Scar) mixed measures ANOVAs. Participant race was the between-subjects factor while condition was the within-subject factors. Overall, no significant differences were found; so, there is nothing meaningful to report for Experiment 2.

For the hit rates, the main effects and interaction had a *p*-value of .33 or above. For false alarm rates, the main effects and interaction had a *p*-value of .57 or above. For the sensitivity measure *d'*, the main effects and interaction had a *p*-value of .32 or above. Lastly, for the response bias measure *c*, the main effects and interaction had a *p*-value of .17 or above.

Table 16

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Race in Experiment 2

	Same-Race Participants				Other-Race Participants			
Condition	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Unaltered	.67	.32	0.75	.01	.59	.37	0.50	.10
	(.05)	(.05)	(.17)	(.08)	(.09)	(.10)	(.37)	(.11)
Mole	.54	.40	0.30	.12	.66	.44	0.48	-.13
	(.05)	(.05)	(.17)	(.07)	(.09)	(.10)	(.34)	(.13)
Scar	.56	.49	0.13	-.08	.59	.41	0.37	-.01
	(.05)	(.05)	(.17)	(.07)	(.08)	(.05)	(.22)	(.11)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 3 Race Analysis

Exploratory race analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 17 for descriptive data). Data were excluded from six participants, who did not provide their race; the remaining participants were categorized either as belonging to the *same-race* as the face stimuli used (Caucasian; $n = 23$) or as *other-race* to the stimuli used ($n = 5$). All four measures were analyzed using separate 2 (Participant Race: Same × Other) x 3 (Condition: Unaltered × Mole × Distinct) mixed measures ANOVAs. Participant race was the between-subjects factor while condition was the within-subject factors. Overall, effects of condition were found, while there was only one effect of race found. Other-race participants were found to have higher false alarms to the stimuli than same-race participants. This finding is not surprising as, as explained previously, participants have fewer false alarms to same-race faces than other-race faces (e.g., Chance & Goldstein, 1996; Meissner & Brigham, 2001).

For the hit rates, the main effect of condition was significant, $F(2,54) = 6.26$, $MSE = 0.02$, $p = .004$, $\eta^2p = .19$; follow up Tukey post-hoc comparisons found no differences in hits between the mole ($M = .58$, $SEM = .04$) and unaltered ($M = 0.63$, $SEM = .03$) conditions, $t(27) = 1.22$, $p = .45$, but greater hits for the distinct condition ($M = 0.73$, $SEM = .03$) compared to the mole condition, $t(27) = 3.36$, $p = .01$, and marginally greater hits compared to the unaltered condition, $t(27) = 2.32$, $p = .07$. The other main effect and interaction had a *p*-value of .23 or above.

For false alarm rates, the main effect of condition was significant, $F(2,54) = 3.74$, $MSE = 0.02$, $p = .03$, $\eta^2p = .12$; follow up Tukey post-hoc comparisons found no differences in false alarms between the mole ($M = .35$, $SEM = .04$) and unaltered conditions ($M = 0.33$, $SEM = .04$), $t(27) = 0.36$, $p = .93$, nor between the distinct ($M = 0.23$, $SEM = .03$) and unaltered conditions, $t(27) = 2.06$, $p = .12$, but fewer false alarms to the distinct condition compared to the mole condition, $t(27) = 2.87$, $p = .02$. The main effect of participant race was marginally significant, $F(1,27) = 4.03$, $MSE = 0.04$, $p = .06$, $\eta^2p = .13$, with greater false alarms for other-race participants ($M = .35$, $SEM = .04$) than same-race participants ($M = .25$, $SEM = .02$). The interaction had a *p*-value of .81.

For the sensitivity measure *d'*, the main effect of condition was significant, $F(2,54) = 13.43$, $MSE = 0.26$, $p < .001$, $\eta^2p = .33$; follow up Tukey post-hoc comparisons found no differences in hits between the mole ($M = 0.67$, $SEM = .13$) and unaltered conditions ($M = 0.86$, $SEM = .12$), $t(27) = 1.14$, $p = .50$, but greater discriminability for the distinct condition ($M = 1.49$, $SEM = .15$) compared to the mole condition, $t(27) = 5.02$, $p < .00,1$ and marginally greater hits compared to the unaltered condition, $t(27) = 3.71$, $p = .003$. The other main effect and interaction had a *p*-value of .31 or above. Lastly, for the response bias measure *c*, the main effects and interaction had a *p*-value of .10 or above.

Table 17

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Race in Experiment 3

	Same Race				Other Race			
Condition	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Unaltered	.57	.27	0.88	.26	.69	.39	0.83	-.08
	(.03)	(.04)	(.11)	(.09)	(.03)	(.07)	(.23)	(.12)
Mole	.54	.29	0.78	.27	.62	.41	0.56	-.02
	(.04)	(.04)	(.12)	(.10)	(.05)	(.07)	(.22)	(.12)
Distinct	.74	.20	1.64	.10	.72	.27	1.33	.07
	(.03)	(.03)	(.14)	(.08)	(.04)	(.06)	(.26)	(.14)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 4 Race Analysis

Experiment 4a

Exploratory race analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 18 for descriptive data). Data were excluded from two participants, who did not provide their race; the remaining participants were categorized either as belonging to the *same-race* as the face stimuli used (Caucasian; *n* = 20) or as *other-race* to the stimuli used (*n* = 9). All four measures were analyzed using separate 2 (Participant Race: Same × Other) x 2 (Condition: Tattooed × Untattooed) mixed measures ANOVAs. Participant race was the between-subjects factor while condition was the within-subject factors. Overall, effects of race were found. Same-race participants had lower false alarms, greater discriminability, and a more conservative response bias for the stimuli. As explained previously, these findings are consistent with the literature (e.g., Chance & Goldstein, 1996; Hugenberg et al., 2010).

For the hit rates, no main effects or interactions had a *p*-value below .14. For false alarm rates, the main effect of participant race was significant, $F(1,27) = 10.60$, $MSE = 0.04$, $p = .003$, $\eta^2 p = .28$, with greater false alarms for other-race faces ($M = .42$, $SEM = .05$) than same-race faces ($M = 0.24$, $SEM = .03$). The other main effect and interaction had a *p*-value of .21 or above. For the sensitivity measure *d'*, the main effect of participant race was significant, $F(1,27) = 4.54$, $MSE = 0.43$, $p = .04$, $\eta^2 p = .14$, with greater discriminability for same-race faces ($M = 1.08$, $SEM = .10$) than other-race faces ($M = 0.68$, $SEM = .15$). The other main effect and interaction had a *p*-value of .27 or above. For the response bias measure *c*, the main effect of participant race was significant, $F(1,27) = 9.33$, $MSE = 0.19$, $p = .01$, $\eta^2p = .26$, with more conservative responses for same-race faces ($M = .25$, $SEM = .07$) than other-race faces ($M = -.13$, $SEM = .10$). The other main effect and interaction had a *p*-value of .26 or above.

Table 18

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Race in Experiment 4a

	Same Race				Other Race			
Condition	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Tattooed	.62	.23	1.18	.28	.66	.39	0.72	-.06
	(.03)	(.04)	(.15)	(.09)	(.03)	(.07)	(.17)	(.14)
Untattooed	.59	.26	0.97	.23	.69	.45	0.64	-.19
	(.03)	(.03)	(.12)	(.08)	(.03)	(.05)	(.12)	(.10)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 4b

Exploratory race analyses were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c* (see Table 19 for descriptive data). Data were excluded from ten participants, who did not provide their race; the remaining participants were categorized either as belonging to the *same-race* as the face stimuli used (Caucasian; $n = 28$) or as *other-race* to the stimuli used ($n = 10$). All four measures were analyzed using separate 2 (Participant Race: Same × Other) x 2 (Condition: Tattooed × Untattooed) 2 (Orientation: Upright × Inverted) mixed measures ANOVAs. Participant race was the between-subjects factor while condition and orientation were the within-subject factors. Overall, no meaningful effects involving race were found, though there was a marginal effect of race in sensitivity. Surprisingly, other-race participants showed slightly better sensitivity than same-race participants which is inconsistent with previous findings and the literature, though, again, the difference did not reach significance.

For the hit rates, the main effect of orientation was significant, $F(1,36) = 6.03$, $MSE = 0.01$, $p = .02$, $\eta^2p = .14$, with greater hits to upright stimuli ($M = 0.60$, $SEM = .03$) than inverted stimuli ($M = 0.60$, $SEM = .02$). The main effect of condition was also significant, $F(1,36) = 11.90$, $MSE = 0.03$, $p = .001$, $\eta^2p = .25$, with greater hits to tattooed stimuli ($M = 0.63$, $SEM = .03$) than untattooed stimuli ($M = 0.52$, $SEM = .03$). Additionally, the Orientation × Condition interaction was also significant, $F(1,36) = 8.17$, $MSE = 0.01$, $p = .01$, $\eta^2p = .19$. Follow-up Tukey post-hoc comparisons showed greater hits to inverted tattooed stimuli ($M = .63$, $SEM = .03$) than the inverted untattooed stimuli

(*M* = .47, *SEM* = .04), *t*(36) = 4.43, *p* < .001. All other main effects and interactions had a *p*-value of .13 or above.

For the false alarm rates, no main effect or interaction reached significance; all had a *p*-value of .11 or above. For the sensitivity measure *d'*, the main effect of orientation was significant, *F*(1,36) = 9.57, *MSE* = 0.20, *p* = .004, η^2p = .21, with greater sensitivity for upright stimuli (*M* = 0.82, *SEM* = .07) than inverted stimuli (*M* = 0.56, *SEM* = .07). The main effect of condition was also significant, *F*(1,36) = 6.72, *MSE* = 0.39, *p* = .01, η^2p = .16, with greater sensitivity for tattooed stimuli (*M* = 0.84, *SEM* = .08) than untattooed stimuli (*M* = 0.54, *SEM* = .09). The main effect of participant race approached significance, *F*(1,36) = 3.36, *MSE* = 0.42, *p* = .08, η^2p = .10, with numerically better sensitivity for other-race stimuli (*M* = 0.80, *SEM* = .10) than same-race stimuli (*M* = 0.58, *SEM* = .06). All interactions had a *p*-value of .28 or above.

Lastly, for the response bias measure *c*, the main effect of condition was marginally significant, *F*(1,36) = 3.86, *MSE* = 0.16, *p* = .06, η^2p = .10, with a numerically more conservative bias for untattooed stimuli (*M* = 0.21, *SEM* = .07) than tattooed stimuli (*M* = 0.07, *SEM* = .06). The Orientation × Condition interaction was significant, *F*(1,36) = 11.01, *MSE* = 0.04, *p* = .002, η^2p = .23. Follow-up Tukey post-hoc comparisons showed a more conservative response bias for inverted untattooed stimuli (*M* = .28, *SEM* = .09) than the inverted tattooed stimuli (*M* = .01, *SEM* = .07), *t*(36) = 3.15, *p* = .02. Lastly, the Orientation × Condition × Race interaction was significant, *F*(1,36) = 4.56, *MSE* = 0.04, *p* = .04, η^2p = .11. Follow up Tukey post-hoc comparisons, however, found no significant differences. The interaction lay in small numeric differences leading to uncorrected *p*-values that reached significance, none of which were meaningful. All other main effects and interactions had a *p*-value of .29 or above.

Table 19

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Race in Experiment 4b

		Same Race				Other Race			
Condition	Orientation	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Tattooed	Upright	.62	.33	0.79	.08	.63	.27	1.04	.18
		(.03)	(.03)	(.10)	(.07)	(.06)	(.05)	(.16)	(.14)
	Inverted	.57	.36	0.58	.11	.70	.37	0.94	-.09
		(.03)	(.04)	(.11)	(.07)	(.05)	(.07)	(.18)	(.15)
Untattooed	Upright	.57	.34	0.64	.13	.59	.31	0.64	.16
		(.03)	(.02)	(.10)	(.06)	(.05)	(.07)	(.12)	(.17)
	Inverted	.47	.36	0.30	.24	.46	.32	0.42	.32
		(.04)	(.04)	(.12)	(.09)	(.07)	(.08)	(.23)	(.17)

Note: Standard Errors are presented in parentheses below their corresponding means.

Appendix G: Distinctiveness Ratings from Experiment 3a

		Male	Female	Total
Overall	Unaltered	3.20	3.33	3.26
Distinctiveness	(*n* = 74)	(0.50)	(0.41)	(0.46)
Ratings	Altered	4.76	4.54	4.65
(*N* = 144)	(*n* = 74)	(0.84)	(0.67)	(0.77)
	Total	3.98	3.94	3.95
		(1.04)	(0.83)	(0.93)
Distinctiveness	Lowest	2.46	2.62	
Rating by	Distinctiveness	2.52	2.76	
Selection	(*n* = 16)	2.64	2.80	
(*N* = 48)		2.66	2.90	
		2.70	2.90	
		2.76	2.94	
		2.80	2.98	
		2.86	2.98	
		2.88	3.00	
		2.88	3.02	
		2.92	3.04	
		2.94	3.06	
		2.94	3.10	
		2.96	3.16	
		2.98	3.16	
		3.04	3.22	
	Total	2.81	2.98	2.89
		(0.17)	(0.16)	(0.18)
	Middle	3.42	3.52	
	Distinctiveness	3.46	3.56	

	(*n* = 16)	3.54	3.62	
		3.56	3.74	
		3.62	3.82	
		3.66	3.82	
		3.70	3.84	
		3.74	3.84	
		3.78	3.84	
		3.80	3.86	
		3.80	3.88	
		3.86	3.96	
		3.98	4.00	
		3.98	4.00	
		4.02	4.04	
		4.06	4.10	
	Total	3.75	3.84	3.79
		(0.20)	(0.17)	(0.19)
	Highest	4.82	4.46	
	Distinctiveness	4.90	4.56	
	(*n* = 16)	4.98	4.64	
		5.04	4.66	
		5.04	4.70	
		5.16	4.78	
		5.28	4.84	
		5.34	4.96	
		5.70	5.06	
		5.80	5.30	
		5.96	5.40	
		5.98	5.48	

	6.14	5.50	
	6.20	5.84	
	6.22	5.90	
	6.36	6.22	
Total	5.56	5.14	5.35
	(0.54)	(0.54)	(0.57)

Note: Standard Deviations are presented in parentheses below their corresponding means.

Appendix H: List of Sources for Tattoo Stimuli

http://drawknife-caecal-multiform.xyz/?u=tpap60a&o=zlbwly0&cid=038810ee-d4e8-421c-bfdb-073a77ab1c31

https://blog.tattoo2me.com/tatuagem-no-pesco%C3%A7o-qual-o-significado-do%C3%AD-e-o-que-fazer-9af344bc37c4

https://br.pinterest.com/pin/486951778449212268/?amp_client_id=CLIENT_ID(_)&mweb_unauth_id=&simplified=true

https://cafemom.com/lifestyle/214871-small-face-tattoos/253186-open_star

https://cafemom.com/lifestyle/214871-small-face-tattoos?epik=dj0yJnU9N3FxcmFORHd3TEZ0NGNVOTJSbnFEVjUtN0lOckxWZWcmcD0wJm49bmJrVDV2YXNCdjg3QnBXdkNCbzZfQSZ0PUFBQUFBR1RzNDln

https://cafemom.com/lifestyle/214871-small-face-tattoos?epik=dj0yJnU9WEZZWURXYm5vbzhURHdQdG55eFZtMjFzc1FVbHlma20mcD0wJm49TmJwMDZJNW80ZF9zMkNqZ0REekhwdyZ0PUFBQUFBR1RzNDNR

https://cafemom.com/lifestyle/214871-small-face-tattoos?epik=dj0yJnU9YzYwUnZQa2plMjdyTjRXaXBhNVkwdk40UldtSmV6aC0mcD0wJm49ZzFsWGd4Q1o3TUFmbkVXWVdUZVRudyZ0PUFBQUFBR1RzNDhr

https://i.pinimg.com/originals/25/fd/aa/25fdaa38fcd8c8978ac20ffa7012c7b9.jpg

https://i.pinimg.com/originals/6b/00/29/6b0029de242376111c13c6a3ea7f20b5.jpg

https://in.thtantai2.edu.vn/tattoos-above-the-eyebrow-5e7t4alu/

https://menshairstylesfix.com/best-medium-length-hairstyles-men-2018/

https://nextluxury.com/mens-style-and-fashion/face-tattoos-for-men/

https://onpointfresh.com/permanent-layer-tattoos-fashion/norman-theuerkorn/

https://outsons.com/amazing-face-tattoos-you-need-to-see/

https://quoters.info/music/lil-peep

https://tattoodi.com/face-tattoos-women/

https://tattooinsider.com/body-tattoos/face-tattoos/

https://tattoos.lovetoknow.com/Facial_Tattoos

https://thestyleup.com/face-tattoos/2/

https://truetattoos.wordpress.com/2016/08/02/colorful-small-stars-tattoo-on-girls-face/

https://wake-me-up-before-you-go.tumblr.com/post/83629787022

https://weheartit.com/entry/274067881

https://women-with-huge-septums.tumblr.com/post/148796364589

https://www.bodytattooart.com/woman-face-tattoo/

https://www.dubuddha.org/face-small-tattoo-and-old-school-neck-tattoos-on-madison-skye/

https://www.eonline.com/news/553046/17-inmates-hotter-than-hot-convict-jeremy-meeks-because-apparently-it-s-feloncrushfriday

https://www.etsy.com/ca/listing/644461216/small-red-roses-set-of-2-roses-temporary

https://www.inkedmag.com/culture/30-face-tattoos-ranked-from-worst-to-best

https://www.loudtv.net/wp-content/uploads/2021/04/aaron-watts.jpeg

https://www.malemodelscene.net/fresh-faces/leo-pride-henrique-resende/

https://www.pinterest.ca/pin/117938083980610115/

https://www.pinterest.ca/pin/231653974569943153/

https://www.pinterest.ca/pin/245586985914788034/

https://www.pinterest.ca/pin/296322850480519721/

https://www.pinterest.ca/pin/330733166394785528/

https://www.pinterest.ca/pin/359091770273999840/

https://www.pinterest.ca/pin/439030663666623966/

https://www.pinterest.ca/pin/455567318533375568/

https://www.pinterest.ca/pin/47428602314092812/

https://www.pinterest.ca/pin/509821620313692343/

https://www.pinterest.ca/pin/510947520221996173/

https://www.pinterest.ca/pin/517702919665881967/

https://www.pinterest.ca/pin/527413806362169749/

https://www.pinterest.ca/pin/575616396104386898/

https://www.pinterest.ca/pin/577938564667689512/

https://www.pinterest.ca/pin/626985579366959913/

https://www.pinterest.ca/pin/687502699373396475/

https://www.pinterest.ca/pin/728527677223841623/

https://www.pinterest.ca/pin/730357264562355512/

https://www.pinterest.ca/pin/861665341211941474/

https://www.pinterest.ca/pin/AS3u6j3Kjpkmk--5QejR5p73DozUodihsJWCbwsn6kp-Cp66pMUwOJ0/

https://www.sinemalar.com/karakter-galeri/5980/nika-boronina/11 https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Men-09.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Men-41.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Men-54.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Men-76.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Men-80.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Women-07.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Women-14.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Women-23.jpg

https://www.tattoomenow.com/tattoo-designs/wp-content/uploads/2020/01/Face-Tattoo-Women-29.jpg

https://www.tattoosforgirl.com/girls-face-tattoos/

https://www.thefashionisto.com/exclusive-2016-ryan-davies-hall/

https://www.the-sun.com/news/210734/teen-girl-with-a-target-tattoo-on-her-face-is-arrested-for-helping-fiance-kill-a-woman-just-hours-after-they-got-engaged/

https://www.thetrendspotter.net/face-tattoos/

https://www.thetrendspotter.net/heart-tattoos-for-men/

https://www.tumblr.com/hannahpixiess/142955197271

https://www.wildtattooart.com/face-tattoos

https://www.wmagazine.com/fashion/milan-fashion-week-fall-2015-best-backstage#51

https://www.yourtango.com/2018316988/pretty-face-tattoos-women

Appendix I: Tattoo Related Questionnaire Responses across Experiments

				E1	E2	E3	E4a	E4b
Do you have any tattoos?	No			30	18	21	18	25
	Yes			13	8	9	13	13
		If yes, how many?[12]		3.15 (4.96)	3.75 (5.42)	3.67 (2.50)	2.69 (2.72)	3.77 (2.28
		If yes, are they generally visible when wearing typical clothing?	No	6	2	3	8	2
			Yes	7	6	6	5	11
Does someone having a tattoo change your perception of them as a person?	No			35	20	27	21	35
	Yes			8	5	2	10	3
		Do you think having a tattoo makes someone fundamentally different than someone who	No	43	24	29	26	36
			Yes	0	1	1	5	2

[12] Overall means are reported; standard deviations are presented in parentheses below their respective means

		does not have a tattoo?						
Thinking specifically about face and neck tattoos, does someone having a face or neck tattoo change your perception of them as a person?	No			21	15	11	11	21
	Yes			21	10	18	20	17
		Again, thinking specifically about face and neck tattoos, do you think having a face or neck tattoo makes someone fundamentally different than someone who does not have a	No	34	20	22	21	30
			Yes	6	3	8	10	8

		face or neck tattoo?					
Why do you think that some people might be critical of tattoos?	Stigma or Deviance		20	15	15	17	24
	Religious or Cultural Differences		12	2	13	7	7
	Not Understood		5	5	0	4	1

Note: Counts of responses are represented unless otherwise noted.

Appendix J: Exploratory Tattoo Analyses

Experiment 4a

Exploratory analyses based on the presence of participants' tattoos were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c*. First, those with visible (n = 5) and non-visible (n =8) tattoos were compared (see Table 20 for descriptive data); all four measures were analyzed using separate 2 (Visible Tattoo: Tattoo Visible × Tattoo Not Visible) x 2 (Condition: Tattooed × Untattooed) mixed measures ANOVAs. Visible tattoo was the between-subjects factor while condition was the within-subject factors. Due to the only differences being found in uncorrected follow-up comparisons for the response bias interaction, both groups were combined and those with tattoos (n = 13) were compared to those without (n = 18); see Table 21 for descriptive data. Data were excluded from one participant, who did not respond to the question; all four measures were analyzed using separate 2 (Tattoo: Tattooed × Not Tattooed) x 2 (Condition: Tattooed × Untattooed) mixed measures ANOVAs. Tattoo was the between-subjects factor while condition was the within-subject factors. No differences were found.

For the comparison within participants with tattoos, for the hit rates, no main effects or interactions had a *p*-value below .31. For the false alarm rates, no main effects or interactions had a *p*-value below .11. For the sensitivity measure *d'*, no main effects or interactions had a *p*-value below .34. For the response bias measure *c*, the Visible Tattoo × Condition interaction was significant, $F(1,11) = 4.75$, $MSE = 0.04$, $p = .05$, $\eta^2p = .30$, but none of the follow-up Tukey post-hoc comparisons were significant. The interaction lay in small numeric differences leading to uncorrected *p*-values that reached significance; the only notable difference was that participants with visible tattoos (*M* = .05, *SEM* = .11) had a more liberal response bias for tattooed stimuli than participants without visible tattoos (*M* = 0.40, *SEM* = .08), $t(11) = 2.62$, $p = .02$. Neither main effect had a *p*-value below .11.

Table 20

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Tattoo Visibility in Experiment 4a

	Tattoo Visible				Tattoo Not Visible			
Condition	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Tattooed	.66	.30	0.96	.05	.55	.20	1.08	.40
	(.08)	(.02)	(.25)	(.10)	(.04)	(.05)	(.25)	(.08)
Untattooed	.65	.25	1.09	.15	.59	.31	0.81	.15
	(.05)	(.04)	(.25)	(.05)	(.05)	(.05)	(.23)	(.09)

Note: Standard Errors are presented in parentheses below their corresponding means.

For the comparison between participants with and without tattoos, for the hit rates, no main effects or interactions had a *p*-value below .32. For the false alarm rates, no main effects or interactions had a *p*-value below .15. For the sensitivity measure *d'*, no main effects or interactions had a *p*-value below .16. For the response bias measure *c*, no main effects or interactions had a *p*-value below .17.

Table 21

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Tattoo in Experiment 4a

	Tattoo				No Tattoo			
Condition	Hit	FA	*d'*	*c*	Hit	FA	*d'*	*c*
Tattooed	.59	.24	1.03	.27	.65	.29	1.12	.15
	(.04)	(.03)	(.18)	(.08)	(.02)	(.06)	(.16)	(.12)
Untattooed	.62	.29	0.92	.15	.64	.33	0.89	.07
	(.04)	(.03)	(.17)	(.06)	(.03)	(.05)	(.11)	(.11)

Note: Standard Errors are presented in parentheses below their corresponding means.

Experiment 4b

Exploratory analyses based on the presence of participants' tattoos were conducted on hits, false alarms, the sensitivity measure *d'*, and the response bias measure *c*. Data were combined for all of those with tattoos (n = 13), as only two of the 13 did not have visible tattoos; data from participants with tattoos were then compared to those without (n = 25); see Table 22 for descriptive data. Data were excluded from ten participants, who did not respond to the question; all four measures were analyzed using separate 2 (Tattoo: Tattooed × Not Tattooed) x 2 (Condition: Tattooed × Untattooed) x 2 (Orientation: Upright × Inverted) mixed measures ANOVAs. Tattoo was the between-subjects factor while condition and orientation were the within-subject factors. Overall, the only differences found were Tattoo × Condition interactions with both false alarms and *d*', though no follow-up comparisons were meaningful.

For the hit rates, the main effect of orientation was significant, $F(1,36) = 11.10$, $MSE = 0.01$, $p = .002$, $\eta^2p = .24$, with greater hits to upright stimuli ($M = 0.60$, $SEM = .02$) than inverted stimuli ($M = 0.54$, $SEM = .02$). The main effect of condition was also significant, $F(1,36) = 8.39$, $MSE = 0.03$, $p = .01$, $\eta^2p = .19$, with greater hits to tattooed stimuli ($M = 0.61$, $SEM = .02$) than untattooed stimuli ($M = 0.53$, $SEM = .03$). Additionally, the Orientation × Condition interaction was also significant, $F(1,36) = 4.95$, $MSE = 0.01$, $p = .03$, $\eta^2p = .12$. Follow-up Tukey post-hoc comparisons showed greater hits to inverted tattooed stimuli ($M = .60$, $SEM = .03$) than the inverted untattooed stimuli ($M = .47$, $SEM = .03$), $t(36) = 3.54$, $p = .01$. All other main effects and interactions had a *p*-value of .41 or above.

For the false alarm rates, the Tattoo × Condition interaction was significant, $F(1,36) = 4.52$, $MSE = 0.03$, $p = .04$, $\eta^2 p = .11$. Follow up Tukey post-hoc comparisons, however, found no significant differences. The interaction lay in small numeric differences leading to uncorrected *p*-values that reached significance, none of which were meaningful. All other main effects and interactions had a *p*-value of .11 or above.

For the sensitivity measure *d'*, the main effect of orientation was significant, $F(1,36) = 9.52$, $MSE = 0.19$, $p = .004$, $\eta^2 p = .21$, with slightly greater sensitivity for upright stimuli ($M = 0.76$, $SEM = .07$) than inverted stimuli ($M = 0.53$, $SEM = .07$). The Tattoo × Condition interaction was also significant, $F(1,36) = 6.56$, $MSE = 0.33$, $p = .02$, $\eta^2 p = .15$. Follow-up Tukey post-hoc comparisons showed no meaningful differences. The main effect of condition was also marginally significant, $F(1,36) = 3.31$, $MSE = 0.33$, $p = .08$, $\eta^2 p = .08$, with sensitivity for tattooed stimuli ($M = 0.73$, $SEM = .07$) greater than for untattooed stimuli ($M = 0.56$, $SEM = .08$). All other main effects and interactions had a *p*-value of .19 or above.

For the response bias measure *c*, the main effect of condition was significant, $F(1,36) = 4.23$, $MSE = 0.16$, $p = .05$, $\eta^2 p = .11$), with a more conservative response bias for untattooed stimuli ($M = .20$, $SEM = .07$) than tattooed stimuli ($M = 0.06$, $SEM = .06$). The Orientation × Condition interaction was also significant, $F(1,36) = 4.97$, $MSE = 0.05$, $p = .03$, $\eta^2 p = .12$. Follow-up Tukey post-hoc comparisons showed a more conservative response bias for inverted untattooed stimuli ($M = .28$, $SEM = .08$) than the inverted tattooed stimuli ($M = .05$, $SEM = .07$), $t(36) = 2.72$, $p = .05$. All other main effects and interactions had a *p*-value of .11 or above.

Table 22

Mean Hit and False Alarm (FA) Proportions, Sensitivity Measure (d'), and Response Bias Measure (c) by Participant Tattoo in Experiment 4b

		Tattoo				No Tattoo			
Condition	Orientation	Hit	FA	*d*'	*c*	Hit	FA	*d*'	*c*
Tattooed	Upright	.63	.39	0.63	-.03	.61	.27	0.98	.18
		(.04)	(.03)	(.12)	(.08)	(.03)	(.03)	(.11)	(.08)
	Inverted	.60	.38	0.64	.04	.61	.36	0.70	.06
		(.04)	(.05)	(.19)	(.11)	(.03)	(.04)	(.11)	(.08)
Untattooed	Upright	.62	.31	0.84	.10	.56	.34	0.60	.16
		(.04)	(.03)	(.14)	(.08)	(.03)	(.04)	(.10)	(.08)
	Inverted	.49	.30	0.57	.31	.45	.38	0.21	.24
		(.06)	(.05)	(.19)	(.12)	(.04)	(.04)	(.12)	(.10)

Note: Standard Errors are presented in parentheses below their corresponding means.