Conditioning Causes Shifts in Group Exemplars' Perceived Prototypicality: Investigating Mechanisms of Stereotype Formation and Change

Scott Allan John Turnbull

B. Psychology (Hons I) (*Newcastle)*

Submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy

School of Psychology,

Faculty of Science and Information Technology

The University of Newcastle, Australia

June 7, 2018

Declaration

1. I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers at the University of Newcastle and at another Australian institution. The work embodied in all chapters of this thesis reflects substantial intellectual input by A/Professor Stefania Paolini and Dr Andrea Griffin, my primary and secondary supervisors at the School of Psychology, the University of Newcastle. Substantial intellectual input into Chapter 3 and 6 of this thesis was provided by Professor David Neumann from Griffith University. The work presented in Study 3.1 and 5.1 reflects joint data collection with Dr Nicholas Harris, a former PhD student at the University of Newcastle. Data from these two studies contributed to Nicholas Harris' PhD thesis at the University of Newcastle; the data reported by Nicholas Harris addressed research questions that are different from those addressed in this thesis. These co-investigators will be named in any research outcome stemming from the research reported in this thesis in an order that reflects their respective contribution, consistent with the Guidelines of the American Psychological Association. Olivia Gritten, Emma Sherwood, Alexandra Allan, Mitchell Woods and Peter Tatnell contributed to data collection for Studies 3.2, 3.3, 5.1 and 6.1; they reported some of these studies data in their honours thesis to test different research questions.

2. The remainder of this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference

has been made in text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968.

Scott Allan John Turnbull

Signature:

Date: 08/06/18

Acknowledgements

I would like to begin by thanking my two PhD supervisors, A/Professor Stefania Paolini and Dr Andrea Griffin. The support, time and effort you provided me have made my PhD journey a wonderful learning experience. In particular your time and effort in providing guidance and feedback during the difficult final stages taught me a lot. Throughout my candidature you provided me with opportunities I never expected to experience, which have guided me on my life journey. I owe you both my deepest gratitude.

I would also like to thank Alex O’Donnell for his guidance and review of my thesis in the final stages. This was a big task and your ability to provide timely and constructive feedback strengthened this thesis greatly. In addition, I would also like to thank Professor David Neumann for his advice, guidance and direction throughout the course of my PhD. Your expertise and review of technical areas gave me confidence in the output produced.

The support my family provided me during my candidature has not gone unnoticed. Without your love, support, encouragement and generosity I would not have made it this far and I thank you for sticking by me.

I owe a great deal of thanks to all my friends who have tolerated my early nights or weeks spent in isolation writing. The motivation to leave my desk and join you has helped me throughout the final stages of my candidature.

Finally, I would like to thank all the wonderful people I have met and worked with during my studies and candidature at the University of Newcastle and beyond, from administrative staff, technical staff and colleagues. The help you provided has always been valued and I hope we stay in touch.

Table of Contents

List of Tables

List of Figures

Abstract

Exemplar perceived prototypicality refers to the goodness of fit between an exemplar and the category prototype (Rosch, 1978). It is key to category activation (Macrae & Bodenhausen, 2000) and generalisation processes (Rothbart & John, 1985). Past research examined the consequences of pairing a negative/anxiety-provoking stimulus with exemplars of stereotyped social categories on perceived exemplar anxiety and exemplar evaluations (Olson & Fazio, 2006; Olsson, Ebert, Banaji, & Phelps, 2005). The present thesis extends this research focus by investigating the effects of associative learning (direct, vicarious; aversive, appetitive) on intergroup categorisation, measured in terms of perceived exemplar prototypicality of Black, White and minimal group exemplars by White participants. This research has the potential to contribute to understanding stereotype formation and change and to help design interventions that increase social integration in society.

Chapter 1 reviews the stereotyping and intergroup categorisation literature, focusing on stereotyping and exemplar perceived prototypicality. Chapter 2 starts by introducing evaluative-fit and emotion-fit mechanisms as possible psychological underpinnings of prototypicality shifts under conditions of pairing group exemplars with valence and emotion. Associative learning is then described as a way to affect evaluative-fit and emotion-fit and, as a result, cause shifts of exemplar prototypicality. In three successive studies, Chapter 3 provides some initial evidence that an outgroup face (conditioned stimulus, CS) paired with an aversive stimulus (unconditioned stimulus, US; i.e., unsafe exemplar), in the form of an uncomfortable electrotactile stimulation, is perceived as being more prototypical of the outgroup following both direct and vicarious aversive conditioning. Chapter 3 also explores the properties of this basic categorisation or

prototypicality shift effect: After establishing a self-reported, post-extinction, CS-specific and generalised prototypicality shift following both direct and vicarious conditioning (Study 3.1), Study 3.2 shows prototypicality shifts on an implicit measure only post-extinction. Study 3.3 shows that backward CS-masking during conditioning eliminates CS-specific prototypicality shifts and replaces it with a generalised (CS-nonspecific) prototypicality shift towards the entire outgroup. Exploring the causal role of extinction in more depth, Chapter 4 reveals that repeated presentations of ethnic cues, rather than repeated presentation of the target face per se, are sufficient to change how that exemplar is categorised. Using a minimal group procedure, Study 5.1 tests, and confirms the prediction that aversive associative learning should cause safe exemplars (i.e. those not paired with an aversive US) to be perceived as being more prototypical of the ingroup after conditioning. In contrast, Study 5.2, which explores group membership along ethnicity lines, unexpectedly reveals that both safe and unsafe exemplars tended to shift away from the ingroup prototype and become less ingroup-like. Drawing upon theories of evaluative fit, Study 6.1 attempts to test the prediction that pairing outgroup faces with positivity should make them become less representative of the outgroup whereas those paired with negativity should become more representative. A newly designed gambling game paradigm sought to test this prediction; this returned non contingent-specific prototypicality shifts when the exemplar was paired with a negatively valenced stimulus. Chapter 7 summarises the results of the empirical chapters and discusses the implications and limitations of this research, including some key methodological challenges. The research reported in this thesis provides initial evidence that associative learning shapes the goodness of fit between stimuli from stereotyped social categories and their category. Hence, the evidence presented suggests that associative learning does not only change exemplar evaluations and affect but the robustness of the effect and underlying mechanisms involved need to be further investigated. Given that perceived prototypicality influences category activation and affects how stereotypes form, change, and are applied, the research reported in this thesis provides a unique and novel perspective on ways in which problematic intergroup relations can develop and possibly change.

Chapter 1: Stereotypes and Intergroup Categorisation

Research presented in this thesis investigates shifts of exemplars' perceived prototypicality. Exemplar prototypicality is a key dimension of social categorisation and stereotyping and is an important determinant of category activation (Blair, Judd & Chapleau, 2004; Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000). Prototypicality is defined as the goodness of fit between an exemplar and a category prototype (Corneille & Judd, 1999; Rosch, 1978). Exemplar prototypicality is investigated because it is a marker of psychological inclusion, or exclusion of an exemplar from a group (Bless & Schwarz, 1998; Corneille & Judd, 1999; Wyler, Sadler & Judd, 2002). In Chapter 1 I describe the mechanisms that underlie stereotyping, including the consequences, structure, activation and function of stereotypes.

Stereotypes and Stereotyping

Stereotypes are summarised attributes, characteristics and behaviours that represent an overarching perception of a group and its group members (Hamilton & Sherman, 1994; Hilton & Von Hippel, 1996; McGarty, Yzerbyt & Spears, 2002; Schneider, 1991). Stereotypes are dependent on perceiver's expectations of a group based on previous experiences and knowledge (McGarty et al., 2002), and are not necessarily an accurate representation (Jussim, Cain, Crawford, Harber & Cohen, 2009). African American exemplars are predominantly referred to in this thesis, due to the experimental material of choice and a centrality in the (American-led) empirical literature on stereotyping that provides the background literature to this work. Traditionally this group's stereotype consists of physical qualities such as dark skin and attributes such as 'being lazy', and 'excelling at sports'.

Stereotyping is the process where a stereotype is applied to an individual without necessarily obtaining information or interacting with the individual (Fiske, 2000). For example, stereotyping is judging a Black exemplar as being lazy (a Black stereotype) without knowing anything about that exemplar (Donders, Correll & Wittenbrink, 2008). In this section I discuss stereotypes and stereotyping in relation to this thesis and my empirical work. Consequences and effects of stereotypes are discussed first. Secondly, I explain stereotype structure, formation, activation and generalisation. I conclude by explaining the psychological and social functions of stereotypes.

The Consequences and Effects of Stereotyping

Stereotypes bias attention and this causes stereotype confirming information to be over emphasised and stereotype disconfirming information to be disregarded (Bastian & Haslam, 2007; Fiske & Neuberg, 1990; Plaks, Stroessner, Dweck & Sherman, 2001; Stangor & Ford, 1992). Selective attention involves attending to and processing information that is consistent with a stereotype, in comparison to information that disconfirms a stereotype (Plaks et al., 2001). Attention and over emphasis on stereotype confirming information cause stereotypes to strengthen, resulting in continued use of intergroup perceptions and judgements (Sherman, Macrae & Bodenhausen, 2000). Attention to stereotype consistent information occurs because information is easy to comprehend and aligns with existing cognitions (Fiske & Neuberg, 1990; Stangor & Ford, 1992). As such, selective attention aligns with the cognitive heuristic function of stereotyping: The cognitive heuristic limits the amount of information a perceiver needs to process, saving time and energy (Macrae, Milne & Bodenhausen, 1994).

Allen, Sherman, Conrey and Stroessner's (2009) research, for instance, is consistent with the idea that selective attention to stereotype confirming information is used to limit demand on cognitive resources. Allen et al. placed one group of participants under cognitive load and another group under no cognitive load. Stereotype consistent and inconsistent information was read by participants, after which stereotyping was measured. Participants, who had limited cognitive resources, processed stereotype consistent information more than inconsistent information. Thus, research indicates information is selectively attended to and retrieved confirming stereotypical beliefs (Allen et al., 2009; Kunda, 1990; Stangor & Ford, 1992).

Stereotyping is problematic because a person is judged on the basis of their stereotype and not on the basis of their personal characteristics and traits. Riach and Rich (2006) demonstrated how stereotypical judgements caused discrimination in workplaces. Two resumes were created that contained identical experience and qualifications, differing only in the gender of applicant's name. Four types of jobs were applied for using the created resumes: an engineer (stereotypically male job), a secretary (stereotypically female job) and two gender neutral roles (accountant and computer analyst). Discrimination was measured via an invitation or rejection to attend an interview. Discrimination occurred when one applicant received an interview invitation and the other did not, despite the two applicants having equivalent qualifications. No discrimination occurred if both received an invitation to an interview and a non-observation was recorded if neither applicant received an invitation to an interview. Riach and Rich results indicated clear discrimination based on gender: In stereotypical male jobs women were rejected from an interview 46% of the time, whilst men were rejected only 23% of the time. In stereotypical female jobs, men were rejected from an interview 59% of the time and women only 16% of the time. Thus, applicants were judged based on their gender rather than their skill set and discrimination occurred. This effect is consistent across workplaces around the world (Riach, 2015) and not limited to gender, but extends to race, religion and age (Bertrand & Mullainathan, 2004; King & Ahmad, 2010).

Another consequence is the influence stereotypes have on behaviour. A common African American stereotype in American society is that this group consists of criminals and should be feared (Welch, 2007). In an extreme real world example, Correll, Park, Judd and Wittenbrink (2002) demonstrated the devastating behavioural consequences stereotypes have in a shoot-don't shoot task. Images of various backgrounds first appeared on a screen: Either a White or Black exemplar appeared on the screen holding a gun or a safe object. If the exemplar was holding a gun participants were required to select the 'shoot' key as quickly as possible and select a 'don't shoot' key if a safe object was being held. Participants were faster at shooting Black armed exemplars than armed White exemplars. Additionally, participants were more likely to mistakenly shoot unarmed Black exemplars compared to unarmed White exemplars. Quicker decisions were made to not shoot a White exemplar than a Black exemplar. Correll, Park, Judd and Wittenbrink (2007) followed up the earlier research by Correll et al., (2002) and found a similar pattern of results. Furthermore, follow up research showed speed and tendency to shoot Black exemplars more than White exemplars was associated with a 'criminal stereotype'. News stories were read by participants before completing the same shoot-don't shoot task. Participants read a news story describing a string of armed robberies. In one condition suspects were described and sketched as being Black males. In the other condition the same information was presented but White males were described and sketched as suspects. When exposed to stereotypical information (Black = criminals), participants were quicker and more likely to mistakenly shoot Black exemplars than White exemplars. However, when exposed to counter stereotypical information (White = criminals), the effect to mistakenly shoot Black over White exemplars disappeared. Hence, the juxtaposition of stereotypical information exacerbated the effect to mistakenly shoot Black exemplars more than White exemplars and suggests that the salience of the stereotype underpins this effect. Whilst shooting is an extreme example of behaviour, it is representative of hostilities towards stereotyped groups and has become more common and reported in recent years (see Mclaughlin, 2016; McCoy, 2014; Swaine, Laughland, Lartey & McCarthy, 2016; Zielinksi, 2015).

Stereotyping has also the consequence of affecting decision making. Decision making based on stereotypes can cause imprecise and over exaggerated decisions to be made and this is most evident in criminal sentencing. Rehavi and Starrs' (2014) recent analysis of the United States of America Bureau of Justice Statistics data show a disparity between sentences handed to White and Black males. African American males received sentences 10 percent longer than White American males who committed similar crimes. Increased sentences for Black males compared to White males when similar crimes were committed is supported by research (see Everett & Wojtkiewicz, 2002; Johnson, 2003; Mitchell, 2005; Mustard, 2001).

A prominent factor that exacerbated sentence length in these investigations was how prototypically Black the defendant was. Research by Blair and colleagues shows inmates who have more Afrocentric features receive harsher sentences compared to inmates with less Afrocentric features despite committing similar crimes (Blair et al, 2004; Blair, Judd, Sadler & Jenkins, 2002). Thus, greater perceived prototypicality of the Black exemplar relative to the Black group resulted in a stronger association with the negative Black stereotype.

Together, the body of research described in this section shows some of the negative consequences that stereotypes have in society. Salient stereotypes cause judgements, behaviours and decisions to be inaccurate and have severe consequences for individuals and groups, such as not hiring the right person, shooting an innocent person, or sentencing someone to a harsh prison sentence. Blair and colleagues (2002; 2004) show that perceived exemplar prototypicality is a key factor in how consequential a stereotype is. Hence, there is a strong link between an exemplar's perceived prototypicality and stereotyping consequences. Understanding how perceived exemplar prototypicality changes provides an opportunity to modify stereotypes and alter their negative consequences that underpin problematic intergroup relations. Next I describe how stereotypes are formed, structured and their content activated and generalised to others.

Stereotype Formation, Structure, Activation and Generalisation

Research on stereotype formation and structure investigates how group schemas are formed and stored/structured in memory (Hamilton & Sherman, 1994; Hilton & Von Hippel, 1996). Stereotype activation and generalisation occurs when group representations are retrieved and used to form an opinion of new exemplars or process some stereotype-related information (Brown & Hewstone, 2005; Dovidio, Gaertner & Kawakami, 2003). Two dimensions are involved in stereotype formation and structure: central tendency and variability.

Central tendency in a group representation refers to the common features and traits that define a group (Dijksterhuis & van Knippenberg, 1999; Linville & Fischer, 1993; Park & Hastie, 1987). High central tendency of a feature or trait means that the feature or trait is regarded as being common or representative of a group (e.g., African Americans have Black skin). Low central tendency of a feature or trait is when a feature or trait is uncommon and not representative of a group (i.e., African Americans have white skin). A group's central tendency may form from continued associations. When a feature or trait is repeatedly paired with an exemplar it causes stereotypical perceptions to form and the feature or trait becomes more associated with the group. Hence, common features and traits make up a group's central tendency. Prototypicality, the key outcome measure used in this thesis, relates to central tendency because increases in exemplar prototypicality mean that the exemplar is perceived as being closer to a group's central tendency.

Perceived group variability is the perceived variance or spread around the group's central tendency. It is how different or heterogeneous a group is perceived to be (Brauer & Er-rafiy, 2011; Hewstone & Hamberger, 2000; Park & Judd, 1990; Voci, 2000). High variability indicates that exemplars within a group are seen to be very different, whereas low variability indicates that exemplars are seen to be very similar. Perceived variability plays a role in stereotype formation because it helps to determine which traits become associated with a group's stereotype. When a trait has low variability among the group it indicates that the trait is common in group exemplars, resulting in closer association with the central tendency of the group. When a trait has high variability among the group it indicates that the trait is not common amongst group exemplars and is likely to be excluded from the group stereotype (Rubin, Hewstone & Voci, 2001). Prototypicality relates to variability because an exemplar that displays traits that are similar to the group stereotype will be perceived as being more prototypical of the group than if they display traits with high variability.

Involved in stereotype structure and activation is the meta-contrast principle (Turner, 1985; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). The meta-contrast principle involves the comparison of two groups. Two groups are compared based on similarities within groups and differences between groups. Under these conditions, similarities within groups and differences between groups are perceptually accentuated, which forms well defined groups with clear boundaries (Deffenbacher, Park, Judd & Correll, 2009; Hogg, 2000). Meta-contrast causes group exemplars to be clustered around the central tendency with little variability between group exemplars when two groups are compared.

Tajfel and Wilkes (1963) demonstrate the meta-contrast effect. They asked participants to estimate category exemplars (i.e., lines) along a descriptive feature (i.e., length). Participants were randomly allocated into three groups: a ‘classified’ group, which categorised lines according to line length; A ‘randomly classified’ group, which categorised lines randomly; And an ‘unclassified’ group, which did not categorise lines. ‘Classified’ group participants perceived the difference between the longest of the short lines and shortest of the long lines as being greater in length than the ‘randomly classified’ and ‘unclassified’ groups. Hence, group differences were perceptually accentuated and moved towards central tendency (see also Corneille, Klein, Lambert & Judd, 2002). This effect resulted in a change of perceived prototypicality and lines were perceived as being closer to the group prototype.

Central to meta-contrast is the grouping of items into discrete categories, which is at the foundation of intergroup categorisation. Intergroup categorisation is a process that extracts individual information, for example physical and behavioural qualities, to form representations of social groups based on similar qualities (Bodenhausen, Kang & Peery, 2012; Hugenberg, & Sacco, 2008). Once established, intergroup categorisation processes compare new individuals to category prototypes to determine if the new individual fits within the category boundary and can be categorised into that group

(Bodenhausen & Peery, 2009; Hilton & Von Hippel, 1996; Macrae & Bodenhausen, 2000).

Intergroup categorisation provides a structured outlook of social environments, but is flexible and subject to change (Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000). By allowing for flexibility, intergroup categorisation can adapt to categorise stimuli that belong to multiple groups and in different contexts (Crisp & Hewstone, 2007). Thus, intergroup categorisation is malleable. As perceived exemplar prototypicality is an important factor in determining whether an exemplar is categorised in, or excluded from a category (Corneille & Judd, 1999, Cikara & Van Bavel, 2014; Rosch, 1978), changes in perceived exemplar prototypicality are likely to influence intergroup perceptions and generalisation.

Member-to-group generalisation, or simply generalisation, occurs when stereotypical perceptions extend from one group exemplar to all group exemplars (Stark, Flache & Veenstra, 2013). This type of generalisation is a form of inductive reasoning. During inductive reasoning opinions about a group are formed based on experiences with individual category exemplars (Kruglanski & Thomson, 1999; Nesbitt, Peng, Choi & Norenzayan, 2001). Within the stereotyping literature, this type of inductive reasoning is called member-to-group generalisation because it refers to generalising an experience from an individual group member or exemplar of a category to stereotypical judgments about their entire group or category (Rothbart & John, 1985; Rothbart, Sriram, & Davis-Stitt, 1996; Sherman, 1996).

Central to this thesis is the role exemplar typicality has in member-to-group generalisation. Rothbart and John (1985) argued that typical group exemplars should be more influential at changing group perceptions compared to atypical group exemplars (Rothbart & Lewis, 1988; Wilder, Simon, & Faith, 1996). This is because typical group exemplars have stronger cognitive connections to a group and stronger connections result in more robust and readier generalisation. Atypical exemplars, instead, have weaker connections to a group and are unlikely to activate a group stereotype, resulting in weaker generalisations (Dunsmoor & Murphy, 2014; Johnston & Hewstone, 1992; Rothbart & John, 1985; Rothbart & Lewis, 1988). For example, Dunsmoor and Murphy (2014) conditioned participants to fear typical (i.e., sparrows) and atypical exemplars (i.e., penguins) of a group. Both typical and atypical exemplars were feared more after conditioning. However, fear generalised to other category exemplars only when conditioning occurred with typical exemplars. Hence, prototypical exemplars influence stereotype development and activation more than atypical exemplars (Brown & Hewstone, 2005).

Deductive reasoning is another form of generalisation that differs from inductive reasoning because information generalises from the group to an exemplar, rather than from an exemplar to the group. During deductive reasoning overarching information about a group is used to form an impression of individual group members (Wason & Johnson-Laird, 1972). In stereotyping, this process is called group-to-member generalisation (Chen & Ratliff, 2015; Gawronski & Quinn, 2013). Stereotypes associated with a group are used to describe and make inferences about individual group exemplars (Ratliff & Nosek, 2011).

Similar to inductive reasoning, prototypicality is important in deductive reasoning. In the work by Blair and colleagues (2002; 2004) prototypical Black exemplars, a group typically associated with danger, were more likely to receive harsher sentences compared to atypical exemplars. Hence, prototypical exemplars, who have stronger connection to their group, result in stronger stereotypical perceptions being activated compared to atypical exemplars.

As prototypicality is an important variable in activating stereotypical perceptions, in driving member-to-group generalisation and group-to-member generalisation, changes in prototypicality have the potential to affect intergroup relations in society. Thus, a greater understanding of the processes that cause prototypicality shifts can potentially help shape intergroup experiences and relations for the better.

Models of Stereotype Representation and Activation

In this section, I discuss models that theorise how stereotypes are stored in memory. This discussion is centred on how prototypicality fits within the prototype, exemplar and mixed models because this thesis investigates how first-hand experiences with group exemplars affect exemplar prototypicality. Other forms of contact such as vicarious contact are briefly discussed as this type of contact has also relevance in this thesis.

Prototype models theorise stereotypes as consisting of a collection of key characteristics and features that are common in all group exemplars and form a group prototype (Cantor & Mischel, 1979; Minda & Smith, 2001; Minda & Smith, 2002; Minda & Smith, 2011; Rosch, 1975; Rothbart & John, 1985; Smith & Minda, 2001). Simply, in these models, the prototype is the group's central tendency or averaged representation. Stereotype activation from a prototype model perspective is best explained by the family resemblance concept (Rosch & Mervis, 1975). When an exemplar has one, and more often several, features in common with the group prototype, it is considered to have family resemblance (Rosch & Mervis, 1975). When an exemplar's features overlap more with one group prototype than another group prototype, the exemplar is categorised into the group with more overlapping features (Fiske & Neuberg, 1990; Operario & Fiske, 2001). The more features an exemplar has in common with the group prototype, the more prototypical it is considered to be. Hence, in a nutshell, prototype models explain how prototypical representations are stored.

Research using reverse correlation tasks support the idea that prototypes are at the core of group representations (Dotsch, Wigboldus, Langner & van Knippenberg, 2008; Dotsch, Wigboldus & van Knippenberg, 2011; Dotsch, Wigboldus & van Knippenberg, 2013; Imhoff & Dotsch, 2013; Ratner, Dotsch, Wigboldus, van Knippenberg & Amodio, 2014). Reverse correlation tasks are a data driven method to visualise internal representations or prototypes. In a reverse correlation task, a base image has random noise superimposed to create one image, and another image is created by inversing pixel noise. Thus, two images are created from the same image using random noise (Dotsch & Todorov, 2012). Two images are presented repeatedly (different random noise/inverse noise applied to images) and participants select which face fits a target category better. Subsequent judgements progressively converge into a specific face that is perceived as being prototypical, or the prototype of the target category.

Figure 1 displays steps used by Dotsch et al. (2008) to create visual representations of prototypes. In section a, a base face is selected. The left hand side shows the base image with pixelated noise applied, and the right hand side shows the inverse pixel noise. Subsequent judgments are aggregated, in this example, to create a Moroccan prototypical representation, created using reverse correlation task. The Moroccans group is an outgroup to Dutch participants, the sample utilised by Dotsch et al. (2008). Highly prejudiced participants in their research perceived this face as more representative of criminals and not trustworthy. On the right hand side of Figure 1, section c, the Chinese prototype face is shown. The same procedure to identify the

Moroccan prototype was used but differed in the target categorisation group (Moroccan vs Chinese). Thus, through subsequent presentations and categorisation decisions an average or prototype representation for a target group is created in this type of studies. This process is expected to mimic spontaneous processes of group prototype formation.

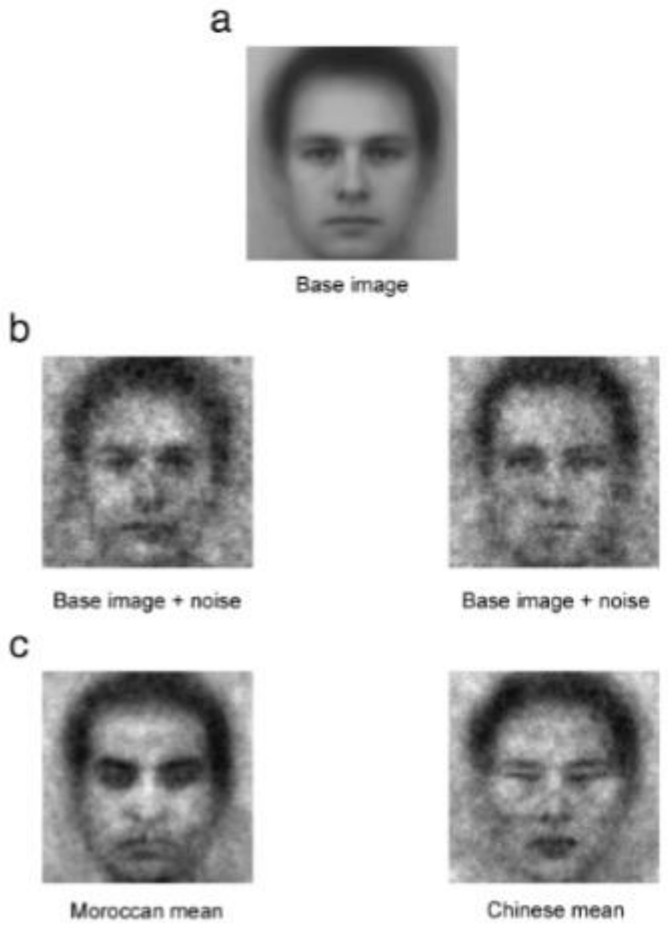


Figure 1. (a) example of base image, (b) example of noise applied to base image, (c) classification image from subsequent presents for Moroccan and Chinese prototype (taken from Dotsch et al., 2008).

In contrast to prototype models, exemplar models theorise that a newly encountered individual is compared against different exemplars rather than a central prototype (Hamilton & Sherman, 1994; Hilton & von Hippel, 1996; Kruschke, 2011). A new exemplar is categorised into a group by being compared to existing group exemplars. If a new exemplar is more similar to one group of exemplars than another group of exemplars, the most similar group's stereotypes are activated (Kruschke, 2011; Nosofsky, Kruschke & Mckinley, 1992; Zaki, Nosofsky, Stanton & Cohen, 2003). Hence, new individuals are compared to existing exemplars when activating stereotypes (Hilton & von Hippel, 1996; Kruschke, 2001).

Blended or mixed models have also been proposed and theorise stereotypes as consisting of both prototypes and exemplars (Hamilton & Mackie, 1990; Hilton & Von Hippel, 1996; Klein, Loftus, Trafton & Fuhrman, 1992; Lech, Gunturkun & Suchan, 2016). Mixed models assume that the category prototype is abstracted by participants and stored in memory along with category exemplars, which is why these models are called blended or mixed models (Medin, Altom & Murphy, 1984). A new exemplar is compared to the group prototype and existing group exemplars to be categorised into a group (Smith & Zarate, 1990; 1992). Similarly, category activation occurs by comparing the new exemplar to the prototype of a group, as well as to the exemplars that are stored in memory. Stereotypes are activated when a threshold is reached either through a match with a prototype or group of exemplars (Lech et al., 2016; Smith & Zarate, 1990; 1992).

The taxonomy of models presented theorises different mechanisms involved in stereotype storage and activation. Prototype models store stereotypes as a prototype that new exemplars are compared to (Minda & Smith, 2011; Rosch, 1975; Rothbart & John, 1985). Stereotypes are stored through group exemplars in exemplar models and stereotype activation results in a comparison with group exemplars (Kruschke, 2011; Nosofsky et al., 1992; Zaki et al., 2003). Blended or mixed models use both prototype and exemplar model assumptions to store and activate stereotypes (Smith & Zarate, 1990; 1992).

The three models discussed previously explain the processes that are used to store and maintain stereotypes through direct or first-hand contact with outgroup members. Other forms of contact with group members are possible (e.g., talking to a friend about a group yet to be encountered) and the social learning theory provides one example that explains alternative processes used to store and maintain stereotypes (i.e., Bandura, 1977). Vicarious contact, which is contact with others through observing an interaction between an ingroup and an outgroup member, for example, provides an opportunity to learn stereotypes without having direct contact (Mazziotta, Mummendey & Wright, 2011; Miller & Dollard, 1941). Vicarious contact also influences stereotypical representation and is investigated in my research.

Exposure to exemplars forms the basis for prototypical representations, which provide a summarised version of similar characteristics and features a group consists of. In this thesis I present exemplars with and without a negative or positive reinforcer to form associations. I investigate how associations with positive/negative stimuli shift perceived exemplar prototypicality and, as a result, possibly change how stereotypes are structured and activated. Next I discuss the cognitive and social functions of stereotypes and how these functions are involved in stereotype formation.

The Psychological Functions of Stereotypes

A variety of psychological functions have been identified and discussed in the literature. The cognitive function of stereotypes is to simplify a complex world that consists of extensive information richness (Cloutier, Mason & Macrae, 2005; Macrae et al., 1994; McGarty et al., 2002; Rothbart & John, 1985). Allport (1954) first proposed that humans are limited in the amount of information they can process and the function of stereotypes was to streamline information processing. Information is processed quicker and easier when a heuristic, schema or summary about a group of exemplars is created (Brubaker, Loveman & Stamatov, 2004; McGarty et al., 2002; Rothbart, Fulero, Jensen, Howard & Birrell, 1978; Wyer & Srull, 1986). Hence, stereotypes as a heuristic save energy and time.

Cognitive efficiency results in stereotype formation because group information is applied to individual group exemplars (Macrae et al., 1994; Rothbart et al., 1978; Wyer & Srull, 1986). For example, a new individual is encountered that belongs to the African American group. This group's cognitive schema will be activated by the encounter and the African American stereotype will summarise the new exemplar without using much additional cognitive resources.

The social function of stereotypes is relevant to motivational and rationalisation processes that explain how social groups are perceived and group ideologies maintained (Rutland & Brown, 2001; Sinclair & Kunda, 2000; Tajfel, 1981). From this stance, people are motivated to explain and justify social relations in intergroup contexts (Yzerbyt, Rocher & Schadron, 1997). Central to this thesis are social identity and self-categorisation theory, which theorise about the motivational processes that underlie group perception (Hogg, 2000; Jost & Banaji, 1994; Jost, Banaji & Nosek, 2004). Social identity and self-categorisation theories propose that individuals are motivated to protect the positive valence associated with the ingroup. To protect positive regard for the ingroup, individuals associate positivity with the ingroup, and in doing so bolster individual self-esteem because of their membership within the group (Brewer & Silver, 1978; Tajfel, Billig, Bundy & Flament, 1971). In a similar vein, the psychological superiority of the ingroup over the outgroup can cause negative outgroup stereotypes and evaluations (Aberson, Healy & Romero, 2000; Tajfel et al., 1971). Stereotypes have social functions in the sense that they satisfy motivations to enhance the ingroup image and maintain social hierarchy (Jost et al., 2004; Rutland & Brown, 2001; Sinclair & Kunda, 2000).

There is research consistent with social identity and self-categorisation theoretical perspective of stereotypes and their social functions (Doosje, van den Bos, Loseman, Feddes & Mann, 2012; Ellemers, Spears & Doosje, 2002; Wirtz, van der Plight & Doosje, 2016). For instance, when intergroup threat is high, self-enhancement motives cause stronger ingroup identification that result in outgroup derogation (Cakal, Hewstone, Guler & Heath, 2016). Outgroup derogation assists the lowering of outgroup status and the enhancement of ingroup status and individual self-esteem. Hence, in line with social identity and self-categorisation theory, positivity is typically associated with the ingroup and this positivity boosts an individual’s self-esteem as a member within the ingroup.

In summary, cognitive and social functions of stereotypes contribute to explain why stereotypes are formed and used in society. Cognitively, stereotypes simplify a complex world and provide a schema that summarises a group to save processing time and energy (Brubaker et al., 2004; McGarty et al., 2002; Rothbart et al., 1978; Wyer & Srull, 1986). Socially, stereotypes provide order to a social world in which individuals are motivated to be associated with a positive group to boost individual self-esteem (Hogg, 2000; Rutland & Brown, 2001; Sinclair & Kunda, 2000; Tajfel, 1981). Changes in perceived exemplar prototypicality provide an opportunity to modify stereotypes and their negative consequences that result in problematic intergroup relations in society. Therefore, understanding how stereotypes function and form provides the foundation to my analysis of the conditions and mechanisms governing changes in exemplar perceptions.

References

Aberson. C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review, 4* (2), 157-173.

Allen, T. J., Sherman, J. W., Conrey, F. R & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology, 45* (5), 1081-1087.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Bandura, A. (1977). *Social learning theory.* New York: General Learning Press.

Bastian, B., & Haslam, N. (2007). Psychological essentialism and attention allocation: Preferences for stereotype-consistent versus stereotype-inconsistent information. *The Journal of Social Psychology, 147* (5), 531-541.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991-1013.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C (2002). The role of Afro- centric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*, 5-25.

Bless, H., & Schwarz, N. (1998). Context effects in political judgement : Assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology*, (28), 159–173.

Bodenhausen, G. V, Kang, S. K., & Peery, D. (2011). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE Handbook of Social Cognition* (pp. 318–336.). India: Replika Press.

Bodenhausen, G. V, & Peery, D. (2009). Social categorization and stereotyping in vivo : The VUCA challenge. *Social and Personality Psychology Compass*, *2*(3), 133–151.

Brauer, M., & Er-rafiy, A. (2011). Increasing perceived variability reduces prejudice and discrimination. *Journal of Experimental Social Psychology*, *47,* 871-877.

Brewer, M. B., & Silver, M. (1978). Ingroup bias as a function of task characteristics. *European Journal of Social Psychology, 8*, 393-400.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, Vol 37* (pp.255-342). San Diego: Elsevier.

Brubaker, R., Loveman, M., & Stamatov, P. (2004). Ethnicity as cognition. *Theory and Society, 33*, 31-64.

Cakal, H., Hewstone, M., Guler, M., & Heath, A. (2016). Predicting support for collective action in the conflict between Turks and Kurds: Perceived threats as a mediator of intergroup contact and social identity. *Group Processes and Intergroup Relations, 19*(6), 732-752.

Cantor, N., & Mischel, W. (1979). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality, 13*,187-205.

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science, 9* (3), 245-274.

Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black-White biracial group members. *Social Psychological and Personality Science, 6* (5), 544-550.

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology, 88* (6), 885-894.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*, 314-1329.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology, 37*, 1102-1117.

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. *Advances in Experimental Social Psychology, 39*, 163-254.

Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorisation of multifaceted stimuli. *Journal of Personality and Social Psychology, 77*(5), 927–941.

Corneille, O., Klein, O., Lambert, S., & Judd, C. M. (2002). On the role of familiarity with units of measurement in categorical accentuation: Tajfel and Wilkes (1963) revisited and replicated. *Psychological Science, 13*(4), 380–383.

Deffenbacher, D. M., Park, B., Judd, C. M., & Correll, J. (2009). Category boundaries can be accentuated without increasing intergroup bias. *Group Processes and Intergroup Relations, 12*(2), 175-193.

Dijksterhuis, A., & van Knippenberg, A. (1999). On the parameters of associative strength: Central tendency and variability as determinants of stereotype accessibility. *Personality and Social Psychology Bulletin, 25,* 527-536.

Donders, N. C., Correll, J., & Wittenbrink, B. (2008). Danger stereotypes predict racially biased attentional allocation, *Journal of Experimental Social Psychology, 44*, 1328-1333.

Doosje, B., van den Bos, K., Loseman, A., Feddes, A. R., & Mann, L. (2012). "My in-group is superior": Susceptibility for radical right –wing attitudes and behaviours in Dutch youth. *Negotiation and Conflict Management Research, 5* (3), 253-268.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3 (5),* 562-571.

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19,* 978-980.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology, 100,* 999-1014.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology, 43,* 116-125.

Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup contact: The past, present, and the future. *Group Processes and Intergroup Relations, 6*(1), 5–21.

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science, 25* (9), 1816-1821.

Ellemers, N., Spears, R., Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53,* 161-186.

Everett, R. S., & Wotjtkiewicz, R. A. (2002). Difference, disparity, and race/ethnic bias in federal sentencing. *Journal of Quantitative Criminology, 18* (2), 189-211.

Fiske, S. T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: evolution, culture, mind, and brain. *European Journal of Social Psychology, 30* (3), 299-322.

Fiske, S.T., & Neuberg, S.L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology*. New York: Academic Press.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology, 49* (1), 120-125.

Hamilton, D. L., & Mackie, D. M. (1990). Specificity and generality in the nature and use of stereotypes. *Advances in social cognition, 3*, 99–110.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–68). Taylor & Francis Group.

Hewstone, M., & Hamberger, J. (2000). Perceived variability and stereotype change. *Journal of Experimental Social Psychology, 36*, 103-124.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237-271.

Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology, 11*, 223-255.

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052–1072.

Imhoff, R., & Dotsch, R. (2013). Do we look like me or like us? Visual projection as self- or ingroup-projection. *Social Cognition*, *31*(6), 806-816.

Johnson, B. D. (2003). Racial and ethnic disparities in sentencing departures across modes of conviction. *Criminology, 41* (2), 449-490.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: (3) Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28*, 360–386.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*, 1–27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25* (6), 881-919.

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp.199-227). Hillsdale, NJ: Erlbaum.

King, E. B., & Ahmad, A. S. (2010). An experimental field study of interpersonal discrimination toward Muslim job applicants. *Personnel Psychology*, *63,* 881-906.

Klein S. B., Loftus J., Trafton R. G., Fuhrman R. W. (1992). The use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology, 63*, 739–753.

Kruglanksi, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry, 10,* 83-109.

Kruschke, J. K. (2001). Towards a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45*, 812-863.

Kruschke, J. K. (2011). Models of attention learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 120-152). New York: Cambridge University Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin 108* (3), 480-498.

Lech, R. K., Gunturkun, O., & Suchan, B. (2016). An interplay of fusiform gyrus and hippocampus enables prototype- and exemplar-based category learning. *Behavioural Brain Research, 311*, 239-246.

Linville, P. W., & Fischer, G. W. (1993). Exemplar and abstraction models of perceived group variability and stereotypicality. *Social Cognition, 11*, 92-125.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Macrae, N. C., Milne, A. B., Bodenhausen, G. V. (1994) Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*. *66*(1) 37-47.

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects: Applying social-cognitive theory to intergroup contact research. *Group Processes and Intergroup Relations, 14* (2), 255-274.

Mclaughlin, K. (2016, March 2). White officer shoots dead unarmed black grocer yards from his Alabama home. *Daily Mail Australia*. Retrieved from http://www.dailymail.co.uk/news/article-3471592/White-officer-kills-black-man-outside-home-historically-African-American-Alabama-neighborhood.html.

McCoy, T. (2014, November 26). Why I killed Michael Brown: Ferguson police officer Darren Wilson explains. *The Sydney Morning Herald: World.* Retrieved from http://www.smh.com.au/world/why-i-killed-michael-brown-ferguson-police-officer-darren-wilson-explains-20141125-11twqo.html.

McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). *Stereotypes as explanations: The formation of meaningful beliefs About Social Groups*. New York: Cambridge University Press.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representation: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, memory and Cognition*, *10*(3), 333-352.

Miller, N., & Dollard, J. (1941). *Social learning and imitation.* New Haven, NJ: Yale University Press.

Minda, J. P., & Smith, D. J. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*(3), 775-799.

Minda, J. P., & Smith, D. J. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(2), 275-292.

Minda, J. P., & Smith, D. J. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 40-64). New York: Cambridge University Press.

Mitchell, O. (2005). A meta-analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology, 21* (4), 439-466.

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from U.S. Federal Courts. *Journal of Law and Economics, 44*, 285-314.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cogntion. *Psychological Review, 108* (2), 291-310.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 211-233.

Operario, D., & Fiske, S. T. (2001). Effects of trait dominance on powerholders' judgments of subordinates. *Social Cognition, 19*, 161-180.

Park, B., & Hastie, R. (1987). Perception of variability in category development: Instance- versus abstraction-based stereotypes. *Journal of Personality and Social Psychology, 53*(4), 621-635.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59* (2), 173-191.

Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information. *Journal of Personality and Social Psychology, 80* (6), 876-893.

Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin, 37*, 1692–1703.

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implication for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology, 106,* 897-911.

Rehavi, M. M., & Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy, 122* (6), 1320-1354.

Riach, P. A. (2015). A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics, 29* (5), 608-619.

Riach, P. A., & Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the English labor market. *The B.E. Journal of Economic Analysis & Policy, 6* (2), 1-20.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192-233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & Lloyd, B. B. (Eds.), *Cognition and categorization* (pp. 27-48). Oxford, England: Lawrence Erlbaum; England.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology, 14*, 237-255.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effect of intergroup contact. *Journal of Social Issues, 41,* 81-104.

Rothbart, M., & Lewis, S. (1988). Inferring category attributes from exemplar attributes: Geometric shape and social categories. *Journal of Personality and*

Social Psychology, 55(6), 861-872.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology, 32*(4), 309-336.

Rubin, M., Hewstone, M., & Voci, A. (2001). Stretching the boundaries: Strategic perceptions of intragroup variability. *European Journal of Social Psychology, 31*, 413-429.

Rutland, A., & Brown, R. (2001). Stereotypes as justifications for prior intergroup discrimination: Studies of Scottish national stereotyping. *European Journal of Social Psychology, 31*, 127-141.

Schneider, D. J. (1991). Social cognition. *Annual Review of Psychology, 42*, 527-561.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70*(6), 1126-1141.

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and stereotyping: Cognitive constraints on the construction of meaningful social impression. *European Review of Social Psychology, 11* (1), 145-175.

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin, 26*, 1329-1342.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition, 8*(3), 243-262.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgement. *Psychological review, 99*(1), 3-21.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*(4), 984-1002.

Stangor, C., & Ford, T. E. (1992). Accuracy and expectancy confirming processing orientations and the development of stereotypes and prejudice. *European Review of Social Psychology, 3*(1), 57-89.

Stark, T. H., Flache, A., & Veenstra, R. (2013). Generalization of positive and negative attitudes towards individuals to outgroup attitudes. *Personality and Social Psychology Bulletin*, *39*, 608-622.

Swaine, J., Laughland, O., Lartey, J., & McCarthy, C. (2016, January 1). Young black men killed by US police at highest rate in year of 1134 deaths. *The Guardian*. Retrieved from http://www.theguardian.com/us-news/2015/dec/31/the-counted-police-killings-2015-young-black-men.

Tajfel, H. (1981). *Human groups and social categories*. London: Academic Press.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-177.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*, 101-114.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E. J. Lawler, *Advances in Group Processes* (pp.77-121). Greenwich, CT.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Voci, A. (2000) Perceived group variability and the salience of personal and social identity. *European Review of Social Psychology, 11* (1), 177-221.

Wason, P., & Johnson- Laird, P. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Welch, K. (2007). Black criminal stereotypes and racial profiling. *Journal of Contemporary Criminal Justice, 23* (3), 276-288.

Wilder, D. A., Simon, A. F. & Faith, M. (1996). Enhancing the impact of counterstereotypic information: Dispositional attributions for deviance. *Journal of Personality and Social Psychology, 71*, 276-287.

Wirtz, C., van der Plight, J., & Doosje, B. (2016). Negative attitudes toward Muslims in the Netherlands: The role of symbolic threat, stereotypes, and moral emotions. *Peace and Conflict: Journal of Peace Psychology, 22* (1), 75-83.

Wyer, N. A, Sadler, M. S., & Judd, C. M. (2002). Contrast effects in stereotype formation and change: The role of comparative context. *Journal of Experimental Social Psychology*, *38*(5), 443–458.

Wyer (Jr), R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review, 93* (3), 322-359.

Yzerbyt, V., Rocher, S., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 20–50). Cambridge, UK: Blackwell.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1160-1173.

Zielinski, C. (2015, April 8). Officer charged with murder after shooting black man eight times as he ran away. *News Corp Australia Network.* Retrieved from http://www.news.com.au/world/north-america/officer-charged-with-murder-after-

shooting-black-man-eight-times-as-he-ran-away/news-
story/767396d2c7c616afdc6a617a542eddf7.

Chapter 2: Evaluative-Fit and Emotion-Fit: An Associative Learning Outlook onto Stereotype Formation and Change

Associative learning refers to the ability to learn relationships between stimuli and events in the environment (Le Pelley, 2004). In this chapter, I will argue that a learning framework can contribute to explain stereotype and category development through learnt associations that arise during exposure to stereotype-relevant stimuli (e.g., intergroup contact). I will explain how associative learning processes might change stereotypes and category perception by altering the perceived goodness of fit between an exemplar and its category.

Several scholars have argued that associative learning provides a unique perspective to investigate social phenomena such as intergroup discrimination, stigmatisation and ingroup favouritism (Paolini, Harris & Griffin, 2016; Walther, Nagengast & Trasselli, 2005; Walther, Weil & Dusing, 2011). In particular, Paolini et al. (2016) suggested a reconceptualization of intergroup contact effects (i.e., changes in group-level judgments following first-hand experiences and second-hand exposure to information about outgroup members) in terms of associative learning. From a learning perspective, intergroup contact is the process through which information about the outgroup and its exemplars is learnt and progressively revised with new information.

In my research, I start from the premise that associative learning processes contribute to negative intergroup relations in society. Past research has shown that associations between valence (positivity or negativity) and outgroup exemplars can establish and change intergroup evaluations (De Houwer, Baeyens, Vansteenwegen, & Eelen, 2000; Mallan & Lipp, 2007; Olson & Fazio, 2006). Similarly, associations between specific emotions (e.g., fear, anxiety etc.) and outgroup exemplars established and changed the anxiety expressed towards outgroup exemplars (Mallan, Sax & Lipp, 2009; Olson & Fazio, 2006; Olsson, Ebert, Banaji & Phelps, 2005; Weisbuch, Pauker, & Ambady, 2009). I propose that associative learning can extend its reach from causing changes in exemplars' (de-)valuations and exemplar-related anxiety to affecting, for better or worse, the development and maintenance of group stereotypes and category representations. Specifically, I expect conditioning, a type of associative learning in which associations between stimuli are formed (Pavlov, 1927), to affect social categorisation and stereotypes through changes in perceived exemplar prototypicality. I propose that an exemplar's perceived prototypicality can be shifted to fit more (or less) with a group's evaluative or emotional modal response. The two mechanisms by which I propose exemplar prototypicality shifts to fit with the group perception are evaluative-fit and emotion-fit; these are discussed next.

Evaluative-Fit

Background and Predictions

In Chapter 1, I described self-categorisation and social identity theory, which theorise about the motivational processes that underlie group perceptions (Tajfel & Turner, 1979, Turner, Hogg, Oakes, Reicher & Wetherell, 1987). Briefly, according to these theories positive valence is typically associated with an ingroup and negative (or less positive) valence with the outgroup (Abrams & Hogg, 1990; Hornsey, 2008). Self-esteem motivations drive the association between positive valence and ingroup, because this association bolsters an individual's self-esteem through membership or belonging to a positive group. In contrast and as a result, negative valence or less positive valence becomes associated with the outgroup (Cakal, Hewstone, Guler & Heath, 2016; Ellemers, Spears & Doosje, 2002).

The fact that groups (e.g., ingroup and outgroups) are associated with different valences provides the basis for an evaluative-fit effect with my predictions, which are visually depicted in Figure 2. There is *evaluative-fit* when the behaviour, actions, and qualities of an individual match or align with their group's perceived valence (Coats, Latu & Haydel, 2007; Harwood et al., 2017). I have established that an ingroup's perceived valence is typically positive and the outgroup's perceived valence is typically negative. Therefore, when an ingroup exemplar becomes associated with a positively valenced stimulus they will be perceived as having high evaluative fit (branch a in Figure 2). Likewise, when an outgroup exemplar becomes associated with a negatively valenced stimulus they will be perceived as having high evaluative-fit (branch d in Figure 2). Based on social identity and self-categorisation theory (Tajfel & Turner, 1979, Turner et al., 1987), I predict high evaluative-fit to increase an exemplar's perceived prototypicality of their group because the exemplar will be perceived as being more consistent with the group perception. In contrast, there is low evaluative-fit or *misfit* when the behaviour, actions, and qualities of an individual do not align with their perceived group's valence (Coats et al., 2007; Harwood et al., 2017). Therefore, when an ingroup exemplar becomes associated with a negatively valenced stimulus they will be perceived as having low evaluative-fit (branch b in Figure 2). Similarly, when an outgroup exemplar becomes associated with a positively valenced stimulus they will be perceived as having low evaluative-fit (branch c in Figure 2). I predict low evaluative-fit to decrease an exemplar's perceived prototypicality of their group because the exemplar will be perceived as not aligning with pre-existing group evaluations. However, if an individual displays negatively valenced behaviours or qualities they are perceived as having low evaluative-fit or *misfit* with the ingroup. In contrast, I have established that outgroups' perceived valence is typically negative. Therefore, if an individual displays negatively valenced behaviours or attributes they are perceived as having high evaluative-fit with these outgroups. However, if an individual displays positive valenced behaviours or attributes they are perceived as having low evaluative-fit or misfit with these outgroups.

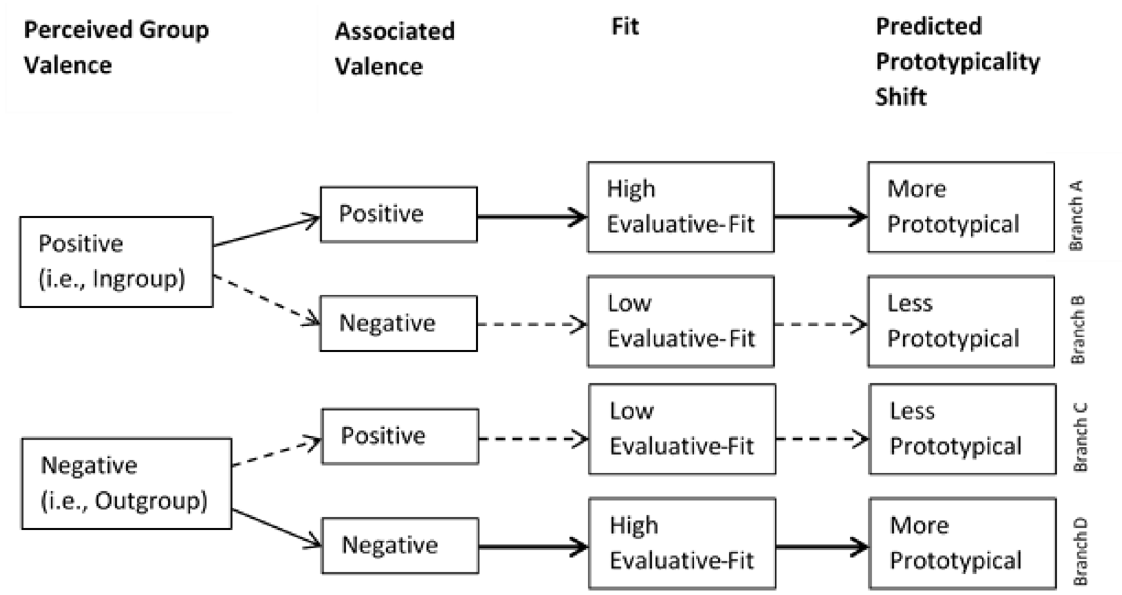


Figure 2. Evaluative-fit predictions for perceived exemplar prototypicality shifts

Evidence for Evaluative-Fit

Research investigating the “black sheep effect” describes the role evaluations have in determining ingroup and outgroup perceptions through evaluative-fit (see Marques & Paez, 1994; Marques, Yzerbyt & Leyens, 1988; Pinto, Marques, Levine & Abrams, 2010). Results from this line of research consistently show that ingroup exemplars associated with negative evaluations are judged and rated harsher than comparatively unlikeable outgroup exemplars (Coull, Yzerbyt, Castano, Paladino & Leemans, 2001). In contrast, ingroup exemplars associated with positive evaluations are treated better than comparatively likeable outgroup exemplars (Marques & Paez, 1994; Marques et al., 1988; Pinto et al., 2010). The act of treating and evaluating ingroup members differently as a function of their valence can be considered as a form of inclusion or exclusion from the ingroup: Negatively valenced ingroup exemplars are distanced or rejected from the ingroup category in order to protect the positivity of the ingroup; positively valenced ingroup exemplars are included in the ingroup category because their positivity enhances the ingroup’s perception. In other words, evaluations are treated psychologically as a way to mark the inclusion/exclusion of an exemplar from the category.

Scholars have investigated the effects exemplar valence has on perceived exemplar fit with their group. Richeson and Trawalter (2005), for example, asked White participants to sort images of positive and negative famous White and Black exemplars into their respective categories (White or Black). For example, their stimulus set included an image of John F Kennedy, as a likeable White exemplar, and an image of OJ Simpson, as a dislikeable Black exemplar. Richeson and Trawalter found that White individuals were faster and more accurate at sorting faces into the “White” category if they were admired White persons, as opposed to disliked White persons. An asymmetry in results was found for the Black category, whereby disliked Black persons were sorted faster and more accurately than admired Black persons (see also Ruys, Dijksterhuis & Corneille, 2008). Similarly, Coats et al. (2007) found exemplar categorisation was based on race when Blacks performed a negatively valenced behaviour and Whites a positively valenced behaviour. However, categorisation of exemplars based on race was disrupted when Blacks performed a positively valenced behaviour and Whites a negatively valenced behaviour. Thus, in line with evaluative-fit mechanisms, in the eyes of White participants, White exemplars who were perceived with positive valence were perceived as being more prototypical of the White category, than negatively valenced White exemplars. Disliked Black exemplars who were perceived with negative valence were perceived as being more prototypical of the outgroup compared to positively valenced Black exemplars (see also Hugenberg & Bodenhausen, 2004; Johnston & Hewstone, 1992; Plaks, Stroessner, Dweck & Sherman 2001; Rothbart, Sriram & David-Stitt, 1996).

Intergroup contact research provides additional evidence that evaluative-fit shapes categorisation. In one study, Paolini, Harwood and Rubin (2010) asked White participants to interact with an outgroup confederate, who behaved in a positive or negative manner towards participants. Negative interactions with the confederate resulted in participants perceiving the confederate as having a better fit with the outgroup compared to participants who experienced a positive interaction with the ethnic individual. Paolini et al. (2014) extended Paolini et al. (2010) and investigated the effects that the valence of direct, television-mediated and imagined contact had on category salience—a measure of group perceptions that encompasses exemplar prototypicality, category awareness and perceived intergroup differences. Paolini et al. (2014) manipulated the type of contact valence participants experienced and found negatively valenced or limited contact resulted in higher category salience judgments towards the outgroup exemplar. Positive or extensive prior contact with the outgroup moderated the effect contact valence had on category salience. The moderation effect suggested prior contact with the outgroup that was positive in nature and protected against the harmful effects of negative contact. Together, this evidence not only highlights positive valence is a marker of the ingroup and negative valence of the outgroup, but evaluative-fit shapes inclusion/exclusion of valenced contact experiences of outgroup members through variations in categorisation

The body of research presented in this section demonstrates that negative valence is typically a marker of outgroup membership and positive valence a typical marker of membership to the ingroup. Furthermore, exemplar categorisation is influenced by evaluative-fit. Positive valence provides a better fit with the ingroup category and negative valence a better fit with the outgroup category. Therefore, positive associations are more likely to result in exemplar categorisation with the ingroup (i.e., greater ingroup prototypicality and lower outgroup prototypicality) and negative associations are more likely to result in exemplar categorisation into the outgroup (i.e., greater outgroup prototypicality and lower ingroup prototypicality).

Emotion-Fit

Background and Predictions

I now transition from discussing evaluative-fit to discussing the effect emotion-fit has on group perception. Throughout our evolutionary history humans have evolved to fear other groups because of threats they posed (Barchas, 1986; Brewer & Caporael, 1990; McDonald, Navarrete & Van Vugt, 2012; Van Vugt & Park, 2010). Threats could be in the form of threat to physical safety, desired outcomes, disease or contamination or in the form of lost opportunities. Threats have evolved to correspond with specific emotions, which in turn are associated with different groups. DeSteno, Dasgupta, Bartlett and Cajdric (2004) provide an example of a threat association with the outgroup. They placed participants into minimal groups and induced an anger, sadness or neutral state experience. Results from DeSteno et al. (2004) showed anger, a threatening emotion involved in many intergroup conflicts, created automatic prejudice to an unknown laboratory-created outgroup. However, sadness and a neutral state, two emotions that are not threatening, had no effect on intergroup bias. These results suggest that threatening emotions, rather than negative emotional valence *per se* was a primary factor in changing responses to an outgroup.

The idea that outgroups are associated with threatening emotions has been further advanced to account for specific threats different outgroups posed. Cottrell and Neuberg (2005) found different outgroups are associated with specific distinct threats, and each threat carried a particular emotion. For example, a threat to desired outcomes causes anger, whilst a contamination threat causes disgust. Cottrell and Neuberg argue that specific outgroups, over time, have become associated with specific threat-related emotions that fit the type of threat an outgroup posed. For example, African Americans elicited higher anxiety affective responses compared to gay men, who elicited higher disgust responses. Dasgupta, DeSteno, Williams and Hunsinger (2009) research aligns with the idea of specific threatening emotions being associated with one outgroup but not another. Dasgupta et al. (2009) showed specific emotions produced increases in intergroup bias towards outgroups when the emotion applied to the group. For example, anger emotions produced increases in bias towards an Arab outgroup, but not to a homosexual outgroup. In contrast disgust produced increases in bias towards a homosexual outgroup, but not an Arab outgroup. Similar effects have been observed in numerous other studies in which other outgroups have been associated with one emotion more than another (see Olatunji, 2008; Terrizzi, Shoock & Ventis, 2010 for further evidence that disgust is associated with homosexuals). Hence, from this alternative stance, discrete emotions provide a better fit with specific outgroups because they have evolutionarily or culturally posed a particular threat and thus ‘carry’ specific emotions.

The notion that some groups are more associated with a particular threatening emotion than other groups provides the basis for emotion-fit effects. Researchers have touched on the existence of an emotion-fit mechanism, without describing the mechanism in detail. For example, the emotion-fit mechanism has been demonstrated across different domains of intergroup research. Dasgupta et al. (2009) demonstrate emotion-fit mechanisms from an intergroup bias angle. Their emotion specific hypothesis predicts that incidentally experienced emotions (i.e., from a source unrelated to the group or exemplar) increase outgroup bias when the emotion is applicable to the outgroup's pre-existing stereotype. In addition, Tapias, Glaser, Keltner, Vasquez and Wickens (2007) demonstrate emotion-fit mechanisms from a stereotypical angle. They hypothesised stereotypical perceptions towards a specific group would elicit discrete emotional responses. Thus, emotion-fit mechanisms suggest outgroups elicit specific emotions that not only affect stereotypical perceptions but intergroup biases too. I base an emotion-fit mechanism on the idea that different groups activate different emotional responses and are associated with specific discrete emotions.

An emotion-fit mechanism suggests that outgroups have a better fit with a specific (threat-related) emotion (Cottrell & Neuberg, 2005; Dasgupta et al., 2009). African Americans for example are associated with anxiety emotions, homosexuals are associated with disgust emotions and Native Americans are associated with pity emotions (Cottrell & Neuberg, 2005; Fiske, Cuddy, Glick & Xu, 2002; Tybur, Lieberman, Kurzban & DeScioli, 2013). These emotions became associated with each group presumably due to the type of evolutionary or cultural threat that each group posed or because their related stereotype makes associations with particular emotions more probable (Fiske et al., 2002). Therefore, if an individual is perceived as behaving threatening in such a way that elicits an anxiety association, the individual would be perceived as having high emotion-fit with the Black outgroup and low emotion-fit with a homosexual outgroup. Similarly, behaviour that causes disgust to be associated with an individual would result in them being perceived as having high emotion-fit with a homosexual outgroup, but low emotion-fit with the Black outgroup. In contrast, an individual who displays non-threatening behaviour or elicits no threat-related emotion (e.g., a safety signal) will have low emotion-fit with the outgroup because non-threatening behaviours and lack of threat-related emotions (aka safety) are not typical of the outgroup; high emotion-fit with the ingroup would be perceived instead, because non-threatening behaviours non-threat related emotions are perceived as being key features of the ingroup.

My predictions about how exemplar prototypicality shifts following conditioning can be explained by emotion-fit and are visually depicted in Figure 3. When an ingroup exemplar is paired with a non-threatening emotion or lack of threat they will be perceived as having high emotion-fit with the ingroup (branch a in Figure 3). Likewise, when an outgroup exemplar is perceived to be paired with a threatening emotion that aligns with the evolutionary threat that outgroup posed (i.e., Blacks and anxiety), high emotion-fit with that specific outgroup will be perceived (branch d in Figure 3). I predict high emotion-fit to increase an exemplar’s perceived prototypicality of their group because the exemplar will be perceived as being more consistent with the group perception. In contrast, when an ingroup exemplar is paired with a threatening emotion they will be perceived as having low emotion-fit (branch b in Figure 3). Similarly, when an outgroup exemplar is paired with a non-threatening emotion low emotion-fit is perceived (branch c in Figure 3). I predict low emotion-fit to decrease an exemplar’s perceived group prototypicality because the exemplar will not be perceived as being consistent or align with the perception of the group.

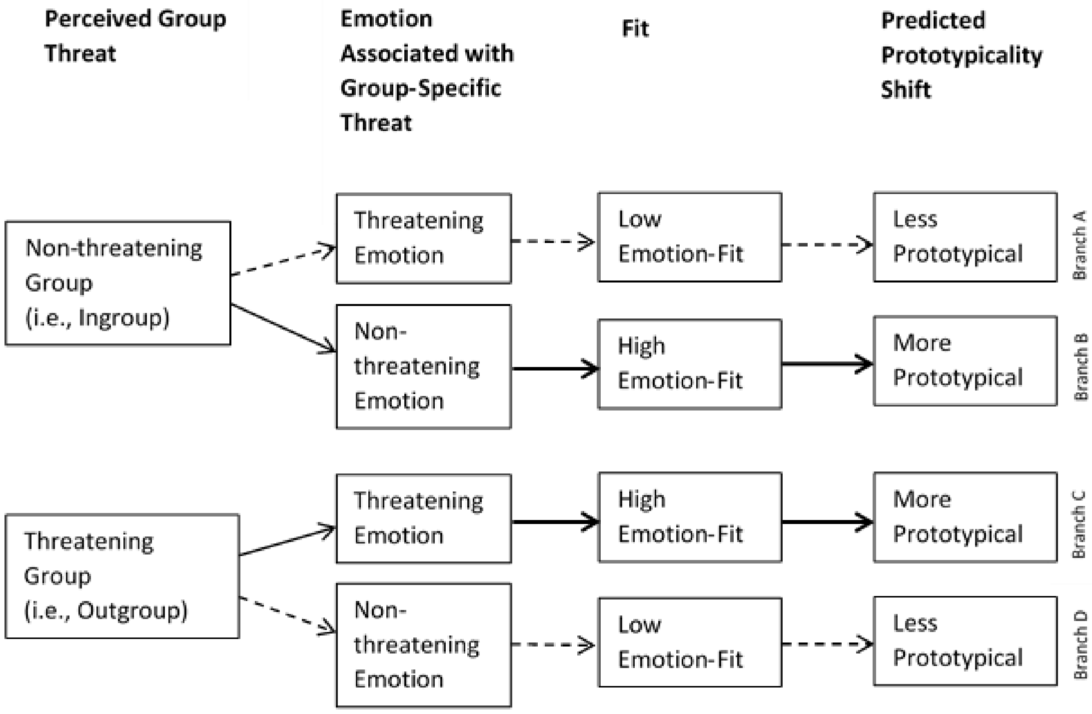


Figure 3. Emotion-fit predictions for perceived exemplar prototypicality shifts

Evidence for Emotion-Fit

Threat related factors provide a better fit with outgroup. For example, in a series of studies Miller, Maner and Becker (2010) demonstrate that threat related cues cause outgroup categorisation rather than ingroup categorisation. White participants were placed into groups that manipulated threat cues in the target (masculinity, movement towards participants and facial expressions) and threat related factors in the participant (i.e., interpersonal threats). Miller et al. (2010) found that participants were more likely to categorise targets that displayed threatening cues as a Black outgroup exemplar than a White ingroup exemplar. This effect was extended to minimal groups. Overall, Miller et al. (2010) evidence suggests threatening cues provide a better fit with the outgroup.

Greater anxiety towards outgroup (vs ingroup) exemplars has been demonstrated by conditioned anxiety being resistant to extinction (Mallan et al., 2009; Olsson et al., 2005; Olsson & Phelps, 2004; Parra, Esteves, Flykt & Ohman, 1997). Anxiety is generated in fear conditioning studies often via the administration of a mild electrical stimulation (other stimuli can also be used such as white noise). Studies show that conditioned anxiety persisted throughout extinction for the outgroup exemplar paired with the electrical stimulation but not the ingroup exemplar (Navarrete et al., 2012; Olsson et al., 2005; Olsson & Phelps, 2004). Therefore, this effect suggests that anxiety is harder to reduce for outgroup exemplars in comparison to ingroup exemplars, which I interpret as evidence for greater emotion-fit between anxiety and the outgroup (vs ingroup).

The idea that anxiety is more associated with outgroups is consistent with aspects of the evolved fear module (Mineka & Ohman, 2002). The core idea underlying this module is stimuli that posed a threat to human evolution are more readily associated with fear than stimuli that did not pose an evolutionary threat. This module allowed early ancestors to quickly adapt to dangerous and life threatening stimuli/situations. Seligman (1971) originally proposed the idea that selective associations towards stimuli or situations (prepared stimuli) that threatened the survival of humans formed. Mineka and Ohman (2002) extended this idea in the form of a fear module, which is assumed to have four key characteristics that stimuli must meet to be considered as evolutionary prepared stimuli (Mineka & Ohman, 2002; Ohman & Mineka, 2001); 1) Selectivity: Stimuli that have been associated with frequent threatening situations activate the fear module. Thus, only certain stimuli activate the module; 2) Automaticity: Fear relevant stimuli are expected to activate an automatic/rapid response when encountered; 3) Encapsulation: The activation of the fear module is encapsulated or resistant to more advanced human cognition; 4) Specific neural circuitry: A specific neural circuit shaped by evolutionary factors are expected to provide an adaptive advantage. Therefore, stimuli that meet all four conditions are considered to be ‘prepared stimuli’, as they posed an evolutionary threat that activates the fear module.

Evidence in support of an evolved fear module typically stems from fear conditioning research using evolutionary feared (i.e., snakes and spiders) and non-feared stimuli (i.e., flowers and mushrooms). For example, fear conditioning results consistent with the evolved fear module are: 1) Selectivity: Resistance to extinction towards snakes and spiders and not flowers and mushrooms demonstrate fear is selective towards evolutionary feared stimuli (Ohman, Fredrikson, Hugdahl & Rimmo, 1976; Schell, Dawson & Marinkovic, 1991); 2) Automaticity: Subliminal or non-conscious presentations following fear conditioning found fear persisted towards snakes and spiders paired with the aversive stimulus but not flowers and mushrooms (Ohman & Soares, 1993). Thus, fear responses were automatically activated in the absence of conscious awareness; 3) Encapsulation: Studies demonstrate a resistance to instructed extinction with fear relevant stimuli (Hugdahl & Ohman, 1977; Soares & Ohman, 1993). Therefore, instructed extinction did not change participants’ perceptions of fear relevant stimuli and demonstrates relative independence between cognition and the evolved fear module; 4) Specific neural circuitry: The amygdala has been proposed as the gateway of neural circuitry that is particularly sensitive to negative and threatening stimuli (LeDoux, 2000; Mineka & Ohman, 2002). Greater amygdala activity towards stimuli that triggered a threat in our evolutionary path (i.e., threatening animals) compared to more recent cultural threats (i.e., guns) has been found (Yang, Bellgowan & Martin, 2012). Hence, a specialised neural circuitry has developed towards evolutionary feared stimuli.

The assumption of evolved fear has been proposed to include fear relevant social stimuli such as outgroups by several prominent researchers, thereby extending the concept of prepared learning to the social domain (Empirical evidence; Navarrete et al., 2012, Ohman & Mineka, 2001; Olsson et al., 2005; Olsson & Phelps, 2004; Theoretical evidence: Paolini et al., 2016). Thus, the evolved fear module suggests evolutionary feared stimuli enter aversive associations more readily, and the module has recently been extended to include social stimuli (i.e., outgroups).

Social stimuli classification as prepared stimuli within the evolved fear module has been questioned (Mallan, Lipp & Cochrane, 2013). Mallan et al. (2013) reviewed evidence and found results from fear conditioning studies are less robust for social stimuli than evolutionary feared animal stimuli, with effects being malleable to cognition. For example, Mallan et al. (2009) conditioned Caucasian participants to fear a Chinese outgroup. Prior to extinction one experimental group were instructed that no more electrical stimulations would be administered, thereby introducing a cognitive influence in this group of participants. Resistance to extinction was found in the group of participants who were not given instructions, whilst anxiety extinguished in the group given instructions. By failing one of the key requirements of the evolved fear module, that being encapsulated from cognition, Mallan et al. (2009) research is consistent with Mallan et al. (2013) rejection of the suggestion that outgroup exemplars exhibit properties of prepared stimuli. Therefore, Mallan and colleagues dispute the applicability of the evolved fear module for intergroup responding. In order to better account for results, Mallan et al. (2013) suggest a co-evolutionary system whereby genes and culture both influence intergroup responding (see also, Mesoudi, 2016).

The socio-cultural threat based approach is line with the idea that genetic and culturally evolved threats together activate specific emotions that affect group perception (Cottrell & Neuberg, 2005). This approach suggests specific emotions are associated with different outgroups depending on the type of threat an outgroup posed (see Cottrell & Neuberg, 2005; Dasgupta et al., 2009; DeSteno et al., 2004). For example, Dasgupta et al. (2009) demonstrated specific emotions are associated with specific outgroups, depending on the type of threat an outgroup posed. Participants completed an emotion induction task, where they described a personal account that made them angry, disgusted or emotionally neutral. Afterwards participants completed an IAT, which measured implicit attitudes towards Gays and Lesbians (Experiment 2) and Arabs (Experiment 3). Results suggested participants induced to feel disgust had more implicit bias towards Gays and Lesbians, whilst participants induced to feel anger had more implicit bias towards Arabs. The association between specific emotions with certain outgroups is a growing body of research that has expanded to include a variety of emotions (i.e., anxiety, anger, disgust and pity) with a range of outgroups (i.e., ethnicity, sexual preference, sex) (see Cottrell & Neuberg, 2005; Cunningham, Forestell & Dickter, 2013; Dasgupta et al., 2009; Kuppens, Pollet, Teixeria, Demoulin, Roberts & Little, 2012; Ramos et al., 2015; Vartanian, Thomas & Vanman, 2013). Based on this body of research both evolutionary and cultural threats contribute to group perception that is consistent with an emotion-fit mechanism.

It is important to point out that whether associations between outgroups and specific emotions are made readier by genetic or social learning factors is not key for my argument. Rather, the key point argued from the research described is that a readiness to associate (varied) threat-responses (e.g., including disgust) and outgroups exists – either through genetic or culturally learnt threats. In other words, based on my prediction of an emotion-fit mechanism, I expect genetic and/or culturally learnt threats to facilitate associations between outgroups and certain (typically negative) emotions via emotion-fit but I am fundamentally agnostic about this readiness' exact origin. Therefore, the key point is threat related emotions are associated more easily with certain outgroups, based on the type of evolutionary/cultural threat posed in the past or present by these groups.

In summary, the body of research presented demonstrates threatening emotions are a typical marker of outgroup membership and non-threatening/safe emotions a typical marker of the ingroup. Furthermore, exemplar categorisation is influenced by emotion-fit. In the eyes of White Australians, anxiety constitutes a marker for the Black outgroup (Mineka & Ohman, 2002; Olsson et al., 2005; Seligman, 1971). Therefore, a Black exemplar that is associated with an anxiety emotion is more likely to be categorised into the Black outgroup (i.e., greater outgroup prototypicality) because of greater emotional-fit (for an 'integral' analogue, please see Weisbuch & Ambady, 2008).

Associative Learning and Conditioning

To investigate the effect evaluative-fit and emotion-fit mechanisms have on exemplar prototypicality, I use conditioning procedures, a form of associative learning, to create associations between social stimuli and valence/emotions. Associative learning is broadly defined as learning a predictive relationship between two stimuli (Mitchell, De Houwer & Lovibond, 2009; Shanks, 1995). Associative learning encompasses a variety of conditioning procedures, including classical, evaluative and fear conditioning. In classical conditioning a conditioned stimulus (CS) becomes associated with an unconditioned stimulus (US) to produce a naturally occurring behaviour (Baeyens, 1998; Le Pelley, 2004; Shanks, 1995). Fear conditioning is a subset of classical conditioning and involves the formation of fear relevant associations between the CS and US (Delgado, Olsson & Phelps, 2006; Ohman & Mineka, 2001; 2003). Evaluative conditioning is another subset of classical conditioning that involves the formation of positive or negative valenced associations between the CS and US (De Houwer, Thomas & Baeyens, 2001; Hofmann et al., 2010; Martin & Levey, 1985). Whilst there are different subsets of conditioning, they all can be used to create valenced/emotional associations with the stimuli I investigate. Thus, these approaches are valid and appropriate for my research questions.

Acquisition is a key component in conditioning procedures that refers to the process of learning a predictive association between two stimuli (Davey, 1992; Harris, 2011). At the very beginning of acquisition, an US produces an unconditioned response. Initially, at this point in time no associations have been formed and the CS does not elicit any response. Throughout the acquisition process, an association begins to form between the conditioned stimulus (previously the neutral stimulus) and US after being paired together. At the end of acquisition, an association between the CS and US has developed and the CS will evoke the CR (conditioned response). Thus acquisition is the process of learning an association between two stimuli and knowledge of the association is known as contingency awareness. In my research, acquisition is the process in which I develop evaluative or emotional associations between social stimuli (CS) and an evaluative/emotionally loaded US.

In contrast to acquisition is the extinction process. Extinction is the process of disrupting a learnt association developed during acquisition so that the CS no longer predicts the US (Bouton, 1994; 2002; 2004; 2014). The extinction process disrupts associations developed during acquisition by reducing the predictive power of the CS, or through learning a new association that competes with the previously learnt association (i.e., the CS no longer predicts the US; Bouton, 2002). Thus, extinction does not eliminate the association developed during acquisition, it renders the CS ambiguous and effects can re-emerge. Evidence that the association developed during acquisition is not permanently extinguished comes from research investigating renewal, spontaneous recovery and reinstatement of associations following extinction. Renewal of a conditioned response occurs when participants experience contextual changes after extinction (Bouton, 2002). The change in context after extinction results in the conditioned response being renewed when the CS is presented in the new context. There are several versions of the renewal effect that vary in the context acquisition, extinction and re-test are presented in. A common type of renewal is when acquisition develops in context A, extinction of the conditioned response occurs in context B, after which the CS is presented once again in context A and evoking the conditioned response (ABA renewal; Holmes & Westbrook, 2014; Neumann, Lipp & Cory, 2007). Other types of renewal effects include ‘ABC renewal’ (Bouton, 2002; Miguez, Cham & Miller, 2012) and ‘AAB’ renewal (Bouton & Ricker, 1994). Spontaneous recovery is the recovery of an extinguished association following a passage of time (Bouton, 2002; Rescorla, 2004). Pavlov (1927) first noted that a response would *spontaneously* recover when the CS was presented again after a period of time since extinction had elapsed. A possible explanation of this effect is that the temporal context changes (Bouton, 2002; 2004): The extinguished response recovers because it is presented outside of the temporal context it was extinguished in (Bouton, 1994; 2002; 2004). Reinstatement is another effect where a conditioned responses re-emerges and occurs when the US is presented following extinction in the absence of the CS. The re-presentation of the CS after the US presentation results in the conditioned response being reinstated (Bouton, 2002; Norrholm et al., 2006). Throughout my research prototypicality measures are collected at different time points to investigate the implication extinction processes have on prototypicality shifts.

In my research, I use conditioning to superimpose valence/emotions onto exemplars as a way to investigate the implication of associative (vs non-associative) processes in social categorisation. In conditioning procedures, valence/emotions can be superimposed onto one exemplar (CS+) whilst another exemplar is never superimposed with valence/emotion (CS-). Through this differential conditioning procedure I cause differential valence/emotions to be associated to different exemplars. This cue-based/contingent-specific approach provides the opportunity to investigate whether exemplar prototypicality shifts take place selectively in one exemplar but not another. That is, associative processes can be investigated by ascertaining whether prototypicality shifts occur in one exemplar (CS+) but not the other (CS-) as a result of superimposed valence/emotion, or whether non contingent specific prototypicality shifts occur that indiscriminately affect all exemplars (non-associative process).

Non-associative changes in prototypicality may be due to sensitisation and habituation effects. Sensitisation effects occur when participants respond to the CS+ and CS- similarly following repeated presentations of the US (Cevik, 2014). Although habituation is a different process to extinction, they share a similarity in that responses may diminish following repeated presentations of the CS+/CS- (Best et al., 2008; Thompson & Spencer, 1966).

Overall, my view is that understanding how group exemplar perceptions develop and change – through associative or non-associative processes – can ultimately shed light on when and how stereotypes are applied and might selectively (or not) apply to certain group exemplars more than others. In summary, in my research acquisition is the process in which participants learn to associate social stimuli with a particular valence/emotion. Extinction is the process of disrupting (but not necessarily removing) the association learnt during acquisition. Renewal, spontaneous recovery and reinstatement could occur and the conditioned response could be activated post-extinction. I use conditioning to create and extinguish associations, so that associative processes (vs non-associative processes) can be investigated, although not essential for testing my research questions. That is, I can determine the selective effect that pairing valence/emotion with one exemplar but not another has on the perception of exemplars, but if non-associative process occur these can also be investigated.

Potentials and Challenges of an Associative Outlook in this Research

The main goal of this research is to investigate prototypicality shifts of exemplars involved in aversive conditioning with a negatively valenced and anxiety provoking stimulus. Understanding how exemplar prototypicality shifts following an association with a stimulus is important because it has the potential to provide greater insight into how stereotypical perceptions and intergroup relations can be improved. Stereotypical perceptions and intergroup relations could be improved via exemplar prototypicality shifts because changes in prototypicality are a precursor for changes in intergroup stereotypes and relations (Rothbart & John, 1985; Rothbart et al., 1996; Sherman, 1996).

I expect exemplar prototypicality to shift in a direction consistent with an evaluative-fit and emotion-fit mechanism. The predicted prototypicality shifts derived from evaluative-fit and emotion-fit mechanisms are linked to two qualitatively different motivations that focus on the ingroup and outgroup respectively. Predictions derived from an evaluative-fit mechanism are largely focused on an individual’s motivation to maintain the ingroup’s positive distinctiveness (Hogg, 2000; Jost & Banaji, 1994; Jost, Banaji & Nosek, 2004). Thus, in order to maintain positive ingroup (and self as ingroup member) distinctiveness the individual would harness the evaluative-fit mechanism with a focus on *valence effects* implicating the *ingroup*. In contrast, predictions derived from an emotion-fit mechanism are largely focused on the motivation to preserve an individual's physical/psychological integrity from threats that could cause harm, whatever their origin (i.e., evolutionary or cultural; Cottrell & Neuberg, 2005). Thus, in order to protect an individual from harm, the emotion-fit mechanism would be harnessed with a focus on *emotion-specificity* towards the *outgroup*. Therefore, evaluative-fit predictions are driven by the motivation for ingroup's positive distinctiveness and emotion-fit predictions driven by the motivation to protect oneself from outgroup threat.

While conceptually distinct, it is important to recognise that it is empirically difficult to distinguish between evaluative-fit and emotion-fit predicted shifts of exemplar prototypicality. Predictions derived from an evaluative-fit mechanism encompass any association that is perceived as being positive or negative in valence. Predictions derived from an emotion-fit mechanism provide a more nuanced approach that focuses on a specific emotion. As emotions carry a positive or negative valence, it is difficult to distinguish whether exemplar prototypicality shifts are caused by the associated valence or by the emotion itself.

For example, the Black outgroup is typically associated with negative valence *and* anxiety (Amodio & Hamilton, 2012; Cottrell & Neuberg, 2005; Hogg, 2000). A Black outgroup exemplar that is associated with a mild electric stimulation – a negatively valenced stimulus that also causes anxiety – is predicted to shift towards the outgroup and be perceived more prototypical of the Black outgroup. The exemplar's shift in prototypicality could be explained by an association with negative valence (evaluative-fit) and/or by anxiety (emotion-fit) generated by the electrical stimulation.

It is difficult to distinguish between the two mechanisms, and at the first instance my research does not attempt to do so. Rather, most of my research tests for exemplar prototypicality shifts using methods that potentially trigger *both* evaluative-fit *and* emotion-fit mechanisms. Initially I do not attempt to distinguish between evaluative-fit and emotion-fit mechanisms as my primary interest was in establishing the existence of a prototypicality shift effect, whatever the underlying mechanisms. Once the prototypicality shift effect has been established with some degree of confidence, I will attempt to differentiate between the two explanations.

My basic method tests for perceived exemplar prototypicality shifts prior to, and after aversive conditioning. One exemplar is paired with a mild electrical stimulation, whilst another exemplar is never paired with the electrical stimulation. The electrical stimulation used throughout aversive conditioning can be considered as being both a negatively valenced stimulus and an anxiety provoking stimulus. Thus, my basic method combines negative valence and anxiety when testing for exemplar prototypicality shifts. In Study 6.1 I changed my basic method and used winning and losing money to create associations rather than a mild electric stimulation in order to better distinguish between evaluative- and emotion-fit. Money loss removed anxiety from any association providing a neater test the effect a negatively valenced association with an exemplar (evaluative-fit) has for exemplar prototypicality shifts. Furthermore, positive associations can also be tested with this alternate method. In summary, the majority of my research investigates exemplar prototypicality shifts following an association between an exemplar and a negatively valenced/anxiety provoking stimulus.

Fear conditioning provides a method of forming associations between two stimuli and should be required when shifting prototypicality. In addition to investigating exemplar prototypicality shifts as a result of conditioning, I sought to explore the mechanisms or components of conditioning that might facilitate prototypicality shifts. Contingency awareness is the knowledge of which exemplar was paired with an unconditioned stimulus (e.g., the mild electrical stimulation; Kattner, Ellermeier & Tavakoli, 2012; Lovibond & Shanks, 2002; Stahl, Unkelbach & Corneille, 2009). Contingency awareness and whether it is required for robust conditioning effects has been debated (see, De Houwer et al., 2001; Field, 2000). It is largely accepted that contingency awareness is required for fear conditioning (Lovibond, Liu, Weildemann & Mitchell, 2011; Lovibond & Shanks, 2002) and evaluative conditioning (Hofmann, De Houwer, Perugini, Baeyens & Crombez., 2010; Pleyers, Corneille, Luminet & Yzerbyt, 2007). Despite that debate being largely resolved, the effect contingency awareness has in modifying perceived exemplar prototypicality remains untested. One could argue that changes in exemplar prototypicality are based on fear/evaluative conditioning, and therefore will require awareness. However, as exemplar prototypicality is malleable, contingency awareness may affect this variable differently to previously observed effects. Throughout my research I explore whether an association between an exemplar and evaluations/emotions needs to be recognised in order to shift exemplar prototypicality. Larger exemplar prototypicality shifts are expected in the contingent aware (vs. contingent unaware) participants because the evaluative-/emotion-fit between the exemplar and unconditioned stimulus will be greater than in contingent unaware participants.

Furthermore, I sought to investigate how exemplar prototypicality shifts generalise to similar exemplars not involved in conditioning. Prior to, and following conditioning when exemplar prototypicality is measured towards faces presented during conditioning, I also measure the perceived prototypicality of a 25% and 50% variation of the faces involved in conditioning and two new exemplars. Similarity gradients have been investigated in both the social (Harwood, Paolini, Joyce, Rubin & Arroyo, 2010) and learning domains (Dunsmoor, Mitfoff & LaBar, 2009; Vervliet, Vansteenwegen & Eelen, 2006), but to the best of my knowledge they have never been investigated in the learning domain with social stimuli. This extension allows me to investigate whether evaluative-fit and emotion-fit mechanisms involved in shifting the prototypicality of exemplars implicated in acquisition extend to similar exemplars. Investigating whether prototypicality shifts as a result of evaluative- and emotion-fit extend to exemplars not involved in conditioning represents a unique approach that has consequences for stereotyping and intergroup relations more broadly. Similarity driven generalisations can contribute to broader societal changes in group representations after exposure to specific exemplars and this aspect of my research clarifies how far the reach is.

Lastly, I investigate whether changes in exemplar prototypicality need to be learnt first-hand, or if exemplar prototypicality shifts occur when one merely witnesses another experiencing an aversive association. If other forms of socially mediated experiences (i.e., vicarious contact, television etc.) are capable of shifting exemplar prototypicality, they would demonstrate that changes in intergroup categorisation consistent with evaluative- and emotion-fit mechanisms can take place without directly experiencing negative valence and anxiety (Bandura, 1977; Mallan et al., 2009; Mazziotta, Mummendey, & Wright, 2011; Miller & Dollard, 1941; Olsson & Phelps, 2004). For example, an individual would not need to experience negative valence and anxiety in conjunction with an exemplar to change perceived exemplar prototypicality. Rather, they would only need to “observe” an association between an exemplar and negative valenced/anxiety provoking stimulus in order to shift exemplar prototypicality. Changes in exemplar prototypicality in the absence of directly experiencing negative valence/anxiety is another method that intergroup categorisation, and as an extension stereotyping, can be shifted.

To summarise, my research investigates (1) whether perceived exemplar prototypicality shifts following an association with an evaluative/emotion loaded stimulus; (2) potential mechanisms responsible for changing perceived exemplar prototypicality; (3) whether prototypicality shifts extend to similar group exemplars; and (4) whether changes in exemplar prototypicality are equal in magnitude when associations are learnt directly or vicariously. Next, I outline how each study in this thesis aims to investigate these areas of interest.

Overview of Studies

My thesis reports seven studies that investigate intergroup categorisation in terms of changes in exemplar prototypicality as a function of associative links with evaluations and emotions. White, Black and minimal group exemplars were paired with a negatively valenced and anxiety provoking stimulus (mild electric stimulation; referred to as the unconditioned stimulus) via a conditioning paradigm. White participants rated perceived exemplar prototypicality towards ingroup (Chapter 5) and outgroup stimuli (Chapter 3-6) prior to, and after aversive conditioning with a negatively valenced and anxiety provoking stimulus (Chapter 3-5), and aversive and appetitive conditioning with a negatively or positively valenced stimulus (Chapter 6). During conditioning an ingroup or outgroup exemplar was paired with an unconditioned stimulus (CS+, unsafe face), whilst another ingroup or outgroup exemplar was never paired with the unconditioned stimulus (CS-, safe face). I expected exemplar prototypicality to shift after being associated with valence and an emotionally loaded stimulus in the direction consistent with evaluative- and emotion-fit. Thus, I propose a novel way in which perceived exemplar prototypicality can shift following associations.

In Chapter 3, I present three studies that investigate perceived outgroup exemplar prototypicality shifts. I first aim to establish the existence of outgroup exemplar prototypicality shifts following an association with a negatively valenced and anxiety provoking stimulus. Across all three studies perceived prototypicality of the outgroup exemplars involved in conditioning are measured prior to, and after conditioning with a self-reported measure (Study 3.1) and an implicit measure (Study 3.2 and 3.3). Study 3.2 and 3.3 extended Study 3.1 by manipulating the time points perceived prototypicality was measured. As in Study 3.1, one group of participants completed post-test prototypicality measures following acquisition and extinction, whilst another group of participants completed post-test prototypicality measures immediately after acquisition and before extinction. Measurements were taken at different time points to explore the effect extinction had on prototypicality shifts. Furthermore, Study 3.2 explored the effects contingency awareness had by analysing prototypicality shifts separately for contingent aware and contingent unaware participants. In contrast, Study 3.3 masked exemplar presentations during conditioning, preventing contingency awareness from occurring. Whether the effects generalised to similar exemplars was investigated by measuring perceived prototypicality of a 25% and 50% variation of exemplars presented during conditioning, and two new exemplars. Lastly, I investigated whether contingency needed to be experienced first-hand, or if viewing another participant learning the exemplar and negative valence/anxiety association was sufficient to shift perceived outgroup exemplar prototypicality.

In Chapter 4 I followed up results from Chapter 3 and continued to investigate outgroup exemplar prototypicality shifts of exemplars involved in conditioning and whether the effects generalised to similar exemplars. In particular, the effect extinction had on outgroup exemplar prototypicality shifts was explored in greater detail to provide insight into the underlying mechanisms involved in prototypicality shifts. The mechanism Study 4.1 focused on was contingency awareness and the role repeated exemplar presentations during extinction had on prototypicality shifts. The extinction procedure was manipulated by changing the type of extinction participants underwent. During extinction one group of participants were presented with repeated exemplar presentations, as in the research reported in Chapter 3. The other group of participants were presented with repeated presentations of distorted images of exemplars that could not be recognised. Manipulating the type of image allowed me to determine whether recognizable outgroup exemplar presentations during the extinction procedure were required to shift outgroup exemplar prototypicality.

In Chapter 5, I reported two studies that sought to determine whether the processes responsible for shifting outgroup exemplar prototypicality extended to the ingroup. Ingroup and outgroup membership was defined by minimal groups in Study 5.1 and based on ethnicity in Study 5.2. In both studies participants underwent aversive conditioning using a negatively valenced and anxiety provoking stimulus paired with ingroup and outgroup exemplars. Perceived exemplar prototypicality of exemplars involved in conditioning and generalisation exemplars was measured prior to, and after acquisition. The effect contingency awareness had was also investigated by analysing prototypicality shifts separately for contingent aware and unaware participants. Thus, in this research I investigated whether ingroup exemplar prototypicality shifted away from the ingroup prototype when paired with a negatively valenced and anxiety provoking stimulus as a result of evaluative- and emotion-misfit and explored underlying mechanisms involved in the shift.

In Chapter 6, I continue to investigate outgroup exemplar prototypicality shifts and whether shifts generalise to similar exemplars. However, in Study 6.1 outgroup exemplar prototypicality shifts are investigated with a significantly modified research method that manipulates valenced association in the absence of anxiety. A gambling cover story was used and a conditioning paradigm implemented in which exemplars were paired with money loss or money gain. Perceived outgroup exemplar prototypicality was measured prior to, and after conditioning as reported in earlier studies. Money loss was used as the unconditioned stimulus in one group of participants, which provided a neater test of evaluative-fit because the anxiety component associated with my previously used unconditioned stimulus was largely removed. Furthermore, money gain was used as the unconditioned stimulus in the other group of participants, providing a first investigation into the effect an association between an outgroup exemplar and positive valence has for exemplar prototypicality shifts. In addition to investigating outgroup exemplar prototypicality shifts following conditioning with a positive and negative valenced stimulus, I also explored the effects contingency awareness had by again analysing contingent aware and unaware results separately.

In Chapter 7, I provide a general discussion of the results reported throughout my thesis. Firstly, I provide an overview of results from each study and expand on previous discussions within each chapter and explain the broader links to other research theories and results, along with broader implications to society. Limitations of the research are discussed and ideas for future research are proposed.

Overall, my research has the potential to identify ways to reduce negative intergroup stereotypes and intergroup attitudes in society. Exemplar prototypicality is a key determinant of category activation (Bruner, 1957; Locke et al., 2005; Medin &

Smith, 1981) and a gatekeeper of generalised changes in intergroup attitudes and stereotypes (Brown & Hewstone, 2005; Rothbart & John, 1985). Thus, any factor capable of shifting exemplar prototypicality has the potential to affect intergroup relations more broadly. Understanding how exemplar prototypicality shifts is important because underlying mechanisms can be manipulated to prevent negative stereotypes from developing initially, and by providing avenues to change pre-existing negative stereotypes. For example, a possible practical implication stemming from my research would be the requirement for structured intergroup contact that is void of any potential negative and threatening stimuli in order to prevent negative stereotypes developing and to change pre-existing stereotypes. Together, my research provides insight into the psychological underpinnings of prototypicality shifts that have practical implications for understanding the development and the reduction of negative intergroup stereotypes and relations in society.

References

Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology, 1*(1), 195-228.

Amodio D. M., & Hamilton, H. K. (2012). Intergroup anxiety effects on implicit racial evaluation and stereotyping. *Emotion, 12*(6), 1273-1280.

Baeyens, F. (1998). Cool whitewine when suddenly the doorbell goes: Reference and evaluation versus expectancy and preparation in Pavlovian learning. *Gedragstherapie, 31*, 7-47.

Bandura, A. (1977). *Social learning theory.* New York: General Learning Press.

Barchas, P. (1986). A sociophysiological orientation to small groups. In E. Lawler (Ed.), *Advances in group processes* (Vol. 3, pp. 209-246). Greenwich: JAI Press.

Best, J. D., Berghmans, S., Hunt, J. J. F. G, Clarke, S. C., Fleming, A., Goldsmith, P., & Roach, A. G. (2008). Non-associative learning in larval zebrafish. *Neuropsychopharmacology, 33*, 1206-1215.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(3), 219-231.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory, 11*, 485-494.

Bouton, M. E. (2014). Why behaviour change is difficult to sustain. *Preventive Medicine, 68*, 29-36.

Bouton, M. E., & Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning and Behavior, 22*(3), 317-324.

Brewer, M. B., & Caporael, L. R. (1990). Selfish genes vs. selfish people: Sociobiology as origin myth. *Motivation and Emotion, 14*, 237-242.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol 37* (pp.255-342). San Diego: Elsevier.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152.

Cakal, H., Hewstone, M., Guler, M., & Heath, A. (2016). Predicting support for collective action in the conflict between Turks and Kurds: Perceived threats as a mediator of intergroup contact and social identity. *Group Processes and Intergroup Relations, 19*(6), 732-752.

Cevik, A. O. (2014). Habituation, sensitization, and Pavlovian conditioning. *Frontiers in Integrative Neuroscience, 8*, 13.

Coats, S., Latu, I. M., & Haydel, L. A. (2007). The facilitative effect of evaluative fit on social categorization. *Current Research in Social Psychology, 12*, 54 – 67.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Coull, A., Yzerbyt, V. Y., Castano, E., Paladino, M. P., & Leemans, V. (2001). Protecting the ingroup: Motivated allocation of cognitive resources in the presence of threatening ingroup members. *Group Processes and Intergroup Relations, 4* (4), 327-339.

Cunningham, E., Forestell, C. A., & Dickter, C. L. (2013). Induced disgust affects implicit and explicit responses toward gay men and lesbians. *European Journal of Social Psychology, 43*(5), 362-369.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice, *Emotion, 9*(4), 585-591.

Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behaviour Research and Therapy, 14*(1), 29-66.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavoir Processes, 26*(2), 237-242.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biological Psychology*, *73*(1), 39-48.

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes, *Psychological Science, 15*(5), 319-324.

Dunsmoor, J. E., Mitroff, S. R., & LaBar, K. S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning and Memory*, *16*, 460-469.

Ellemers, N., Spears, R., Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*, 161-186.

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*, 13-36.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902.

Harris, J. A. (2011). The acquisition of conditioned responding. *Journal of Experimental Psychology: Animal Behavior Processes, 37*, 151-164.

Harwood, J., Joyce, N., Chen, C-Y., Paolini, S., Xiang, J., & Rubin M. (2017). Effects of past and present intergroup communication on perceived fit of an outgroup member and desire for future intergroup contact. *Communication Research, 44* (4), 530-555.

Harwood, J., Paolini, S., Joyce, N., Rubin, M., & Arroyo, A. (2011). Secondary transfer effects from imagined contact: Group similarity affects the generalization gradient. *British Journal of Social Psychology, 50*, 180-189.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390–421.

Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology, 11*, 223-255.

Holmes, N. M., & Westbrook, F. R. (2014). ABA renewal is greater when extinction occurs in the same context as cue pre-exposure. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*(3), 369-379.

Hornsey, M. J. (2008). Social identity theory and Self-categorization theory: A historical review. *Social and Personality Psychology Compass, 2* (1), 2014-222.

Hugdahl, K., & Ohman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 3*(5), 608-618.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization. *Psychological Science, 15*(5), 342–345.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: (3) Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28*, 360–386.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*, 1–27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25* (6), 881-919.

Kattner, F., Ellermeier, W., & Tavakoli, P. (2012). Both trace and delay conditioning of evaluative responses depend on contingency awareness. *Learning and Motivation*, *43*(1-2), 35–47.

Kuppens, T., Pollet, T. V., Teixeira, C. P., Demoulin, S., Roberts, S. G., & Little, A. C. (2012). Emotions in context: Anger causes ethnic bias but not gender bias in men but not in women. *European Journal of Social Psychology, 42*(4), 432-441.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155-184.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of*

Experimental Psychology Section B: Comparative and Physiological Psychology, 57 (3), 193 – 243.

Locke, V., Macrae, C. N., & Eaton, J. L. (2005). Is person categorization modulated by exemplar typicality? *Social Cognition, 23*(5), 417–428.

Lovibond, P. F., Liu, J. C., Weidemann, G., Mitchell, C. J. (2011). Awareness is necessary for differential trace and delay eyeblink conditioning in humans. *Biological Psychology, 87,* 393-400.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behaviour and Processes, 28,* 3-26.

Mallan, K. M., & Lipp, O. V. (2007). Does emotion modulate the blink reflex in human conditioning? Startle potentiation during pleasant and unpleasant cues in the picture–picture paradigm. *Psychophysiology, 44*(5), 737-748.

Mallan, K. M., & Lipp, O. V., & Cochrane, B. (2013). Slithering snakes, angry men and out-group members: What and whom are we evolved to fear? *Cognition and Emotion, 27*(7), 1168-1180.

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology, 45*(6), 1303–1307.

Martin, I., & Levey, A. B. (1985). Conditioning, evaluations and cognitions: An axis of integration. *Behaviour Research and Therapy, 23,* 167-175.

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367* (1589), 670-679.

Marques, J. M., & Paez, D. (1994). The ‘black sheep effect’: Social categorization, rejection of ingroup deviates, and perception of group variability. In W. Strobe, & M. Hewstone (Eds), *European Review of Social Psychology* (Vol 5, pp. 37-68). Chichester, UK: Wiley.

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The “black sheep effect”: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*, 1–16.

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects: Applying social-cognitive theory to intergroup contact research. *Group Processes and Intergroup Relations, 14* (2), 255-274.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(4), 241–253.

Mesoudi, A. (2016). Cultural evolution: Integrating psychology, evolution and culture. *Current Opinion in Psychology*, *7*, 17-22.

Miguez, G., Cham, H. A., Miller, R. R. (2012). Spontaneous recovery and ABC renewal from retroactive cue interference. *Learning and Behavior, 40*(1), 42-53.

Miller, N., & Dollard, J. (1941). *Social learning and imitation.* New Haven, NJ: Yale University Press

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between “us” and “them”. *Journal of Personality and Social Psychology, 99*(1), 62-77.

Mineka, S., & Ohman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry, 52*, 927-937.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associtive learning. *Behavioral and Brain Sciences, 32*(2), 183-198.

Navarrete, C. D., McDonald, M. M., Asher, B.D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an out-group face in a minimal group context. *Evolution and Human Behavior, 33* (5), 590-593.

Neumann, D. L., Lipp, O. V., & Cory, S. E. (2007). Conducting extinction in multiple contexts does not necessarily attenuate the renewal of shock expectancy in a fear-conditioning procedure with humans. *Behaviour Research and Therapy, 45*(2), 385-394.

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning and Memory, 13*(6), 681-685.

Ohman, A., Fredrikson, M., Hugdahl, K., & Rimmo, P. A. (1976). The premise of equipotentiality in human classical conditioning: conditioned electrodermal responses to potentially phobic stimuli. *Journal of Experimental Psychology: General*, *105*(4), 313-337.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Ohman, A., & Mineka, S. (2003). The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *Current Directions in Psychological Science, 12*(1), 5-9.

Ohman, A., & Soares, J. J. (1993). On the automatic nature of phobic fear: Conditioned electrodermal responses to masked fear-relevant stimuli. *Journal of Abnormal Psychology, 102*(1), 121-132.

Ohman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following non-recognized fear-relevant stimuli. *Journal of Experimental Psychology: General, 127*, 69-82.

Olatunji, B. O. (2008). Disgust, scrupulosity and conservative attitudes about sex: Evidence for a mediational model of homophobia. *Journal of Research in Personality, 42*, 1364-1369.

Olson, M. A, & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science, 309*(5735), 785–787.

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828.

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, *36* (12), 1723-1738.

Paolini, S., Harris, N. C., & Griffin, A. S. (2016). Learning anxiety in interactions with the outgroup: Towards a learning model of anxiety and stress in intergroup contact. *Group Processes and Intergroup Relations*, *19* (3), 275-313.

Paolini, S., Harwood, J., Rubin, M., Husnu, S., Joyce, N., & Hewstone, M. (2014). Positive and extensive intergroup contact in the past buffers against the disproportionate impact of negative contact in the present. *European Journal of Social Psychology*, *44*, 548–562.

Parra, C., Esteves, F., Flykt, A., & Ohman, A. (1997). Pavlovian conditioning to social stimuli: Backward masking and the dissociation of implicit and explicit cognitive processes. *European Psychologist, 2*(2), 106-117.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.

Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of Personality and Social Psychology*, *99*(1), 107–119.

Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information. *Journal of Personality and Social Psychology, 80* (6), 876-893.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis) liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*(1), 130-144.

Ramos, M. R., Barreto, M., Ellemers, N., Moya, M., Ferreira, L., & Calanchini, J. (2015). Exposure to sexism can decrease implicit gender stereotype bias. *European Journal of Social Psychology, 46*(4), 455-466.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning and Memory, 11*, 501-509.

Richeson, J. A, & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology*, *89*(4), 517–530.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effect of intergroup contact. *Journal of Social Issues*, *41,* 81-104.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology, 32*(4), 309-336.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis)categorization of unattractive brides and attractive prostitutes. *Experimental Psychology, 55*(3), 182–188.

Schell, A. M., Dawson, M. E., & Marinkovic, K. (1991). Effects of potentially phobic conditioned stimuli on retention, reconditioning, and extinction of the conditioned skin conductance response. *Psychophysiology, 28*(2), 140-153.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2,* 307-320.

Shanks, D. R. (1995). *The psychology of associative learning*. Great Britain: Cambridge University Press.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70*(6), 1126-1141.

Soares, J. J., & Ohman, A. (1993). Preattentive processing, preparedness and phobias: Effects of instruction on conditioned electrodermal responses to masked and non-masked fear-relevant stimuli. *Behavioural Research therapy, 31*(1), 87-95.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404–420.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific America, 223*(5), 96-102.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of social conflict. In W. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp.33-47). California: Brooks/Cole.

Tapias, M. P., Glaser, J., Keltner, D., Vasquez, K., & Wickens, T. (2007). Emotion and prejudice: Specific emotions toward outgroups. *Group Processes & Intergroup Relations, 10*(1), 27-39.

Terrizzi, J. A., Shook, N. J., & Ventis, W. L. (2010). Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. *Personality and Individual Differences, 49*(6), 587-592.

Thompson, R. F., & Spencer, W.A. (1966). Habituation: a model of phenomenon for the study of neuronal substrates of behaviour. *Psychological Review, 73*, 16-43.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review, 120*(1), 65-84.

Van Vugt, M., & Park, J. H. (2010). The tribal instinct hypothesis: Evolution and the social psychology of intergroup relations. In S. Sturmer, & M. Snyder (Eds.), *The psychology of prosocial behaviour: Group processes, intergroup relations, and helping* (pp.13-32). Malaysia: Wiley-Blackwell.

Vartanian, L. R., Thomas, M. A., & Vanman, E. J. Disgust, contempt, and anger and the stereotypes of obese people. *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity, 18*(4), 377-382.

Vervliet, B., Vansteenwegen, D., & Eelen, P. (2006). Generalization gradients for acquisition and extinction in human contingency learning. *Experimental Psychology, 53*(2), 132-142.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19*, 175-196.

Walther, E., Weil, R., & Dusing, J. (2011). The role of evaluative conditioning in attitude formation. *Current Directions in Psychological Science, 20*(3), 192-196.

Weisbuch M., & Ambady, N. (2008). Affective Divergence: Automatic responses to others' emotions depends on group membership. *Interpersonal Relations and Group Processes, 95*(5), 1063-1079.

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science*, *326*(5960), 1711–1714.

Yang, J., Bellgowan, P. S. F., & Martin, A. (2012). Threat, domain-specificity and the human amygdala. *Neuropsychologia, 50*(11), 2566-2572.

Chapter 3: Direct and Vicarious Aversive Learning Cause Shifts in the Perceived Prototypicality of Outgroup Exemplars

Research on stereotyping and prejudice flourishes in psychology and related disciplines. Ingroup members are typically regarded positively and outgroup members negatively (Turner, Hogg, Oakes, Reicher & Wetherell, 1987). Investigations on the affective and evaluative consequences that aversive associations have on social groups are well documented (Mallan, Sax, & Lipp, 2009; Olson & Fazio, 2006; Olsson, Ebert, Banaji, & Phelps, 2005; Weisbuch, Pauker, & Ambady, 2009). However, the knowledge about how aversive associations shape social categorisation is still scant. The present research uses aversive conditioning to examine the dynamics of perceived exemplar prototypicality, a well-established holistic marker of the categorisation process (Corneille & Judd, 1999; Rosch, 1978). To this end, I paired one outgroup face with a mild electro-tactile stimulation (the *unsafe* face/CS+), and another (control) outgroup face with no electro-tactile stimulation (the *safe* face/CS-). Extinction followed acquisition, which involved presenting repeated unreinforced presentations of the unsafe and safe face (CS+ and CS-) until no differential affective responding towards the unsafe and safe face (CS+ and CS-) is observed, as measured through skin conductance responses. Perceived outgroup exemplar prototypicality was investigated by comparing prototypicality after extinction, compared to before acquisition.

Past Research and This Research's Paradigm

The key aim of this research is to investigate and establish the existence of a perceived prototypicality shift of outgroup exemplars towards the outgroup prototype as a function of aversive conditioning. Cognitive categories are not rigidly defined mental sets (Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000). Whether exemplars are psychologically included in, or excluded from a category depends on their goodness of fit with the prototype of the category, or simply, their perceived prototypicality (Corneille & Judd, 1999; Rosch, 1978). For example, it is because of appreciable variations in perceived prototypicality that four legged chairs are more readily categorised as 'furniture' than 'stools', and police officers are more readily categorised as 'security guards' than 'bouncers'. Prototypicality, however, is not a static property. It is vulnerable to a host of contextual factors; for example inclusion-exclusion linguistic frames (Bless & Schwarz, 1998), context categories (Corneille & Judd, 1999), and judgment standards (Wyer, Sadler, & Judd, 2002). Many argue that it is this very malleability that makes exemplar prototypicality so consequential for cognitions in general (Macrae & Bodenhausen, 2000; Rosch, 1978).

Exemplar prototypicality is an important determinant of category activation (Bruner, 1957; Locke, Macrae & Eaton, 2005; Medin & Smith, 1981; Oakes, Haslam, & Turner, 1994). Exemplar prototypicality is pivotal to both inductive and deductive social reasoning, which was described in greater detail in the previous chapters (Rothbart & John, 1985; Rothbart, Sriram, & Davis-Stitt, 1996). This means that exemplar prototypicality will determine which category distinction among several potentially available (e.g., White vs. Black; young vs. old; male vs. female) becomes cognitively active or not during information processing. It will determine whether category information is used or not during processing of category exemplars and whether exemplar information generalises to the category as a whole.

Negatively valenced or anxiety provoking outgroup exemplars should have a higher goodness-of-fit with the outgroup category's prototype (Huttenlocher, Hedges, & Duncan, 1991; Turner et al., 1987) and, as a consequence, they should be included more readily and with less uncertainty in the outgroup representation; whereas positively valenced or non-anxiety provoking outgroup exemplars should be psychologically low in outgroup prototypicality and thus deemed for exclusion from stereotyped outgroup categories (Richeson & Trawalter, 2005). Therefore, I expect aversive conditioning to cause a perceptual shift of outgroup exemplars paired with an aversive stimulus towards the outgroup category's prototype, so that paired exemplars are perceived to be more prototypical after, compared to before aversive conditioning. The expected direction of these shifts aligns with an evaluative- and/or emotion-fit mechanism: *because* for White people the Black outgroup is typically associated with negativity, anxiety, and fear (Coats, Latu & Haydel, 2007; Cottrell & Neuberg, 2005; Dasgupta, DeSteno, Williams & Hunsinger, 2009), novel associations of outgroup exemplars with negativity, anxiety or fear should facilitate group-level cognitions and increase perceived prototypicality.

Past social categorisation research has manipulated negativity and anxiety that is integral to the outgroup exemplars (i.e., negativity that people experience as being intrinsically associated with the outgroup exemplar; Bodenhausen, 1993) and found evidence that it functions as a powerful marker of exemplars' membership to stereotyped social categories. For example, Richeson and Trawalter (2005) showed that White individuals were faster and more accurate at categorising as 'Black' faces of disliked than admired famous Black people (e.g., O. J. Simpson vs. Will Smith; see also Ruys, Dijksterhuis, & Corneille, 2008). Miller, Maner, and Becker (2010) detected a similar outgroup negativity bias when comparing categorisations for angry (vs. happy) male Black faces, as well as threatening (vs. neutral) male voices of exemplars of a laboratory created outgroup (see also Dunham, 2011; Hugenberg, 2005; Hugenberg & Bodenhausen, 2003; 2004). Hence, disliked, angry and threatening Black exemplars were treated by White individuals as being more prototypical of the outgroup, or more 'outgroup-like', than admired, happy, or safe Black exemplars.

I plan to show that incidentally-induced (vs. integrally held) negativity is sufficient to make outgroup exemplars become more outgroup-like. Changes in their included-excluded status would prove that outgroup exemplars do not need to *be* disliked, angry, or threatening to be perceived as more outgroup-like. Rather, it is sufficient for them to merely *happen* to coincide temporally and/or spatially with a disliked object or a threatening emotion to become more closely associated with the outgroup. Perceptual prototypicality shifts of the unsafe exemplar under incidental negativity/anxiety would signal that the (merely superimposed) negativity/anxiety has become psychologically integrated into the exemplar representation and used as a predictor of group membership. Hence, through incidentally inducing negativity/anxiety with a conditioning procedure I can investigate what information becomes integrated to (vs. excluded from) the cognitive representation of the outgroup exemplar and outgroup.

Basic, Generalised, and Socially Mediated Prototypicality Shifts

In my empirical work I will investigate three distinguishable prototypicality shifts. I refer to *basic* prototypicality shifts, when describing prototypicality shifts of *targets faces* presented during conditioning. Basic prototypicality shifts were investigated by measuring the perceived prototypicality of unambiguously Black computer-generated faces of neutral expression prior to, and after, incidentally—as opposed to integrally—pairing them with an aversive stimulus (i.e., a mildly uncomfortable electro-tactile stimulation). I hypothesise the face paired with the aversive stimulus—the unsafe face—will be perceived to be more outgroup-like or more prototypical of the Black category compared to the safe face after conditioning as compared to before conditioning because negativity/anxiety provides a better fit with the outgroup.

Discrete learning experiences with individual outgroup exemplars can adversely affect broad intergroup relations and contribute to the formation of negative stereotypes in society but only to extent that they transcend the specific outgroup exemplars immediately involved in the learning experience (Blair, Judd & Chapleu, 2004; Blair, Judd, Sadler & Jenkins, 2002; Brown & Hewstone, 2005). From the perspective of the present paradigm, negativity/anxiety should be readily applied to outgroup exemplars unrelated to the negative learning experience and extend to exemplars evoking no immediate threat *when* they resemble the unsafe face involved in the learning experience (i.e., generalisation to 'safe' exemplars; Pettigrew & Tropp, 2006; Ranganath & Nosek, 2008; Verosky & Todorov, 2013; Walther, 2002). Hence, I expect aversive conditioning to shape the inclusion/exclusion status in the outgroup category of exemplars that are both related (vs. unrelated) to the exemplars immediately involved in the aversive experience.

To test for *generalised* prototypicality shifts, I measure (pre/post-test) prototypicality not only of the target faces directly involved in conditioning, but also of *target face variations* and *new exemplars*. A 25% and 50% variation face for the unsafe and safe face was created that moved along a gradient progressively away from the Black prototype. Two new prototypically Black exemplars unrelated to the target faces were also created. I hypothesised that unsafe (vs. safe) face *variations* would shift in a similar direction to the unsafe exemplar and would be perceived as being more prototypical of the Black category after, compared to before conditioning. I hypothesised that *new exemplars* who were perceived as being more similar to the unsafe (vs safe) face would also shift in a similar direction to the unsafe exemplar and would be perceived as being more prototypical of the Black category after, compared to before conditioning. The magnitude of the generalised prototypicality shift was expected to follow a gradient and be more pronounced in faces perceived as being more similar to the target faces (i.e., a larger shift in the 25% generalisation face compared to the 50% generalisation face). If evidence of generalisation is not found it would suggest prototypicality shifts are isolated to specific exemplars and impact on group stereotypes is not as broad as expected.

I also reasoned that outgroup prototypicality shifts should have greater explanatory power for intergroup cognitions if they are readily transmitted between individuals in society. That is, if they are not circumscribed to first-hand experiences (Olsson & Phelps, 2004; Weisbuch et al., 2009), but extend to when one merely witnesses others displaying discomfort, or experienced threat paired with outgroup faces (Bandura, 1977; Miller & Dollard, 1941). To pursue this additional focus, half of my participants experienced the electro-tactile stimulation-plus-outgroup face pairings first hand (direct learning), the other half of the participants watched another ingroup member (a white female individual) receive the electro-tactile stimulation-plus-outgroup face pairings instead (vicarious learning; Olsson & Phelps, 2004). This approach allowed me to ascertain whether prototypicality shifts can be socially transmitted or whether, in order to materialise, they have to be acquired first-hand. I expected prototypicality shifts to extend to *when one merely witnesses* others displaying comfort or experiencing threat. Therefore, I expected prototypicality shifts to occur in both direct and vicarious aversive learning conditions but did not have any expectations for a difference in the size of the shifts because previous research found direct and vicarious learning to be comparable in magnitude (see also Olsson & Phelps, 2004).

Exploring the Interplay of Affect and Cognition

I expect evaluative-fit and emotion-fit to be a driving mechanism of exemplar prototypicality shifts. These mechanisms do *not* require, at a logical level, intervening changes in affect or emotions to take place for prototypicality shifts to materialise. The fit between an exemplar's group and negativity and/or an applicable emotion should suffice for the effect to present. However, given the novelty of this area of research, a secondary aim of this chapter is to explore the extent to which changes in prototypicality or cognitions *require* changes in exemplar-level affect/emotions or reflect cognition-heavy (vs. affect-heavy) processes. Hence, to check the interplay of cognition and affect, I will investigate the role of extinction, measurement controllability, contingency awareness and various mediational/moderational factors and test their role in prototypicality effects. In the literature there is debate over the degree to which cognitions and affect are related or independent from one another (Park & Judd, 2005). I therefore begin this discussion by providing a short summary of the two sides to this debate as a broader context to my analysis of the interplay of cognition and affect and the relative dominance of cognition in my research.

The Broader Debate in the Background

A body of historical and recent work argues that cognition and affect are relatively independent systems (Amodio & Frith, 2006; Zajonc, 1980) and another body of research that they are relatively dependent systems (Storbeck & Clore, 2007). In line with a notion of independent systems is neuropsychological (Amodio & Frith, 2006; Bush, Luu & Posner, 2000; Steele & Lawrie, 2014) and behavioural research (Amodio & Devine, 2006). Neuropsychological research found distinct neural circuits and brain regions are associated with cognitive vs. affective processing of members of social outgroups. For example, the posterior region of the rostral anterior cingulate cortex is more associated with cognitive processing and the anterior region of the rostral cingulate cortex is more associated with affective processing (Amodio & Frith, 2006; Bush et al., 2000; Steele & Lawrie, 2014). In contrast, other neuropsychological (Cunningham & Van Bavel, 2005; Cunningham, Zelazo, Packer & Van Bavel, 2007; Lieberman, Hariri, Jarcho, Eisenberger & Bookheimer, 2005) and behavioural research (Tajfel, 1970; Tajfel, Billig, Bundy & Flament, 1971) provide evidence that cognitive and affective processing are relatively intertwined. For example, minimal group paradigm studies demonstrate that negative affect (or reduced liking) of outgroup (vs. ingroup) members can be activated by simply categorising exemplars into different trivial lab based groups. Therefore, one branch of research provides evidence for relatively independent systems of cognitive and affective processing, whilst another branch of research provides evidence for relatively dependent systems.

Conditioning research has also provided ground for this debate. The notion of an evolved fear module (Mineka & Ohman, 2002; Seligman, 1971), and its application to the social domain (Navarrete et al., 2012, Ohman & Mineka, 2001; Olsson et al., 2005; Olsson & Phelps, 2004; Paolini, Harris & Griffin, 2016) traditionally aligns with the notion of relatively independent systems. From an evolutionary perspective, the evolved fear module posits that stimuli that threatened survival enter aversive associations more readily than stimuli that did not challenge survival (Ohman & Soares, 1993; Ohman & Mineka, 2003). As indicated in earlier chapters, key proponents of the module include an independent behavioural, mental and neural system that is automatically activated by affect/emotions; the module would rely on a specialised neural circuit that is encapsulated from cognition. Hence, many proponents of the evolved fear module argue that affect is a relatively independent system that is largely unaffected by cognition. Mallan and colleagues however dispute the treatment of social stimuli as prepared stimuli (Mallan, Lipp & Cochrane, 2013). Mallan et al. (2009) for example provided evidence that affect towards conditioned social stimuli were malleable to cognition. They used a typical conditioning procedure and paired either an ingroup or outgroup face with an electrical stimulation, but never another ingroup or outgroup face. In addition, to check the degree of cognitive insulation the type of information given to participants was manipulated. One group of participants were told prior to extinction that no electrical stimulations would be presented, whilst the other group were given no prior instructions. Fear readily extinguished for participants who were conditioned with outgroup faces and given prior instructions that no electrical stimulation would be administered, and for ingroup faces. In contrast, resistance to extinction was found in participants who were conditioned with outgroup faces and given no prior instructions. This implies outgroup faces do not act as prepared stimuli because instructions, which involve human cognition, modified conditioned fear of outgroup faces during extinction. Hence, by failing a key requirement of an evolved fear module, that being encapsulated from more advanced human cognition, Mallan et al. (2013) argue social stimuli are not prepared stimuli and their research suggests relative dependent systems between cognition and affect when social stimuli are involved.

The ‘How’ to Exploring the Interplay in This Research

To investigate the interplay between cognition and affect in my research, I explore the impact of extinction, affect-/cognitive-laden mediators/moderators, measurement controllability and contingency awareness have.

Extinction is a key factor that can be investigated to explore the relationship between cognition and affect. It has been argued that extinction disrupts previously learnt associations, which results in conditioned responses becoming more ambiguous or less influential/predictive than during acquisition (Bouton, 2002; Hermans, Craske, Mineka & Lovibond, 2006; Shechner, Hong, Britton, Pine & Fox, 2014). Previous associations are not necessarily unlearnt, which is evident by anxiety towards stimuli being spontaneously recovered (Bouton, 2002; Rescorla, 2004), reinstated (Bouton, 2002; Norrholm et al., 2006) or renewed (Bouton, 2002; Holmes & Westbrook, 2014; Neumann, Lipp & Cory, 2007). Rather, extinction creates new associations that compete with previously learnt associations. From this angle, learning new associations during extinction (aka a cognitively driven process) disrupts or reduces affect responding; therefore extinction is an ideal context where to study the affect-cognition interplay. I will explore the implication of extinction by looking at prototypicality shifts straight after acquisition compared to after extinction. A larger prototypicality shift effect present after acquisition (vs after extinction) is labelled in my research as result due to ‘affect’, i.e., affect is higher after acquisition because of developed associations; whereas a larger prototypicality shift effect present after extinction (vs after acquisition) is labelled in my research as a result due to ‘cognition’, i.e., affect is not as prominent because of developed associaitons are extinguished and decisions are based more on cognitions.

There are a host of other *mediators/moderators*, including number of repeated exemplar presentations during extinction, exemplar level anxiety and absolute levels of anxiety after acquisition that I briefly explore to provide further insight into the relationship between affect and cognition in my focal effects. The number of repeated exemplar presentations during extinction was used to test the idea that cognition might be influential in prototypicality shifts. For example, research on extended extinction studies suggest re-acquisition is less likely to occur with greater number of extinction trials (Weidemann & Kehoe, 2003). The implication of this for my research is following a greater number of extinction trials my measure of prototypicality is less likely to be influenced by affect. Exemplar-level anxiety and absolute level of anxiety at post-acquisition were other factors used to test the idea that affect might be influential in prototypicality shifts if they are larger post- acquisition. I briefly explore these factors to determine if they influenced the degree of affective and cognitive processing involved in prototypicality shifts.

Another way of exploring the relationship between cognition and affect in this research is in terms of *controllability* of categorisation responses as a function of *type* of dependent measure. Human behaviours operate at different levels of control and these variations can be gauged with explicit (deliberate mindful control) and implicit measures (spontaneous mindless control; Cunningham, Raye & Johnson, 2004; Gawronski & Bodenhausen, 2006; Maass, Castelli & Arcuri, 2000). Determining whether exemplar prototypicality shifts are influenced by participant’s intentional control over their responses provides insight into the level of cognitive and deliberate processes required. Prototypicality shifts detected on explicit measures imply a high degree of cognitive elaboration because participants are aware of the decisions being made. In contrast, prototypicality shifts detected on implicit measures would imply that a high degree of cognitive processing is not required, because these measures are more dependent on factors outside conscious awareness. Thus, detecting prototypicality shifts using different dependent measures provides insight into the degree of cognitive processing (self-reported/high vs. implicit/low).

Contingency awareness is another factor that can be used to explore the relationship between cognition and affect in prototypicality shifts. As described in Chapter 2, contingency awareness typically refers to the explicit (cognition-laden) detection of which exemplar is paired with an electrical stimulation in a conditioning procedure (Kattner, Ellermeier & Tavakoli, 2012; Lovibond & Shanks, 2002; Stahl, Unkelbach & Corneille, 2009). A large body of empirical research demonstrates that conditioning effects on exemplar-level affect and evaluations are more robust when participants are aware of the predictive relationship, than when they are not (Hofmann, De Houwer, Perugini, Baeyens & Crombez, 2010; Lovibond, Liu, Weildemann & Mitchell, 2011; Lovibond & Shanks, 2002; Pleyers, Corneille, Luminet & Yzerbyt, 2007). The effect contingency awareness has on prototypicality shifts however has not been investigated. If contingency awareness moderates prototypicality shifts in that the shift is larger among contingent aware than unaware participants, this result would indicate a stronger modulation of cognitions. If on the other hand prototypicality shifts are observed in the absence of contingency awareness, it would indicate that prototypicality shifts are driven by more automatic/affective processes. Hence, because contingency awareness requires a large amount of cognitive processing, it can be used to investigate the interplay between affect and cognition in my paradigm.

In summary, understanding the relationship between cognitive and affective processing is not a key aim of this research but is explored to begin understanding factors that influence prototypicality shifts. As these are secondary aims my research was not always adequately equipped to provide closure and caution needs to be taken interpreting this data (i.e., contingency awareness analyses have small group sizes). There is a body of literature that suggests cognitive processing is relatively independent from affective processing, whilst other research argues for a relatively dependent system. I investigate the role of extinction, measurement controllability, contingency awareness and other possible factors as a first step in understanding the interplay between cognitive and affective processing involved in exemplar prototypicality shifts.

Study 3.1

Method

Participants and Design

Participants were 66 students (22 male, 44 female; *M* = 21.26 years, *SD* = 3.92) from a large regional Australian university. All participants were White and reported an Anglo-Saxon background. They received a monetary compensation (AUS $25) or partial course credit for their participation. Participants were randomly assigned to one of the two types of aversive learning (direct *n* = 34, vicarious *n* = 32). In this first study, post-test prototypicality measures were collected exclusively after the extinction procedure (i.e., after extinction).

Materials

Outgroup faces were developed using the face morphing software FaceGen v3.3.1 and are observable in Figure 4. Two *target faces* were used in the conditioning task as safe versus unsafe faces (counterbalanced) to test for basic prototypicality shifts, while controlling for non-associative processes (Rescorla, 1988); these were two Black 25yrs-old male faces of neutral expression and front orientation. To test for generalised prototypicality shifts, I used six other faces of the same race, age, gender, and expression as the target faces. Four *variations* were 25% and 50% variations of each target face and moved progressively away from the Black prototype along a generalisation gradient (towards a computational average of all FaceGen White, Black, Middle-eastern and Asian faces). Two *new* (prototypically Black) *faces* were also generated as unrelated to the target faces (see Figure 4).

Fifteen pilot participants rated the eight faces as part of a larger set of Black and other ethnicity faces along perceived prototypicality and anxiety (1 = *not at all*, 6 = *very much*). The four faces (two targets and two new faces) were rated as high in perceived

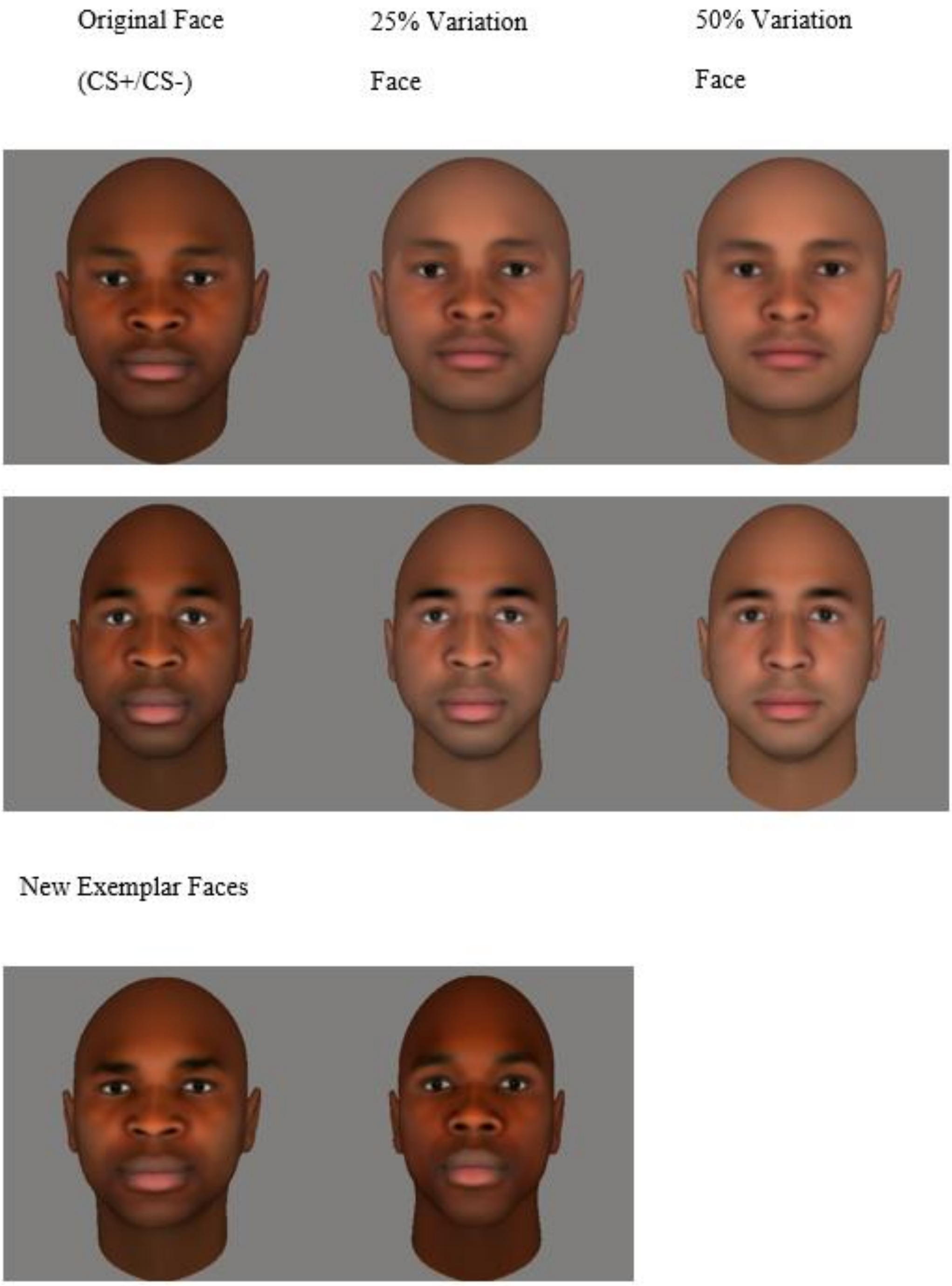


Figure 4. Target, variation and new exemplar faces

prototypicality (Grand *M* = 5.58, *SD* = .87) and low in anxiety (Grand *M* = 2.33, *SD* = 1.15), and statistically comparable along these dimensions, all *p'*s > .05. The two configurally-related faces associated with each of the two target faces (25% and 50% variations) were significantly different in prototypicality, both *p'*s < .05, and always followed a gradient along target, 25%, and 50% (set 1, *M*s [*SDs*] = 5.73[.80], 5.33 [.90] and 4.87 [1.19] ; set 2, *M*s [*SDs*] = 5.20 [1.37], 5.07 [1.22], and 4.13 [1.36]); face differences in anxiety were comparatively slimmer, both *p'*s < .13, but also followed a gradient (set 1, *M*s [*SDs*] = 2.27 [1.03], 1.93 [1.03], and 1.87 [.83]; set 2, *M*s [*SDs*] = 2.33 [1.23], 2.13 [1.30], and 2.00 [1.00]). Hence, selected outgroup faces were suitable to test for basic prototypicality shifts of the two target faces as well as generalised shifts in prototypicality of the face variations and new faces (Lissek et al., 2010).

Procedure and Measures

To minimise response biases caused by repeated measurements, participants provided pre-test exemplar ratings in a first laboratory session approximately five days prior to attending an individual laboratory session (second laboratory session). As part of a larger on-line questionnaire, participants indicated the extent to which each of the eight (randomly ordered) Black faces were prototypical of Black people in general (*prototypically Black*: 1 = *not at all*, 7 = *very much*). I also asked participants to rate how anxiety provoking the faces would be if they were to meet them (*anxious*: 1 = *not at all*, 7 = *very much*) and how similar each pair of faces were (*similar*: 1 = *not at all*, 7 = *very much*).

Once in the laboratory for their individual testing session, participants were seated in front of a computer screen. Initially a work-up procedure was completed to select a level of shock participants regarded as uncomfortable but not painful (Lovibond, Saunders, Weidemann, & Mitchell, 2008). The level of shock was determined by participants progressively sampling different intensities. Type of aversive learning (direct vs. vicarious) was then manipulated adapting Olsson, Phelps, and colleagues' methods (Olsson & Phelps, 2004; Olsson, Nearing, & Phelps, 2007). During the acquisition task, the two target faces appeared on the screen five times each for 10s (inter-stimulus-interval M = 30 s, range 20-40 s). For direct learning participants, one target face (unsafe face) always co-terminated with a 200 ms electro-tactile stimulation delivered to the participant's finger at the level selected during the work-up procedure and the other target face (safe face; stimuli counterbalanced) was never paired with the electro-tactile stimulation. Target faces were presented in a randomised order for direct participants. Vicarious aversive learning participants were led to believe they would also undergo electro-tactile stimulation themselves after watching a video of a study similar to the one in which they were about to participate. Hence, for these participants, the study consisted of the same sequence of Black faces on the left side of the screen and a 6-minute video filmed for the purpose of this study on the right displaying a young White female individual seated in front of a computer screen and presenting signs of bodily arousal (facial expressions and posture) during unsafe face presentation and signs of bodily/face relaxation during safe face presentation. Target faces were presented in a fixed pseudo-randomised order in the vicarious learning condition due to technical restrictions associated with the video.

I checked for differential conditioning by connecting all participants to psychophysiological equipment (skin conductance electrodes) at the very beginning of the second laboratory session and assessing changes in physiological activation for the duration of the conditioning task and during presentation of the eight faces immediately prior to, and after, the acquisition. Thus, pre-test and post-test skin conductance responses were collected immediately prior to and after acquisition (but before extinction), by presenting the unsafe and safe face twice and each generalisation face once using the timing described for conditioning. Skin conductance was measured via an ADInstruments Model ML116 GSR amplifier and standard MLT116F electrodes attached to the distal phalanges of the middle and ring fingers of the left hand.

Respiration was also monitored to check for artefacts using an ADInstruments MLT1132 piezo respiration belt transducer attached around the chest. All physiological data were recorded using computer interfaces with an ADInstruments Powerlab Model 8/30 data acquisition system.

Before the extinction procedure occurred, all physiological equipment was removed from participants to make them more comfortable. During subsequent physiological extinction, the safe and unsafe target faces were presented – always an identical number of times – in the absence of any electro-tactile stimulation until participants displayed no further reduction in physiological activation over four consecutive presentations (the number of presentations was set to a minimum of 5 and a maximum of 25 presentations per target face). At the end of extinction, post-test ratings of exemplar prototypicality (as well as self-reported exemplar anxiety) were recorded again and compared with pre-test face ratings from the first laboratory session. Participants were finally debriefed, thanked and dismissed.

Results and Discussion

Overview of Analyses

Skin conductance responses (SCRs; collected pre-/post-acquisition and post-extinction) and self-reported anxiety (collected at pre-test and post-extinction) were first analysed and treated as manipulation checks to ascertain that negativity/anxiety was acquired and extinguished as a function of acquisition and extinction respectively.

Main analyses for my prototypicality measures (collected at pre-test and post-extinction), was assessed via self-reports. I first compared pre-test and post-extinction measures for target faces to determine whether exemplar prototypicality shifted following an association with negativity/anxiety and extinction. Type of learning was included as a between subjects variable to determine if exemplar prototypicality shifts were similar in magnitude between direct and vicarious learning conditions. I then extended these analyses to variations of faces and new exemplars to determine whether prototypicality shifts generalised to similar exemplars.

In this chapter a secondary aim was to explore the interplay between affect (negativity/anxiety) and cognitions (prototypicality). To this end, a mediational analysis was carried using exemplar specific change in SCRs from pre-acquisition to post-acquisition and exemplar specific changes in exemplar prototypicality from pre-test to post-extinction to determine if greater SCR activity resulted in greater prototypicality shifts. For readers with an interest in affective learning, ancillary results for pre-/post-acquisition changes in generalisation data on SCRs are reported in the Appendix P1.

Checking Effective Acquisition and Extinction

To investigate the extent to which changes in affect played a role in the prototypicality shifts, I checked for successful acquisition and extinction. Successful acquisition and extinction was first checked with SCRs collected immediately after acquisition. SCRs were recorded with a sampling rate of 1k/s and analysed using ADInstruments Chart v5 following standard guidelines (Boucsein et al., 2012; Fowles, Christie, Edelberg, Grings, Lykken & Venables, 1981). Specifically, first interval responses were used as my measure of physiological arousal towards exemplars and this was defined as the amplitude of the skin conductance response that began 1 – 4 seconds after the CS+ or CS- face onset. Amplitude was defined as the distance measured in microsiemens (μS) between the trough and apex of the response curve. The minimum amplitude for a response was 0.05 μS. Pre-test SCRs elicited during the CS+ and CS- was calculated as the mean response across the two presentations of each CS type (randomised order) before acquisition occurred. Post-test SCRs during the CS+ and CS-was calculated as the mean activity across two presentations of the CS+ and CS- after acquisition occurred. SCRs elicited through extinction were calculated in the same manner. Across all acquisition and extinction blocks a response was considered missing if breathing or movement artefacts were present. Data considered missing were not replaced with any values. As commonly performed to normalise the SCRs as much as possible, a square root transformation of CS+ and CS- data was performed on averaged SCR data (Dunsmoor, Murty, Davachi & Phelps, 2015; Esteves, Parra, Dimberg & Ohman, 1994). A constant of 1 was added to data prior to the square root transformation.

A 2 face type x 2 time (pre-test, post-test) x 2 learning type mixed model ANOVA was conducted on SCRs, with face type and time as the repeated measures to check if *target* faces became associated with the aversive stimulus. I found that direct and vicarious participants displayed similar levels of magnitude on skin conductance responses from pre- to post-test towards the unsafe face but not the safe face, evidenced by a face type x time interaction F (1,64) = 40.12, p <.01 η_p^2 = .39 and non-significant face type x time x learning interaction. A paired samples t-test confirmed physiological activity was higher for the unsafe face at post-test (M = 1.38, SD = .44) than at pre-test (M = 1.03, SD = .10), t (65) = -6.47, p <.001. A smaller level of physiological activity was observed for the safe face at post-test (M = 1.10, SD = .18) than at pre-test (M = 1.04, SD = .13), t (65) = -2.29, p = .025. Hence, higher SCR responses towards the unsafe face and is indicative of successful acquisition and negativity/anxiety being more associated with the unsafe face than the safe face (these anxiety learning effects are discussed extensively elsewhere; Harris, Griffin & Paolini, 2016a; Harris, Griffin & Paolini, 2016b).

SCR results suggest acquisition was successful and the unsafe face was perceived to be more negative/anxiety provoking following acquisition compared to the safe face. I expanded the model and included *variations* of the safe and unsafe target faces uninvolved in conditioning using a 2 learning type x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time (pre-, post-test) mixed-model ANOVA with face type, generalisation gradient, and time as repeated measures. A significant face type by time by generalisation gradient interaction was observed, F (2, 128) = 13.70, p < .001, η_p^2 = .18. I examined the 3-way interaction by testing for the face type by time interaction at each level of the generalisation gradient, with unsafe and safe face breakdown previously presented.

A face type by time interaction was detected for the 25% variation faces, F (1, 65) = 20.80, p < .001, η_p^2 = .24. The unsafe face had higher SCR's at post-test (M = 1.16, SD = .28) than at pre-test (M = 1.03, SD = .09), t (32) = -4.39, p <.001. In contrast, there was no difference between the safe face at pre-test (M = 1.02, SD = .04; post-test M = 1.02, SD = .07), t (65) = -.74, p = .461. Together, these results suggest anxiety generalised to the 25% variation of the unsafe but not the safe face.

For the 50% variation faces a time main effect was observed, F (1, 65) = 15.28, p < .001, η_p^2 = .19. Across both learning types a non-associative change in negativity/anxiety of both the unsafe and safe faces was found from pre-test (M = 1.04, SD = .07) to post-test (M = 1.13, SD = .15). This result suggests that although discriminative changes in anxiety were detected for target exemplars, non-associative changes in anxiety were detected for 50% variation faces in both direct and vicarious learning conditions.

To test whether negativity/anxiety generalised to new exemplars, participants pre-test face similarity ratings were used to create a new between subjects variable as described in the prototypicality analysis section. The new variable distinguished between those who perceived each new exemplar as being more similar to the unsafe face than the safe face, as being more similar to the safe face than the unsafe face, or equally similar to both. Individual analysis were carried out for each new Black exemplar using a 2 learning type x 3 new face similarity (similar to the unsafe face, similar to the safe face, equally similar to the unsafe and safe face) x 2 time mixed model ANOVA with time as the repeated measure on SCR data. For the first new exemplar (see Figure 4 in main text), a significant new face similarity x time interaction was detected, F (2, 59) = 15.34, $p < .001$, $\eta_p^2 = .3$. This interaction was followed up by looking at each level of face similarity separately. Unexpectedly, when the new face was perceived as being more *similar* to the *safe* face than the unsafe face, an *increase in SCR* was found from pre-test (M = 1.06, SD = .12) to post-test (M = 1.45, SD = .38), F (1, 20) = 22.47, $p < .001$, $\eta_p^2 = .53$.. For the second new exemplar a significant new face similarity x time interaction was detected, F (2, 59) = 14.49, $p < .001$, $\eta_p^2 = .33$, and subsequently followed up. Unexpectedly, again when the new face was perceived as being more similar to the safe face than the unsafe face, an increase in SCR was found from pre-test (M = 1.05, SD = .13) to post-test (M = 1.65, SD = .63), F (1, 21) = 17.93, $p < .001$, $\eta_p^2 = .46$. New exemplar SCR data suggests anxiety was higher when new exemplars were perceived as being more similar to the safe face than the unsafe face. This effect is in contrast to contingent specific SCR results that found higher SCR responses for the unsafe face than the safe face.

To check for extinction, I ran a 2 learning type (direct vs vicarious) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. Extinction data analyses help to provide clarity over the role that affect has on cognitions because prototypicality shifts observed post-extinction would suggest a partially independent system that does not require changes in affect to hold. SCR data during the first block of extinction refers to the first two presentations of the unsafe and safe face and the ten trial block refers to the last two presentations of the unsafe and safe face up to the tenth trial (i.e., Presentation 7, 8, 9 and 10 of faces during extinction. I chose these two timepoints because all participants were exposed to a minimum of ten extinction trials. A main effect of face type, F (1, 64) = 37.68, p <.001, η_p^2 = .37 and a face type x extinction block interaction were detected, F (1, 64) = 13.05, p = .001, η_p^2 = .17. The main effect of face type suggests the unsafe face had higher SCRs throughout the first 10 extinction trials (*M* = 1.18, *SD* = .23) than the safe face (*M* = 1.01, *SD* = .02). The face type x extinction block interaction was followed up with a paired sample *t*-test and found the unsafe face significantly decreased in SCR activity at the ten trial block (*M* = 1.13, *SD* = .23) compared to the first block in extinction (*M* = 1.23, *SD* = .23), t (65) = 3.73, p <.001. The decrease in SCR is indicative of successful extinction. I also supplemented this analysis in two ways. First, by investigating whether the last two presentations of the unsafe and safe face varied to post- acquisition data collected before extinction. Second, by determining whether the last two presentations of the unsafe and safe face increased significantly from 0. Supplemental data analyses conducted on SCR suggest anxiety extinguished and are discussed extensively in Appendix P1. Altogether, SCR data suggests higher negativity/anxiety developed during acquisition extinguished following extinction.

In addition, I checked for successful extinction with self-reported anxiety data collected after extinction. A 2 learning type (direct, vicarious) x 2 face type (safe, unsafe) x 2 time (pre-, post-extinction) mixed model ANOVA with face type and time as repeated measure factors was performed and detected no significant effects, all *p*'s >

.10. This suggests there were no differences between the unsafe and safe face and extinction was successful.

In summary, acquisition was successful and changes in negativity/anxiety towards the unsafe face were found post- acquisition. Extinction data suggests the association developed during acquisition was made ambiguous by the end of extinction.

Testing Basic and Generalised Prototypicality Shifts

I tested for *basic* shifts in *target faces*' outgroup prototypicality with a 2 learning type (direct, vicarious) x 2 face type (safe, unsafe) x 2 time (pre-, post-test) mixed-model ANOVA with face type and time as repeated measures on self-reported prototypicality. I detected a significant face type by time interaction, $F\ (1, 64) = 23.81$, $p < .001$, $\eta_p^2 = .27$ (see Figure 5). As predicted, the unsafe face was regarded more prototypically Black after extinction ($M = 5.94$, $SD = 1.04$) than before acquisition ($M = 4.08$, $SD = 1.66$), $F\ (1, 65) = 60.05$, $p < .001$, $\eta_p^2 = .52$; no change occurred for the safe face from pre-test ($M = 4.23$, $SD = 1.62$) to post-test ($M = 4.21$, $SD = 1.63$), $F < 1$. As the 3-way interaction between face type, time and learning condition was non-significant, $F < 1$, these prototypicality shifts were of comparable magnitude in the direct and vicarious conditions[1]. Hence, the outgroup face that was associated—directly or vicariously—with incidental negativity displayed the expected increase in prototypicality; no changes occurred for the safe target face.

I tested the extent to which changes in prototypicality caused by aversive conditioning generalised to *variations* of the safe and unsafe target faces uninvolved in conditioning using a 2 learning type x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed-model ANOVA with face type, generalisation gradient, and time as repeated measures. A significant face type by time by generalisation gradient interaction was observed, $F\ (2, 128) = 5.41$, $p < .001$, $\eta_p^2 = .08$. I examined the 3-way interaction by testing for the face type by time interaction at each level of the generalisation gradient. As the effect in Figure 5 is the same as that reported earlier for the target faces, the follow up analyses focused on the data for the 25% and 50% variations.

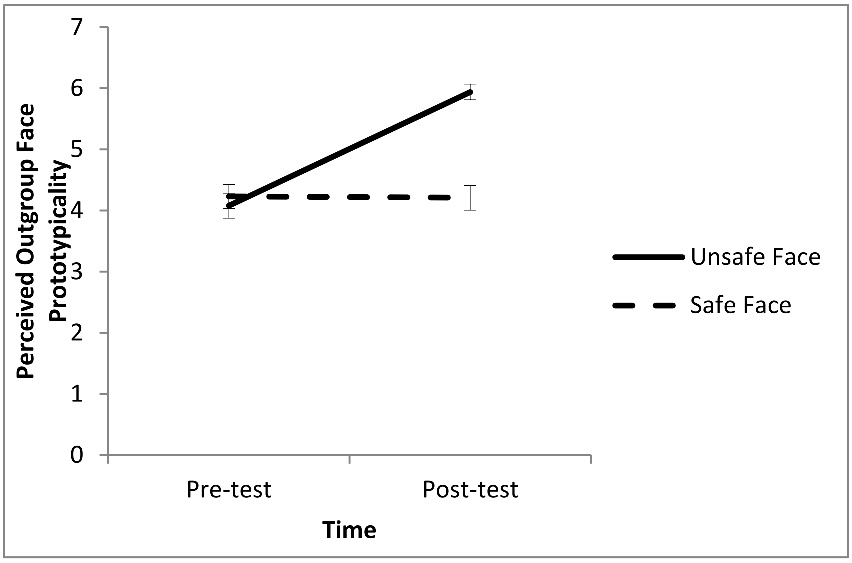


Figure 5. Face type x time interaction on perceived prototypicality of target outgroup faces

A face type by time interaction was also detected for the 25% variations, F (1, 65) = 49.08, $p < .001$, $\eta_p^2 = .43$ (see Figure 6). Consistent with a discriminative generalisation effect, the 25% unsafe face variation also increased in prototypicality over time, F (1, 65) = 29.77, $p < .001$, $\eta_p^2 = .31$ (pre-test $M = 2.86$, $SD = 1.44$; post-test $M = 4.47$, $SD = 1.73$). The 25% safe face variation instead displayed a contrast effect with a significant decrease in prototypicality over time, F (1, 65) = 10.64, $p < .01$, $\eta_p^2 = .14$ (pre-test $M = 2.68$, $SD = 1.43$; post-test $M = 2.06$, $SD = .97$). Hence, in addition to shifting the prototypicality of the faces directly involved in the aversive experience, direct and vicarious aversive learning also made the variation face configurally most similar to the unsafe target face perceptually more prototypically Black, and the variation face configurally most similar to the safe target face perceptually less prototypically Black.

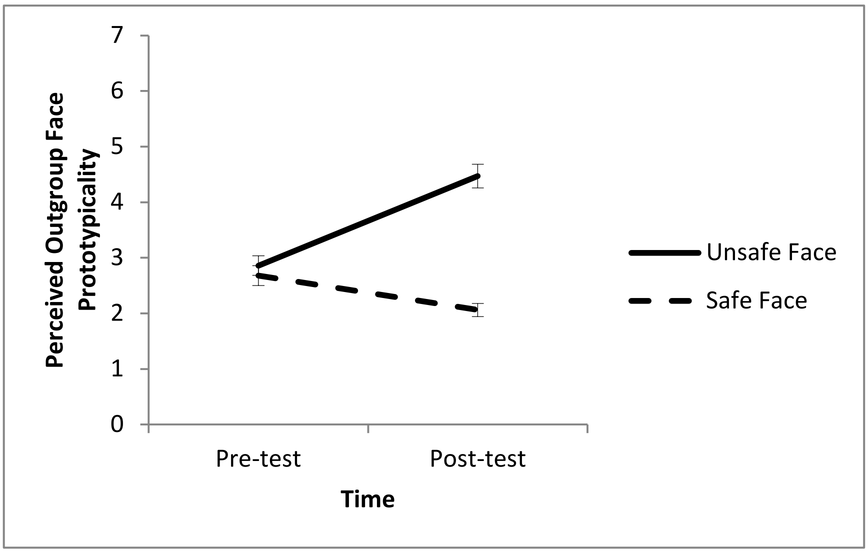


Figure 6. Face type x time interaction on perceived prototypicality of 25% variations of target faces

For the 50% face variations, the face type by time interaction was non-significant, $F < 1$. There was only a weak trend for an overall reduction of prototypicality across face type over time (see Figure 7); face type $p = .11$ (unsafe M = 4.07, SD = 1.93; safe M = 3.80, SD = 2.13); time $p = .15$ (pre-test M = 4.14, SD = 1.92; post-test M = 3.74, SD = 2.15). Altogether, the generalisation tests on variation faces indicated that discriminative (unsafe vs safe face) prototypicality shifts extended from the target faces to the 25% face variations, but not to the 50% face variations[2].

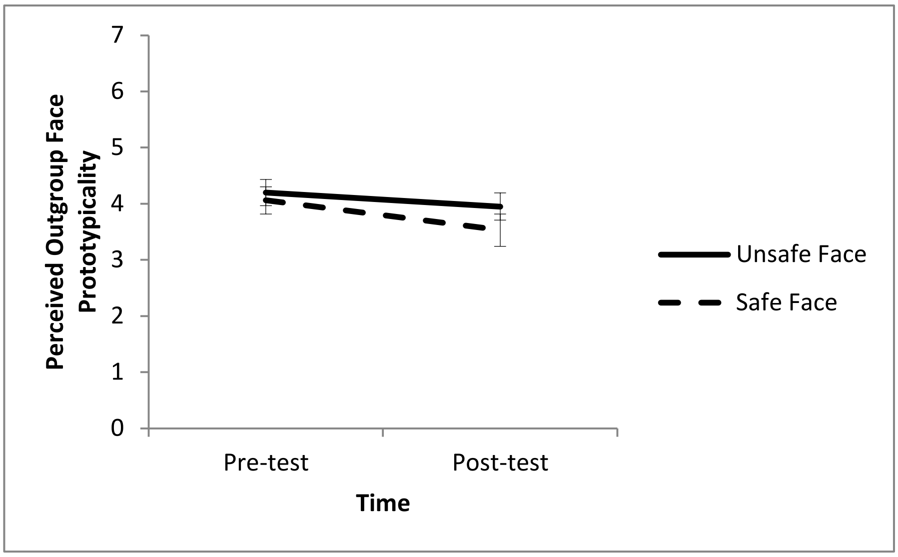


Figure 7. Face type x time interaction on perceived prototypicality of 50% variations of target faces

To test whether discriminative prototypicality shifts generalised to new exemplars, participants' pre-test face similarity ratings were used to create a new between subjects variable. The new variable distinguished between those who perceived each new exemplar as being more similar to the unsafe face than the safe face (new exemplar 1 N = 25, new exemplar 2 N = 16), as being more similar to the safe face than the unsafe face (new exemplar 1 N = 22, new exemplar 2 N = 23), or equally similar to both (new exemplar 1 N = 19, new exemplar 2 N = 27)[3]. Individual analysis were carried out for each new Black exemplar using a 2 learning type x 3 new face similarity (similar to the unsafe face, similar to the safe face, equally similar to the unsafe and safe face) x 2 time mixed model ANOVA with time as the repeated measure on self-reported prototypicality data. For the first new exemplar, a significant learning type x time, F (1, 59) = 4.81, p = .01, η_p^2 = .08 and learning type x new face similarity x time interaction was detected, F (2, 59) = 3.84, p = .027, η_p^2 = .12. I followed up the 3 way interaction by looking at each level of learning type and new face similarity separately. In the direct learning condition, when the new exemplar was perceived as being more similar to the safe face a decrease in prototypicality was observed from pre-test (M = 6.20, SD = .92) to post-test (M = 4.20, SD = 1.32), $t(9) = 4.05$, $p = .003$. All other effects were non-significant, all p's > .05. For the second new exemplar a main effect of time, $F(1, 59) = 4.12$, $p = .047$, $\eta_p^2 = .07$, and a learning type x time interaction was detected, $F(1, 59) = 4.01$, $p = .05$, $\eta_p^2 = .06$. I followed the learning type x time interaction by analysing each level of learning type separately. In the direct condition I found prototypicality decreased from pre-test (M = 5.47, SD = 1.11) to post-test (M = 4.61, SD = 1.33), $t(33) = 3.51$, $p = .001$. No effects were found in the vicarious learning condition, all p's > .929.

In summary, these prototypicality results provide initial evidence that outgroup exemplars paired with an incidental aversive experience become more outgroup-like – as reflected in changes in self-reported prototypicality ratings. Configurally closely related exemplars were also rated higher on outgroup prototypicality following acquisition and extinction, demonstrating that the observed effect can generalise to other, similar outgroup exemplars. These prototypicality shift effects were of a comparable magnitude in the direct and vicarious learning conditions, indicating that vicarious experiences with an outgroup member are just as effective at changing exemplar perceptions, as a firsthand experience. There is little evidence for the effects of the unsafe face prototypicality generalising to faces that are perceived as being similar.

Exploring the Role of Mediators in Basic Prototypicality Shifts

I tested whether SCR changes in face anxiety, collected after acquisition but before extinction, mediated prototypicality shifts. For this, I entered an index of pre- vs. post-test change in negativity/anxiety to the safe face, relative to the unsafe face for target and generalisation faces, respectively[4], as covariate in our key mixed-model analyses (see Baron & Kenny, 1981; Judd, Kenny & McClelland, 2001; Yzerbyt, Muller, & Judd, 2004 for an extensive discussion of mediation tests with within-subject factors). Prototypicality shifts remained substantially unchanged when controlling for changes in physiological anxiety immediately after acquisition: basic shift, from F (1, 64) = 23.81, $p < .001$, $\eta_p^2 = .27$ to F (1, 63) = 18.14, $p < .001$, $\eta_p^2 = .22$; generalised shift, from F (2, 128) = 5.41, $p < .01$, $\eta_p^2 = .078$ to F (2, 126) = 6.85, $p < .01$, $\eta_p^2 = .098$); hence, there was no evidence that changes in affect associated with exemplars as a result of conditioning were pivotal for our prototypicality effects. This additional step was used to explore the relationship between cognition and affect in basic and variation prototypicality shifts and results suggest a partially independent system.

Overall, results from Study 3.1 provided evidence for the existence of an exemplar prototypicality shift following an association with a negative/anxiety provoking stimulus. Evidence for prototypicality shifts to generalise to similar exemplars was also provided. Despite establishing the existence of the prototypicality shift, the underlying psychological mechanisms need to be investigated further.

Study 3.2

Study 3.1 provided evidence that prototypicality shifts did not require changes in exemplar-level negativity/anxiety to hold post-extinction. Study 3.2 provided a better test of the relationship between cognition/affect by varying the position of the post-test prototypicality measurement. To investigate this possibility, I varied the position of the post-test prototypicality data collection between-subjects to be either immediately *after* acquisition but *before* the extinction procedure (or 'after acquisition' condition), or after both conditioning *and* extinction procedures as in Study 3.1 (or 'after extinction' condition). If exemplar-specific heightened negativity/anxiety plays a key role in modifying an exemplar's prototypicality, I would expect the magnitude of the prototypicality shift to be larger when negativity/anxiety is at its highest i.e., after acquisition. However, if exemplar-specific heightened negativity/anxiety does not play a key role, prototypicality shifts collected after acquisition and extinction (after extinction condition) would be of a similar magnitude.

To investigate the controllability of the prototypicality shift, I used an implicit measure of categorisation to measure perceptual changes in exemplar prototypicality. White participants completed a speeded sorting task (Richeson & Trawalter, 2005; Ruys et al., 2008): They sorted White and Black exemplars as quickly and as accurately as possible between the White and Black category (Black exemplars presented in this task were the same as presented during conditioning). The more prototypical an outgroup face is of the outgroup category, the faster it should be sorted into that category because of its psychological proximity to the group prototype. Hence, I expected the outgroup unsafe face (and its variations) to be sorted significantly faster than the outgroup safe face and its variations after conditioning. Finally, in this study I investigated the role of several other potential underlying mechanisms of this novel prototypicality shift effect. To this end, I compare different levels of contingency awareness and used mediational analysis to investigate the role played by repeated exemplar presentations during extinction, and overall anxiety at post-test (as indexed by the average face anxiety).

Method

Participants and Design

Participants were 61 students (20 male, 41 female; *M* = 20.67, *SD* = 3.59) from a large regional Australian university. Similar to Study 3.1, participants reported as being from a White, Anglo-Saxon background. Again, participants received monetary compensation (AUS $25) or partial course credit for their participation. Participants were randomly assigned to one of four conditions of a 2 learning type (direct n = 31, vicarious n = 30) x 2 post-test position (after acquisition n = 31, after extinction n = 30) between-subject experimental design[6].

Procedure and Measures

The procedure and materials used in Study 3.2 were similar to those used in Study 3.1 and the same Black target, generalisation and new exemplars were used again. Response biases from repeated measurements were again minimised by collecting pre-test data some time before post-test data collection (between 5 and 28 days; M = 11.4, SD = 5.35). This time, however, these data were also collected in the laboratory (vs. online in Study 3.1) so to access a specialised software (E prime 2.0) for the speeded sorting task. This initial laboratory session involved participants sitting in front of a computer and completing two computerised tasks: a speeded sorting task, which yields a well-established implicit measure of perceived exemplar prototypicality (e.g., Richeson & Trawalter, 2005; Ruys et al., 2008), and an online questionnaire.

The second laboratory session involved participants completing the work up procedure described in Study 3.1. Type of aversive learning (direct vs. vicarious) was again manipulated between-subjects, using the same procedure described in Study 3.1. The key difference from Study 3.1 involved the manipulation of the position of post-test data collection. Half the participants completed post-test data collection (sorting task and online questionnaire) before the extinction procedure (after acquisition condition), while the other half completed post-test data collection after acquisition and extinction (after extinction condition). The same extinction procedure described in Study 3.1 was used, but this time the *number of repeated presentations* during extinction were recorded to determine the mediational effect of the length of extinction.

During the speeded sorting task, participants were presented with individual Black and White targets, generalisation and new exemplar faces at the centre of the computer screen. White faces were chosen from the bank of pilot tested faces (see Study 3.1) and paralleled the Black face set in terms of target, generalisation and new exemplar faces (i.e., 2 target faces, 4 variation faces and 2 new exemplars). These faces were selected on the basis of comparable prototypicality and anxiety ratings to the Black faces[5]. Participants were instructed to sort each individual face as quickly and as accurately as possible into the “Black” or “White” categories by pressing the green (left handed “S” key) or blue buttons (right handed “L” key) on the keyboard. The category labels were presented in the top left and right corners of the screen and corresponded to the location of the relevant key. Each face was presented 14 times and response keys and category labels were counterbalanced for half of the presentations. Faces were inverted for 25% of the presentations to increase task difficulty and task engagement (Richeson & Trawalter, 2005).

The online questionnaire was used in this study to collect self-reported prototypicality, anxiety and similarity data. In addition, *contingency awareness* was collected to determine the mediational effect this variable had. Post-test prototypicality and anxiety data collection for the online questionnaire occurred either immediately after acquisition or after extinction, like for the sorting task data collection. Contingency awareness was collected post-extinction for all participants. Participants indicated the extent to which each of the eight (randomly ordered) Black faces were prototypical of Black people in general (*prototypically Black*: 1 = *not at all*, 7 = *very much*). I also asked participants to rate how anxiety provoking the faces would be if they were to meet them (*anxious*: 1 = *not at all*, 7 = *very much*) and how similar each pair of faces were (*similar:* 1 = not all, 7 = *very much*). Self-reported anxiety provides a direct measure of anxiety that does not rely on skin conductance responses, which often extinguish quicker compared to evaluations (Lipp, Oughton & LeLievre, 2003). To check whether participants were aware of the unsafe face-electrical stimulation pairing, a series of funnelling questions were asked post-test after the extinction procedure (adapted from Clark & Squire, 1998; Page, 1973; see Appendix O). These questions involved asking participants (1) if they reacted the same way to all faces (2) whether there was a pattern in which face was paired with the electrical stimulation (3) which face was paired with the electrical stimulation and (4) how confident were in their selection of the face paired with the electrical stimulation. In order to be considered contingent aware participants needed to accurately select the face paired with shock (the two target faces were presented as options), with a high degree of confidence (i.e., On a scale 1 -7, a rating of 4 or more on how confident they were the face they selected was paired with the electrical stimulation). Participants also needed to accurately describe the association between the exemplar and electrical stimulation in an open-ended response. Fourteen participants were classified as contingent non-aware and 16 contingent aware in the after acquisition condition; 7 participants classified as contingent non aware and 22 contingent aware in the after extinction condition.

Results and Discussion

Overview of Analyses

SCRs (collected pre-/post-acquisition and post-extinction) and self-reported anxiety (collected at pre-test and either post- acquisition or post-extinction) were first analysed and treated as manipulation checks to ascertain that negativity/anxiety was acquired and extinguished as a function of acquisition and extinction respectively.

Main analyses for my prototypicality measures (collected at pre-test and either post- acquisition or post-extinction), was assessed via self-reports and a speeded sorting task. I first compared pre-test and post-test measures for target faces to determine whether exemplar prototypicality shifted following an association with negativity/anxiety and extinction. Post-test measures were collected at either post-acquisition or post-extinction and this manipulation was included as a between subjects variable to provide a better indication of the interplay between affect and cognition. Type of learning was also included as a between subjects variable to determine if exemplar prototypicality shifted in a similar magnitude between direct and vicarious learning conditions. I then extended these analyses to variations of faces and new exemplars to determine whether prototypicality shifts generalised to similar exemplars.

In this chapter a secondary aim was to explore the interplay between affect (negativity/anxiety) and cognitions (prototypicality). To this end, a mediational analysis was carried using post-test acquisition data for both exemplars to provide a measure of general anxiety and exemplar specific changes in exemplar prototypicality from pre-test to post-extinction to determine if anxiety in general resulted in greater prototypicality shifts. For readers with an interest in affective learning, ancillary results for pre-/post-changes in generalisation data on SCRs and self-reported anxiety are reported in the Appendix P2.

Checking Effective Acquisition and Extinction

To ascertain whether changes in prototypicality were linked to changes in affective processes, I examined the acquisition and extinction of SCR's. Firstly, to check for acquisition, I performed a 2 learning type (direct vs indirect) x 2 face type (unsafe and safe face) x 2 time (pre and post-acquisition) mixed model ANOVA with face type and time as repeated measures on SCR data. A time main effect was detected, F (1, 59) = 4.53, p < .001, η_p^2= .43. Higher SCR's were found at post- acquisition (M = 1.29, SD = .03) than at pre-acquisition (M = 1.02, SD = .35), indicating a non-associative effect where both faces were perceived as more negative/anxiety provoking. All other effects were non-significant, p's > .462.

SCR results suggest both the unsafe and safe face were more negative/anxiety provoking following acquisition (i.e., a non-associative effect). I also expanded the model used to test *target* face negativity/anxiety to include *face variations*. In addition to shifting the negativity/anxiety of the faces directly involved in the conditioning, a non-associative generalisation of negativity/anxiety of the 25% and 50% variation faces were found. *New exemplar* SCR results suggest that new exemplars were perceived as being more negative/anxiety provoking irrespective of whether they were more similar to the unsafe or safe face. These analyses are reported in Appendix P2.

To ascertain whether changes in prototypicality were linked to changes in affective processes, I examined the acquisition and extinction of self-reported anxiety. I performed a 2 learning type (direct vs indirect) x 2 post-test position (after acquisition vs after extinction) x 2 face type (unsafe and safe face) x 2 time (pre and post-test) mixed model ANOVA with face type and time as repeated measures. One participant was excluded from this analysis as their post-test data were missing. A marginal 3-way interaction involving post-test position, face type and time was observed, F (1, 56) = 3.53, p = .065, η_p^2= .06. This effect was followed up by examining the two post-test positions sections of the design separately. I also expanded this analysis to include generalisation effects. A full description of the results are presented in Appendix P2, which demonstrate no generalisation of negativity/anxiety.

Firstly, to check for successful acquisition I looked at the data for participants in the after acquisition condition, who provided their self-reported ratings of face anxiety immediately *after* acquisition and *before* anxiety was extinguished. I found a marginal 2-way interaction involving face type and time, F (1, 29) = 3.99, p = .055, η_p^2 = .12: Participants reported more anxiety towards the unsafe face at post-test (M = 3.77, SD = 1.77) than pre-test (M = 2.87, SD = 1.53), t (29) = 3.73, p = .001; no differences were observed in the anxiety ratings towards the safe face at post-test (M = 3.53, SD = 1.59) and pre-test (M = 3.10, SD = 1.52), t (29) = 1.58, p = .125. A paired samples t-test was conducted to compare the pre- and post-test anxiety ratings for each target face. There was no difference in post-test values between the unsafe and safe, t (29) = 1.27, p = .214. Similarly, there was no difference in pre-test values between the unsafe and safe face, t (29) = -1.49, p = .147. Although there was no difference between the unsafe and safe face at post-test (and pre-test), this does not limit my interpretations. My research is interested in *changes* within each face, which is better highlighted by the significant interaction. The model was expanded to include face variations, however no significant results were found for the 25% and 50% variation faces. I also investigated the effects for each new exemplar separately. The first new exemplar was more anxiety provoking, irrespective of whether it was more similar to the unsafe or safe face or equally similar, at post-test (M = 3.96, SD = 1.62) than at pre-test (M = 3.00, SD = 1.74), F (1, 24) = 12.08, p = .002, η_p^2 = .34. No effects were found for the second new exemplar. These results suggest contingent specific acquisition took place as anxiety increases were observed for the unsafe (vs. safe) face before the extinction procedure was conducted.

After extinction participants, who provided self-reported anxiety following the extinction procedure, reported the unsafe and safe face to be equal in anxiety and not statistically different from pre-test levels: i.e., a non-significant 2 way interaction involving face type and time, F (1, 29) = .59, p = .448, η_p^2 = .02, and non-significant main effects of face type and time, both Fs < 1. As I found comparable anxiety ratings between the unsafe and safe faces among post-extinction participants that did not differ from pre-test ratings, these findings provide evidence that self-reported anxiety was extinguished.

I also checked for extinction with SCR data. I ran a 2 learning type (direct, vicarious) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. SCR data during the first block of extinction refers to the first two presentations of the unsafe and safe face and the ten trial block refers to the last two presentations of the unsafe and safe face up to the tenth trial (i.e., Presentation 7, 8, 9 and 10 of faces during extinction. I chose these two timepoints because all participants were exposed to a minimum of ten extinction trials. As expected there was no main effect of face type F (1, 57) = 2.99, p = .089, η_p^2 = .05 or face type x extinction block interaction F (1, 57) = .23, p = .635, η_p^2 = .00. I also supplemented this analysis in two ways. First, by investigating whether the last two presentations of the unsafe and safe face varied to post- acquisition data collected before extinction. Second, by determining whether the last two presentations of the unsafe and safe face increased significantly from 0. For readers with an interest in affective learning, supplemental extinction analyses are reported in Appendix P2, which confirm negativity/anxiety extinguished. Altogether, SCR data suggests higher negativity/anxiety developed during acquisition extinguished following extinction.

In summary, self-reported anxiety data provides evidence that the face paired with an electrical stimulation was more anxiety provoking following acquisition.

However, SCR data did not align with self-reported anxiety data and a non-associative effect was found with both the unsafe and safe face being perceived as more negative/anxiety provoking. The differing results could be due to measurement sensitivity and further research is needed. Both self-reported anxiety and SCR data provided evidence for successful extinction of the association developed during acquisition. Extinction results suggest participants' affective responding to the outgroup faces did *not* map onto their cognitive responding to these faces because where there was negativity/anxiety acquisition there was no prototypicality shift and where there was negativity/anxiety extinction there was a prototypicality shift

Testing Basic and Generalised Prototypicality Shifts

Latencies during the speeded sorting task were used to investigate prototypicality shifts towards the unsafe and safe exemplars. Incorrect categorisation responses were excluded from the latency data analysis and the mean reaction time of the latencies for the correctly categorised unsafe and safe exemplars were log-transformed to normalise the data (Richeson & Trawalter, 2005; Ruys et al., 2008). A response was judged incorrect if the face was sorted into the wrong category based on the target category creation (described in procedures and methods section). Responses quicker than 300 ms were excluded from the analyses and extremely long responses (> 3 SD) were rescored to the third standard deviation value for each group. On average, participants incorrectly sorted faces 7.04% of the time (SD = 3.91); a learning type x post-test position between subjects ANOVA confirmed there was no systematic difference in errors as a function of manipulations, all p's > .25. The average time taken to sort faces were: The unsafe exemplar at pre-test 558 ms (SD = 106 ms) and at post-test 525 ms (SD = 87 ms); the safe exemplar at pre-test 555 ms (SD = 98 ms) and post-test 528 ms (SD = 82 ms).

To test for *basic* prototypicality shifts among the *target* faces involved in conditioning, the sorting task data was analysed with a 2 learning type x 2 post-test position x 2 face type x 2 time mixed model ANOVA with face type and time as repeated measures. A significant interaction involving post-test position, face type and time was detected and followed up below, F (1, 57) = 10.00, p = .003, η_p^2 = .15.

Similar to Study 3.1, prototypicality shifts were detected in the after extinction condition. In this condition a significant two way interaction involving face type and time was detected, F (1, 29) = 12.00, p = .002, η_p^2 = .29, which is displayed in Figure 8. Participants were faster at sorting the unsafe face at post-test (M = 2.70, SD = .06) than pre-test (M = 2.74, SD = .07), t (29) = 3.83, p = .001, while comparable speeds were observed when sorting the safe face at post-test (M = 2.71, SD = .05) and pre-test (M = 2.72, SD = .06), t (29) = 1.01, p = .321. A paired samples *t*-test indicated there was no difference in post-test values between the unsafe and safe face, t (29) = -1.60, p = .121. There was a difference in pre-test values between the unsafe and safe face however, t (29) = 2.08, p = .047. Although there was no difference between the unsafe and safe face at post-test, but a difference at pre-test, it does not limit my interpretations. My research is interested in *changes* within each face, which is better highlighted by the significant interaction. Hence, consistent with Study 3.1 participants displayed a perceived prototypicality shift after extinction, but this time on an implicit measure. As the four way interaction involving learning, post-test position, face type and time was not significant, F (1, 57) = 1.26, p = .267, η_p^2 = .02, this result suggests comparable prototypicality shifts for those who learned first-hand and vicariously, replicating the findings from Study 3.1.

In the after acquisition condition no prototypicality shift was observed (see Figure 9): The face type by time interaction was non-significant, F (1, 30) = 2.38, p = .133, η_p^2 = .07. This pattern of dissociation along the post-test position factor indicates that prototypicality shifts are not present simultaneously with face-specific negativity/anxiety immediately after acquisition.

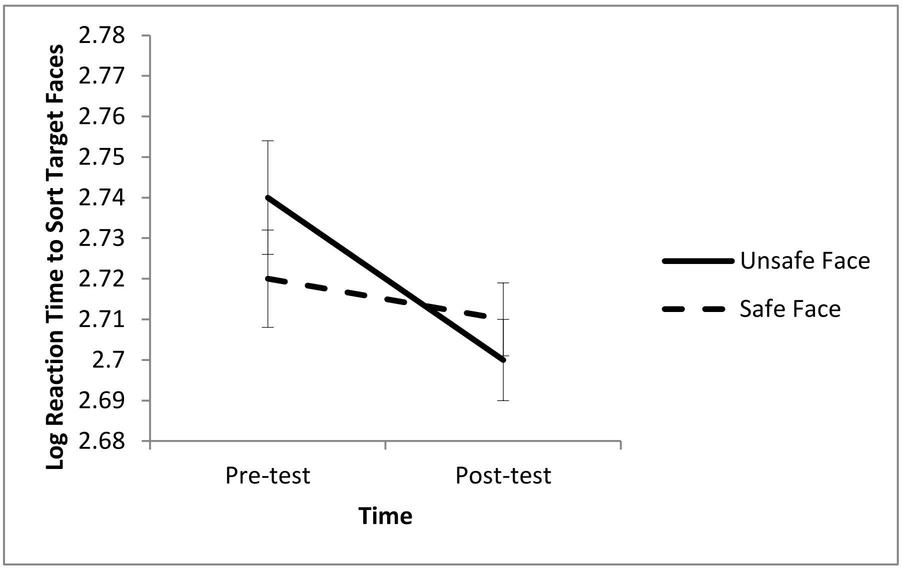


Figure 8. Face type x time interaction on log-transformed reaction times when sorting target faces after the extinction procedure

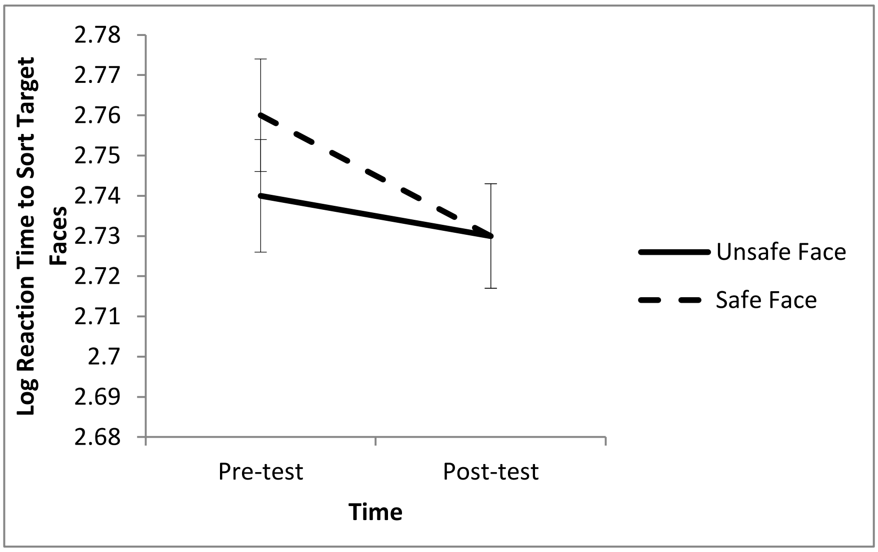


Figure 9. Face type x time interaction on log-transformed reaction times when sorting target faces after acquisition

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A 2 learning type by 2 post-test position by 2 face type by 3 generalisation gradient (target, 25%, 50%) by 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on sorting task data. Unexpectedly, no effects were found. All interactions involving face type x time x generalisation interaction were non-significant, *p'*s > .318; hence, no further analyses were carried out on the variation data.

To test whether discriminative prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face (new exemplar 1 N = 27, new exemplar 2 N = 28), safe face (new exemplar 1 N = 18, new exemplar 2 N = 27), or equally similar (new exemplar 1 N = 16, new exemplar 2 N = 6). The process used for computing similarity ratings was the same to that described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 learning type x 2 post-test position x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on the sorting task data. For the first new Black face, a main effect of time was found, F (1, 48) = 6.32, p = .015, η_p^2 = .12; Faces were rated quicker at post-test (M = 2.74, SD = .08) than at pre-test (M = 2.71, SD = .07), suggesting a non-associative form of generalisation across both faces and conditions. No effects were found for the second new exemplar, all *p*'s > .05.

Prototypicality shifts were also tested for using self-reported measures. To test for *basic* prototypicality shifts among the target faces involved in conditioning, a 2 learning type (direct, indirect) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with face type and time as the repeated measures was conducted on self-reported prototypicality data. One participant's data was excluded from this analysis due to a technical error. A time main effect was found, F (1, 56) = 4.88, p = .031, η_p^2 = .08. Participants rated both the unsafe and safe face as more prototypical of the outgroup at pre-test (M = 6.63, SD = .44) than at post-test (M = 6.51, SD = .51). Unexpectedly, all other effects were non-significant, all p's > .09. This result is inconsistent with the previous study and is explained in the General Discussion of this chapter as being likely due to a meta-contrast effect.

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A 2 learning type by 2 post-test position by 2 face type by 3 generalisation gradient (target, 25%, 50%) by 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted. A generalisation main effect was found, F (2,112) = 400.09, p < .001, η_p^2 = .88; the expected generalisation gradient was found with target faces rated highest in prototypicality (M = 6.57, SD = .44), followed by the 25% variation (M = 5.28, SD = .94) and the 50% variation (M = 4.37, SD = 1.05). Unexpectedly, all other effects were non-significant, all p's > .318, which suggest no generalisation effects to face variations; hence, no further analyses were carried out on the variation data.

To test whether discriminative prototypicality shifts generalised to *new exemplars*, participant's similarities ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face, safe face, or equally similar. The process used for computing similarity ratings described in Study 3.1 was used. Individual analyses were carried out for each new Black face, using a 2 learning type x 2 post-test position x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on self-reported prototypicality data. No effects were found for the two new exemplars, all *p*'s > .122.

In summary, discriminative prototypicality shifts were found when analysing sorting task data. In Study 3.2 I found that the unsafe face was perceived as being more prototypical of the Black outgroup at post-extinction compared to pre-test, similar to the time point that prototypicality shifts were detected at in Study 3.1. Similar to Study 3.1, Study 3.2 found prototypicality shifts post-extinction were comparable between participants who experienced the electrical stimulation directly or witnessed it vicariously. Prototypicality shifts were not found post- acquisition (prior to extinction), suggesting independent systems between cognition and affect may operate. This result also suggests the involvement of the extinction procedure in shifting exemplar prototypicality. In contrast to Study 3.1, there was limited evidence for generalised prototypicality shifts. Discriminative prototypicality shifts were only detected on the implicit measure, and non-associative shift was found on the self-reported prototypicality measure. This result is also in contrast with Study 3.1, but can be explained due to the order of measures which may have created a meta-contrast effect.

Exploring the Role of Mediators/Moderators in Basic Prototypicality Shifts

To analyse the effects contingency awareness[7] had, a 2 learning type (direct vs indirect) x 2 post-test position (after acquisition vs after extinction) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with face type and time as repeated measures was run on the sorting task. The face type x time x post-test position interaction held, F (1, 51) = 5.31, p = .025, η_p^2 =.09. The expected interaction involving face type x time x extinction x contingency awareness was non-significant, F (1, 51) =

1.03, p = .314, η_p^2 =.02. Despite the detection of a non-significant 4 way interaction, planned comparisons were carried out[8].

I investigated the effects contingency awareness had by analysing data separately for contingent aware and unaware participants in a 2 learning type x 2 post-test position x 2 face type x 2 time mixed model ANOVA with face type and time as the repeated measures with sorting task data. No significant effects were found in contingent unaware participants and no further analyses were carried out, all p's > .118. In contrast, participants who were contingent aware demonstrated a time main effect, F (1, 35) = 8.94, p = .005, η_p^2 =.20. More importantly, a significant 3 way interaction involving face type, time and post-test position was found and subsequently followed up by looking at each level of post-test position separately, F (1,35) = 8.36, p = .007, η_p^2 = .19. In the after extinction condition, contingent aware participants (n = 22; 12 direct and 10 indirect participants) sorted the unsafe face faster at post-test than pre-test compared to the safe face and the basic prototypicality shift was reproduced, F (1, 20) = 19.80, p < .000, η_p^2 =.497. No effects were found in the after acquisition group. However, power issues limit the interpretations that can be extracted from this analysis.

The analyses carried out on the sorting task data were replicated for the self-reported prototypicality data. To analyse the effects contingency awareness had a 2 learning type (direct vs indirect) x 2 post-test position (after acquisition vs after extinction) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with face type and time as repeated measures was run on self-reported prototypicality data. The time effect previously observed for basic prototypicality shifts held in the expanded model, F (1, 52) = 4.06, p = .049, η_p^2 =.07. The expected interaction involving contingency awareness, however, failed to detect any significant differences. Despite this I performed planned contrasts as outlined previously with the sorting task data analysis of basic prototypicality shifts involving contingency awareness.

In contingent aware participants, the time main effect observed for basic prototypicality shifts was again found, F (1,34) = 4.44, p = .042, η_p^2 =.12. All other effects were non-significant, p > .124. No effects were detected in contingent unaware participants.

To explore the role that anxiety in general had for prototypicality shifts, I performed a mediational analysis with absolute levels of anxiety in after extinction participants because this is where changes in prototypicality were detected. Absolute levels of anxiety were calculated as the mean post-test skin conductance arousal level in response to the unsafe and safe faces collected immediately after acquisition had occurred. A mediational analysis was performed by entering the average skin conductance level (SCR) as a covariate into a face type x time x learning type ANCOVA (Baron & Kenny, 1986; Judd et al., 2001; Yzerbyt et al., 2004). The average SCR anxiety score nullified the face type x time interaction for after extinction participants, from F (1, 28) = 12.08, p= .002, η_p^2 =.301 to F (1, 27) = .001, p = .973, η_p^2 = .000. This finding suggests anxiety learning does not need to be face-specific. Rather, heightened anxiety *in general* is a factor in the prototypicality shift.

The number of trials it took for after extinction participants to extinguish their psychophysiological reported anxiety response towards the unsafe face varied across participants (M = 12.69, SD = 3.36), but showed no relationship with overall (average) anxiety after acquisition (r = .07, p = .709). This means that participants varied in the number of repeated presentations of the target faces they were exposed to during the extinction procedure. We capitalised on this inter-individual variability to examine whether changes in face prototypicality were mediated by simple repeated exposure to the target faces. A mediational analysis was conducted with the number of trials it took for after extinction participants to extinguish entered as a covariate into a face type x time x learning type ANCOVA. This factor also nullified the face type x time interaction from F (1, 28) = 12.08, p= .002, η_p^2 =.301 to F (1, 27) = .477, p = .496, η_p^2 = .02. This finding indicates that the number of repeated non reinforced presentations of the faces during the extinction procedure is a factor in the prototypicality shifts.

In summary, the magnitude of prototypicality shifts was larger amongst contingent aware participants in the after extinction condition on the implicit measure of prototypicality and not on the self-reported measure – possibly due to a meta-contrast effect. As prototypicality shifts were driven by contingent aware participants, this suggests that awareness of the pairing between outgroup exemplar and shock contributes to the effect – presumably by drawing attention to face differences. Absolute levels of negativity/anxiety collected post- acquisition suggested developed associations do not need to be contingent specific. Rather, heightened negativity/anxiety *in general* is a factor in the prototypicality shift. Number of repeated presentations throughout extinction was also demonstrated to be involved in prototypicality shifts, suggesting cognitions play an important role in shifting exemplar prototypicality. Taken together, the results suggest that heightened anxiety contributes, but not uniquely, to the prototypicality shift.

Study 3.3

Study 3.2 provided evidence that exemplar prototypicality shifts were stronger among participants aware of the face-stimulation pairing. Study 3.3 sought to further disentangle the unique role contingency awareness has in prototypicality shifts. I used the same design from Study 3.2 in Study 3.3 – Direct and vicarious conditioning towards an unsafe and safe Black exemplar occurred and prototypicality was measured at pre-test and at post-test either after acquisition or after extinction. However, I made one change to the design of Study 3.2 to investigate the effects contingency awareness has. I implemented a backwards masking procedure during conditioning – a procedure that prevents explicit *awareness* of the stimuli's contingencies even though the contingency *is* still *presented* to participants. During conditioning, the unsafe and safe face are presented briefly (typically < 100ms), and replaced with a second stimulus (a mask) that disrupts the awareness of the original stimulus by preventing further visual processing beyond the target stimulus duration. The mask comprised of parts of the unsafe and safe target faces, which were split into smaller pieces and reassembled in a scrambled formation in the shape of a face. As a consequence, the faces are no longer consciously perceived during conditioning, even though the contingency between the unsafe face and negative/anxiety provoking stimulus is left intact (Breitmeyer, 2007; Enns & Di Lollo, 2000). Hence, Study 3.3 extended Study 3.2 by preventing contingency awareness from developing, rather than relying on self-reported data to evaluate the effect of this psychological underpinning.

Another addition to Study 3.3, relative to Studies 3.1 and 3.2, was the inclusion of group-level measures in the form of feeling thermometers in the self-reported questionnaire. Feeling thermometers measure explicit evaluations towards the Black group by indicating on a scale of 0-100 how warmly or coldly they felt towards the Black group (Haddock, Zanna & Esses, 1993; Hugenberg & Bodenhausen, 2004). This is a further manipulation check in addition for checking acquisition and extinction. I included a group level measure of evaluations in this study because the mask retained the key ethnicity indicator, the Black skin. If participants associated the Black skin of the mask with the negative/anxiety provoking experience, I am likely to find non-associative conditioning. If non-associative conditioning occurs and both the unsafe and safe face are perceived more negative/anxiety provoking, I could find non-associative prototypicality shifts that generalise to any face that has Black skin.

Method

Participants and Design

Participants were 64 White-Anglo Saxon students (16 male, 45 male; *M* = 20.63, *SD* = 3.15) from a large regional Australian university. Five participants were excluded from the analysis as they indicated some awareness of the faces before the mask[9], leaving 59 valid participants. Similar to Study 3.2, participants were randomly assigned to one of four conditions in a 2 learning type (direct *n* = 28, vicarious *n* = 31) x 2 post-test position (after acquisition *n* =29, after extinction *n* = 30), but all participants underwent a masking procedure.

Procedure

The procedure and materials from Study 3.2 were used for the current study with some notable exceptions. First, in the current study, a masking procedure was implemented during acquisition. During acquisition the target face was presented for a period between 17 and 27 milliseconds and then replaced with a mask to limit the awareness of the face (Breitmeyer, 2007; Enns & Di Lollo, 2000)[10]; previous research has shown masking to be effective for stimulus presentation times ranging from 12 to 106 milliseconds (Bacon-Mace, Mace, Fabre-Thorpe & Thorpe, 2005) and both learning conditions had the target faces presented within this timeframe (Monitor refresh rate: 85 Hz). The mask was made from parts of the unsafe and safe face scrambled randomly; the outline of the face was retained in ways to limit recognition of the target faces (see Appendix Q for mask used; the same mask was used for the unsafe and safe face). The target faces in both conditions were immediately replaced with the same mask for the remaining 10 seconds.

The online questionnaire remained the same as for Study 3.2, but a feeling thermometer was included to assess group based evaluations of the Black outgroup (Haddock et al., 1993)—i.e., generalised affective learning; it required participants to rate their overall feeling towards the outgroup between 0 (*very cold*) and 100 (*very warm*) degrees, with the scale increasing in 10 degree increments. Self-reported prototypicality and the speeded sorting task used in Study 3.2 were used again to measure prototypicality shifts.

Results and Discussion

Overview of Analyses

Skin conductance responses (SCRs; collected pre-/post-acquisition and post-extinction) and self-reported anxiety (collected at pre-test and either post-acquisition or post-extinction) were first analysed and treated as manipulation checks to ascertain that negativity/anxiety was acquired and extinguished as a function of acquisition and extinction respectively. For readers with an interest in affective learning, ancillary results for changes in pre-acquisition and post-acquisition generalisation data on SCRs and self-reported anxiety are reported in the Appendix P3. In addition, feeling thermometer data was measured pre-acquisition and either at post-acquisition or post extinction and used to determine changes in group-level affect.

Main analyses for my prototypicality measures (collected at pre-test and either post-acquisition or post-extinction), were assessed via self-reports and a speeded sorting task, followed. These analyses were the same described for Chapter 2, with the only difference in methodology being the masking procedure during acquisition.

Conditioning and the Influence of Negative Affect/Anxiety

Effective acquisition was checked for using SCR that was scored and calculated as described in Study 3.1. I performed a 2 learning type (direct, indirect) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with face type and time as repeated measures. A time main effect was detected that demonstrated higher SCR activity at post-test (*M* = 1.26, *SD* = .31) than at pre-test (*M* = 1.01, *SD* = .04), F (1,56) = 35.56, $p < .001$, $\eta_p^2 = .39$. Unexpectedly, all other effects were non-significant, all *p*'s > .163. This is a non-associative effect whereby both the unsafe and safe face were perceived as being more negative/anxiety provoking at post-test than at pre-test and aligns with prototypicality shifts of both the unsafe and safe face.

SCR results suggest both the unsafe and safe face were more negative/anxiety provoking following acquisition (i.e., a non-associative effect). I also expanded the model used to test *target* face negativity/anxiety to include *face variations*. In addition to shifting the negativity/anxiety of the faces directly involved in the acquisition, non-associative generalisation of negativity/anxiety of the 25% and 50% variation faces were found. *New exemplar* SCR results suggest that new exemplars were perceived as being more negative/anxiety provoking irrespective of whether they were more similar to the unsafe or safe face. For readers with an interest in affective learning, ancillary results for pre-/post-acquisition changes in generalisation data on SCRs are reported in the Appendix P3

I also checked for effective acquisition and extinction using self-reported anxiety. Self-reported anxiety data was analysed using a 2 learning type x 2 post-test position x 2 face type x 2 time mixed model ANOVA with face type and time as repeated measures. I expected post-test position to be involved in all interactions because anxiety should be higher before extinction if conditioning was effective. In contrast to Study 3.2, I found no evidence of a significant three way interaction involving post-test position, face type and time, F (1, 55) = .54, p = .467, η_p^2 = .010, with all other results also non-significant, p's > .138. I expanded the model to include *face variations*, however no significant results were found, all p's > .112. Similarly, no effects were found when checking for anxiety generalisation to new exemplar, all p's > .05. Hence, masking the contingency between the unsafe face and the mildly aversive outcome prevented anxiety being associated with target, variation and new faces when measured through self-report.

I also checked for extinction with SCR data. I ran a 2 learning type (direct, vicarious) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. SCR data during the first block of extinction (first two presentations of the unsafe and safe face) and the ten trial block (the last two presentations of the unsafe and safe face up to the tenth trial) was chosen as all participants were exposed to a minimum of ten extinction trials. As expected there was no main effect of face type, F (1, 54) = .12, p = .731, η_p^2 = .00 or face type x extinction block interaction F (1, 54) = .25, p = .623, η_p^2 = .00. These two results suggest there is no difference between the unsafe and safe face throughout extinction. I also supplemented this analysis in two ways. First, by investigating whether the last two presentations of the unsafe and safe face varied from post-extinction to post-acquisition data collected before extinction. Second, by determining whether the last two presentations of the unsafe and safe face increased significantly from 1. Supplemental data analyses conducted on SCR suggests negativity/anxiety extinguished and are discussed extensively in Appendix P3.

Altogether, SCR data suggests higher negativity/anxiety developed during acquisition towards the unsafe and safe faces, the effects were extinguished following extinction.

In addition to checking for effective acquisition and extinction with SCRs and self-reported anxiety measures, I assessed group-level changes (as opposed to face specific) with the feeling thermometer data. If negative associations generalised to all Black exemplars, I expected participants in the after acquisition condition to report less positive affect towards Black people compared to participants in the after extinction condition. Effective extinction of generalised anxiety to the entire group would be demonstrated by no difference in feelings at the two time periods in the after extinction condition. A 2 learning type x 2 post-test position x 2 time mixed model ANOVA was conducted on the overall feeling towards Black people, with time being the repeated measures. As expected, a marginally significant post-test position x time interaction was found on feeling thermometer data, $F = (1, 55) = 3.28$, $p = .076$, $\eta_p^2 = .056$. Feeling thermometer data collected before extinction showed a tendency for participants to report less positive evaluations towards Black people in general at post-test ($M = 68.07$, $SD = 16.34$) compared to pre-test ($M = 71.19$, $SD = 17.60$); t (28) = -1.67, $p = .107$. Participants in the after extinction condition instead showed comparable evaluations reactions towards Black people post-extinction (M =74.82, $SD = 13.07$) and pre-test (M =73.39, $SD = 13.00$), t (29) = .81, $p = .423$. Together these results provide some evidence that the aversive stimulus was causing participants to generalise negativity/anxiety to Black people and the extinction procedure was successful at removing the generalised negativity/anxiety.

Testing Basic and Generalised Prototypicality Shifts

To investigate basic prototypicality shifts, correct response latencies during the sorting task were calculated as in Study 3.2 for the target faces and analysed using a 2 learning type x 2 post-test position x 2 face type x 2 time mixed model ANOVA with face type and time as the repeated measures[11]. The mean error rate was 6.83 (*SD* = 4.18) and experimental manipulations again had no effects on errors, all *p'*s > .38. The average time it took participants to sort faces were: The unsafe exemplar at pre-test 562 ms (*SD* = 103 ms) and at post-test 525 ms (*SD* = 81 ms); the safe exemplar at pre-test 538 ms (*SD* = 90 ms) and post-test 513 ms (*SD* = 84 ms). Contrary to Studies 3.1 and 3.2, I found no significant interaction between face type x time, F (1, 55) = .60, p = .443, η_p^2 = .01, or face type x time x post-test position, F (1, 55) = .10, p = .756, η_p^2 = .00. Whilst there was no contingent specific shift in prototypicality, I did find a non-discriminative prototypicality shift towards both faces, as demonstrated by a time main effect, F (1, 55) = 13.65, p = .001, η_p^2 = .20. Together, both faces were sorted quicker at post-test (*M* = 2.71, *SD* = .07) than at pre-test (*M* = 2.73, *SD* = .06). Hence, masking the contingency between the face type and the outcome (shock vs no shock) eliminated any evidence of a contingent specific prototypicality shift and, instead, caused a non-associative prototypicality shift towards both faces. This finding provides an experimental demonstration that contingency awareness is necessary to increase the outgroup-like status of the unsafe outgroup face.

Despite there being no contingent specific shifts in prototypicality, I extended the previous analysis and included face variations in a 2 learning type x 2 post-test position x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation gradient and time as repeated measures. The time main effect previously found held, and more importantly, I found a generalisation gradient x time interaction that I followed up by looking at each face variation separately, F (2, 110) = 3.60, p = .032, η_p^2 = .06 A non-associative prototypicality shift in the same direction as target faces was found for 25% variation faces as evidenced by a time main effect, F (1, 55) = 26.76, $p < .001$, $\eta_p^2 = .33$ (pre-test $M = 2.79$, $SD = .07$; post-test $M = 2.75$, $SD = .07$). A similar time main effect was found for 50% variation faces, F (1, 55) = 18.39, $p < .001$, $\eta_p^2 = .25$ (pre-test $M = 2.86$, $SD = .10$; post-test $M = 2.81$, $SD = .08$). Consistent with the non-associative/non-discriminative *basic* prototypicality shifts, both the unsafe and safe face *variations* shifted in prototypicality in the expected direction along the expected gradient.

To test whether the non-associative prototypicality shifts extended to *new exemplars*, participant's similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face (new exemplar 1 $N = 21$, new exemplar 2 $N = 24$), safe face (new exemplar 1 $N = 17$, new exemplar 2 $N = 30$), or equally similar (new exemplar 1 $N = 22$, new exemplar 2 $N = 6$). I conducted a 2 learning type x 2 post-test position x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure carried out individually for each exemplar using sorting task data. The time main effect observed for variation faces held for the first new exemplar, but was further qualified by post-test position and new face similarity. I investigated this effect further by looking at each level of post-test position and new face similarity separately. All faces were perceived as more prototypical of the outgroup, but this effect was more pronounced in the before extinction, direct learning group. No effects were found for the second new exemplar. Together, these results suggest that prototypicality shifts generalised to one new exemplar, but not another.

I also tested for *basic, face variation and new exemplar* prototypicality shifts using self-reported prototypicality data. I first conducted a 2 learning type (direct, indirect) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with face type and time as the repeated measures on self-reported prototypicality data. No significant effects were observed, all p's > .235. I expanded the model to include *face variations* and detected no effects involving face type or time. Similarly, new exemplar prototypicality shifts were not detected. Together, self-reported prototypicality results are not consistent with prototypicality shifts detected on the sorting task data and may be due to a meta-contrast effect.

In summary, contingent specific shifts of exemplar prototypicality were not found. Although discriminative shifts of prototypicality were not found, I did find non-associative/non-discriminative shifts of prototypicality on the implicit measure that shows both the unsafe and safe face were perceived as more prototypical of the Black outgroup following conditioning – In both learning and post-test position conditions. A non-associative prototypicality shift generalised to *face variations* and *new exemplars*. No effects were found on the self-reported prototypicality measure, but this may be due to the position of measurements that resulted in a meta-contrast effect.

Overall, the results in Study 3.3 revealed acquisition was not contingent specific with SCR data, and rather generalised to all exemplars that had Black skin. Self-reported and SCR data both suggest negativity/anxiety were comparable at the conclusion of extinction. This was further supported through feeling thermometer data, which demonstrated lower feelings of warmth towards the Black group after-acquisition compared to after-extinction. I argue that the Black skin used in the mask became associated with the negative/anxiety provoking stimulus and resulted in non-associative conditioning. This can explain why a non-associative prototypicality shift occurred, as it was the Black skin rather than the unsafe face that was perceived as more negative/anxiety provoking. However, this effect could also be explained by task repetition. Research suggests that task repetition results in faster response times and this is an alternate explanation that cannot be ruled out (Pashler & Baylis, 1991).

General Discussion

Extant literature indicates that pairing negativity with outgroups has a *direct* detrimental impact on intergroup relations because it worsens affective responding to outgroup exemplars (Mallan et al., 2009; Olson & Fazio, 2006; Olsson et al., 2005; Walther, 2002). The present research makes a significant contribution to this literature by identifying novel *indirect* detrimental effects through changes in the social categorisation of outgroup exemplars. In particular, my research demonstrates that first hand and vicarious pairing of incidental negativity/anxiety with outgroup exemplars causes exemplars to be perceived as more prototypical of the outgroup, or more outgroup-like. This effect is consistent with evaluative- and emotion-fit mechanisms as the aversive stimulus paired with the outgroup fits pre-existing outgroup expectations of being negative/anxiety provoking. There is a plethora of research indicating the powerful implications of high levels of prototypicality in stereotyping, discriminatory responses, and intergroup bias (Blair et al., 2004; Blair et al., 2002; Brown & Hewstone, 2005; Ma & Correll, 2011; Rothbart & John, 1985). Thus, there is a robust evidence and theory-driven basis to expect prototypicality shifts, as I isolated them here, to shape important downstream behaviours and cognitions towards outgroup exemplars and the outgroup as a whole.

Summary and Implications of Key Findings

This set of studies first establishes the existence of exemplar prototypicality shifts. Participants in Study 3.1 rated the unsafe outgroup face as more outgroup-like after aversive conditioning; this effect was found after the extinction procedure. Study 3.2 provided further evidence for the prototypicality shift effect using an implicit measure. Importantly, Study 3.2 found that the prototypicality shift effect was present after the extinction procedure. Hence, the timepoint prototypicality shifts were found at in Study 3.1 was replicated in Study 3.2. However, it further revealed that the effect was not present before the extinction procedure. Therefore, these results suggest that the additional associations learnt and the additional cognitive processing implicated in extinction (Bouton, 2002; Hermans et al., 2006) play an important role in shifting exemplar prototypicality. Although prototypicality shifts only occurred post-extinction, the exact cause extinction has in this effect is not yet known. Together, results suggest that negativity/anxiety not only serve as a powerful marker of outgroup membership (Miller et al., 2010; Richeson & Trawalter, 2005), but actively shapes the inclusion-exclusion of individual outgroup exemplars from the outgroup—after the aversive associations have become more complex post repeated exposure to the target faces.

The establishment of exemplar prototypicality shift effect is consistent with evaluative-fit and emotion-fit mechanisms. My argument for the mechanisms underlying exemplar prototypicality shifts focused around the incidental negativity and anxiety embedded in my conditioning procedure. Negative valence is typically a marker of outgroup membership because of the psychological motivation to perceive the ingroup as being positive in order to maintain self-esteem as being part of a positively valenced group (Tajfel & Turner, 1979, Turner et al., 1987). This psychological alignment may have contributed to the observed prototypicality shifts because negativity is experienced as being more normative for the Black outgroup. In addition, recent work on emotion applicability and intergroup bias by Dasgupta, DeSteno, and colleagues (Dasgupta et al., 2009; DeSteno, Dasgupta, Bartlett & Cajdric, 2004) suggests anxiety is particularly applicable to threatening outgroups like my target outgroup (i.e., Black people are stereotyped as aggressive), and is less applicable to non-threat based outgroups (e.g., the elderly; Cottrell & Neuberg, 2005). This psychological alignment may have contributed to the observed prototypicality shifts because anxiety is experienced as being more normative for the Black outgroup. Hence, the observed shifts in prototypicality may be due to exemplar negativity, but also to the applicability of the specific negativity (e.g., anxiety) to the outgroup.

My tests of generalisation mapped onto traditionally distinct operationalisations of generalisation in learning and social psychology (Lissek et al., 2010; Pettigrew & Tropp, 2006; Ranganath & Nosek, 2008; Verosky & Todorov, 2013). In Study 3.1, the evidence indicated generalised prototypicality shifts towards the 25% variation face, but not towards the 50% variation. While outgroup exemplars immediately or closely associated with incidental negativity or anxiety became more outgroup-like, inherently safe exemplars – which had never been associated with the aversive experience -- shifted away from the category prototype (i.e., a contrast effect). Study 3.2 provided evidence for a non-associative prototypicality shift towards a new exemplar, but no other generalisation effects were found. Study 3.3 provided evidence for a non-associative prototypicality shift and all outgroup exemplars were perceived as being more outgroup-like post-extinction when the aversive association could not be pinpointed to a specific outgroup exemplar due to the masking procedure. I suspect the aversive experience became associated with the ethnicity cue (i.e., Black skin), which caused all Black exemplars to be perceived as more outgroup-like. Together, Study 3.1 and 3.3 provided evidence that prototypicality shifts generalise to exemplars not involved in conditioning.

It could be argued that downstream processes as a result of generalised prototypicality shifts were investigated in Study 3.3 through group-level feeling thermometer data, which I used as a manipulation check. Feeling thermometer data suggests that as a result of conditioning, evaluations of the outgroup were worsened following acquisition and brought back to a neutral state following extinction. I used feeling thermometer data as a form of manipulation check to determine whether a non-associative form of conditioning occurred and negativity/anxiety generalised to the entire group (as a result of the aversive pairing with Black skin of the mask). I incorporated this measure because it has been used as a sensitive measure of societal prejudice and evaluations (Hugenberg & Bodenhausen, 2004; Wittenbrink, Judd & Park, 2001). However, to accurately measure the downstream consequences of a generalised prototypicality shift, a stereotyping measure should be included rather than a measure of affective change. To measure changes in stereotyping, participants could be provided with a verbatim that is either stereotypically consistent or inconsistent with the Black group and asked to provide ratings about how each description fits exemplars at pre- and post-test. A similar measure to that proposed will provide a better test of the effects generalised prototypicality shifts have for stereotyping, rather than the affective measure used in Study 3.3.

Notwithstanding limited power, evidence was provided that suggests these categorisation effects do not require negativity-outgroup pairings to be experienced first-hand; our effect of interest was observed equally when aversive experiences were merely witnessed in similar others. This effect is consistent with previous research that demonstrates how vicarious contact can be used to change intergroup attitudes. For example, Mazziotta, Mummendey and Wright (2011) found video-based vicarious contact improved attitudes towards the outgroup. Hence, our results reaffirm how powerful other-mediated experiences (Bandura, 1977; Miller & Dollard, 1941; Mazziotta et al., 2011; Olsson & Phelps, 2004), increasingly central in technologically advanced societies, can be in explaining the development and the wide-spread communication of negative stereotypes in society (Weisbuch et al., 2009).

A secondary issue that this research began to address is the complex and debated affect-cognition interplay (Amodio & Devine, 2008; Cunningham et al., 2004). Exemplar prototypicality shifts were detected only after any physiological and self-reported anxiety had extinguished, suggesting that cognitive and affective responding are at least partly independent processes. In particular, the results suggest that categorisation changes following aversive conditioning do not need affective correlates to be long-lasting or co-occurring. Rather, once established through a discrete aversive experience, they can endure well after negative affect and evaluations towards the unsafe face have subsidised or new associations have made the predictive relationship less clear. Therefore, my results point towards some relative independence between changes in categorisation and changes in affect associated with outgroup targets.

The interplay between cognition and affect was investigated further via the level of cognitive processing involved in the controlled-automatic nature of prototypicality shifts. I demonstrated that prototypicality shifts could be detected using both explicit (self-reports) and implicit (speeded computerised sorting task) measures of intergroup categorisation. The detection of prototypicality shifts on a self-reported measure provided evidence that cognitive processes were involved in the prototypicality shift. However, prototypicality shifts were also detected on an implicit measure, suggesting changes in prototypicality were detected at levels with limited cognitive processing. This invariance of effects across measures suggests that participants' awareness of evaluating exemplar prototypicality was not critical to detect our key effect.

Possible boundary conditions to detecting exemplar prototypicality shifts on different types of measures were found in Study 3.2 and 3.3. In these two studies exemplar prototypicality shifts were not found on self-reported measures and ceiling effects were detected instead. I believe ceiling effects were found because I introduced the sorting task, which included an intergroup (Black and White) context (i.e., a meta-contrast ratio; Oakes et al., 1994; Oakes, Turner & Haslam, 1991; Turner et al., 1987) otherwise not available to Study 3.1 participants. The ceiling effect is highlighted by more pronounced self-reported ratings in Study 3.2 (pre- and post-test ratings means > 6 out of a possible 7) than in Study 3.1 (pre-test ratings means <4.5, post-test ratings <6). White and Black exemplar distinctions were more pronounced by the inclusion of an intergroup context, which lead to a meta-contrast effect and ceiling effects.

Given the caveat of limited power, contingency awareness was found to play a key role in my prototypicality shifts, which would suggest that cognitive processing plays an important role in my effect. In study 3.2 I tested the effects contingency awareness had through a series of funnelled questions that became more targeted in determining which exemplar was paired with the aversive stimulus (Clark & Squire, 1998; Page, 1973; Wardle, Mitchell & Lovibond, 2007). Although self-reported questionnaires have been used in previous research, my aposteriori analysis resulted in small sample sizes that should be followed up with well powered studies that could incorporate better tests of contingency awareness that do not involve recall i.e., dial, pointer setups (Lipp, 2006; Purkiss & Lipp, 2001). Results from Study 3.2 suggested exemplar prototypicality shifts were present in contingent aware participants, but not contingent unaware participants. Study 3.3 manipulated contingency awareness through a masking procedure that prevented contingency awareness from developing. Study 3.3 failed to find discriminative prototypicality shifts (but found non-associative/non-discriminative shifts) in the absence of contingency awareness. Together, these results provide preliminary evidence that suggest awareness between the aversive stimulus and outgroup exemplar pairing is needed in order for exemplar prototypicality shifts. Thus, the effect is consistent with previous research that indicates contingency awareness plays a vital role in conditioning (Hofmann et al., 2010; Pleyers et al., 2007).

Other mediators/moderators were also shown to be influential in exemplar prototypicality shifts. I showed that repeated presentations throughout extinction and general anxiety mediated the prototypicality shift effect found in Study 3.2. It is possible that repeated presentations of the unsafe exemplar in the presence of some anxiety helps to consolidate the shift in prototypicality by focusing attention to the unsafe exemplar. Hence, there is some indication that repeated presentations of the unsafe exemplar with no aversive stimulation, but in the presence of some anxiety (nonspecific to the unsafe face), underpin the prototypicality shift. I investigate these ideas further in the research reported in the next chapter.

Limitations and Future Research Ideas

Differential conditioning was not always found across my studies and non-associative processes could be used to explain results. In Study 3.2 differential conditioning was not detected in SCR but was on self-reported anxiety. In Study 3.3 contingency awareness was removed during conditioning and prevented differential conditioning from occurring. Whilst contingent specific/discriminative prototypicality shifts were detected in Study 3.2 on the sorting task, non-associative/non-discriminative prototypicality shifts were detected on the self-reported measure (see below for meta-contrast explanation). Study 3.3 detected non-associative prototypicality shifts on both measures. Where non-discriminative conditioning and prototypicality shifts occurred, sensitisation and habituation could explain results. During acquisition, sensitisation could have occurred and participants responded to both faces with heightened negativity/anxiety (Cevik, 2014). During extinction, habituation could have occurred and negativity/anxiety responses diminished following repeated presentations. Thus, a weakness of these studies is effects could be due to non-associative processes, rather than associative processes.

The exploratory nature around some factors in my data analysis can be interpreted as a limitation. Factors that were of an exploratory nature did not inform decisions around required sample sizes, which led to power issues and possible false-negative results (Cohen, 1992). This shortcoming is particularly relevant to the confidence in the null finding for learning type across studies and the null effect in the top level ANOVA for contingency awareness analysis in Study 3.2. Future research should provide a more stringent test of my exploratory analyses by taking these factors into account when planning the study design and required sample sizes. Thus, exploratory analyses were undertaken to provide a first look at mechanisms involved in prototypicality shifts and should be followed up with further, well powered studies.

Another limiting factor in my interpretation of the results is lack of clarity over the mechanisms implicated in exemplar prototypicality shifts. I interpret the lack of temporal co-variation between affect and cognition (i.e., no evidence prototypicality shift following conditioning but prior to extinction) as meaning no effect of affect on cognition. However, it is possible that negativity/anxiety towards the unsafe face re-emerges following extinction and is present at the time prototypicality is measured. For example, negativity/anxiety could have been renewed or spontaneously recovered post-extinction (Bouton, 1994; 2002; 2004; 2014). Conditioning (acquisition) and extinction occurred in the same context because physiological equipment was attached and participants were only required to watch the screen. When measuring prototypicality, physiological equipment was removed and participants were required to make decisions. It is possible that participants perceived the collection of prototypicality measures as being in a different context to what conditioning and extinction occurred in.

This raises the possibility that AAB renewal occurred and negativity/anxiety were present were prototypicality measures were collected (Bouton, 2002; Bouton & Ricker, 1994). Similarly, negativity/anxiety could have spontaneously recovered following the passage of time since extinction (Bouton, 2002; Rescorla, 2004).

Whilst my prototypicality and anxiety measures were collected in close proximity, a simultaneous method of measuring prototypicality and anxiety is required to address the possibility that negativity/anxiety was renewed or spontaneously recovered. Methodologically, a simultaneous method is difficult because drawing attention to both processes at the same time point is likely to influence results. For example, in Study 3.2 the introduction of white faces in the stimulus set for the implicit measure of categorisation might have inadvertently caused a meta-contrast effect. The meta-contrast effect caused an intergroup context that resulted in ceiling effects being detected on the self-reported measure. Hence, these processes are malleable and easily influenced by probing questions. As prototypicality is malleable future research could ask participants to rate prototypicality first, after which anxiety is immediately measured. Furthermore, SCR is a good measure of affect that is non-invasive and future research could ask participants to undergo post-acquisition test phase again following the collection of prototypicality measures. Rescorla (2004) used a similar procedure to determine if extinguished effects spontaneously recovered and this could be used in a similar manner. This might help to provide clarity around the factors implicated in prototypicality shifts, which are studied more extensively in the next chapter.

While I believe the present research advances our understanding of the social psychological and associative bases of stereotype formation, several intriguing research questions await further research. The current studies cannot determine conclusively if prototypicality is shifting at the exemplar-level representation or at the group-level representation, or both. For example, it is possible that conditioning affected the overall group representation, making it shift towards the unsafe exemplar (rather than the other way round as I have argued so far). There are several exemplar-based models of category representation that align with this idea (e.g., Smith & Zarate, 1992). These models theorise that group representations and prototypes are formed through interactions with and exposure to members from that group. As interactions with members from the group increase, perceptions and thoughts about the group also develop and change. In this process, it is possible that the unsafe exemplars from my aversive conditioning procedure are weighted more heavily in constructing the overall representation of the group, leading to the observed prototypicality shifts. Alternatively, the prototypicality of the unsafe exemplar may have shifted towards a more stable, enduring group prototype. The method in the current studies does not allow me to distinguish between these two scenarios, but future research should try to discriminate between the two.

A desirable further extension will be to test these effects in the context of ingroup members. In the present line of research, I have shown that an outgroup exemplar paired with an aversive stimulus becomes more outgroup-like. To the extent to which ingroups are treated as intrinsically positive/safe (Cottrell & Neuberg, 2005; Turner et al., 1987), I expect the opposite effect to occur towards ingroup members; that is, I expect ingroup members paired with incidental negativity/anxiety to become *less*, rather than more, prototypical of the ingroup (i.e., less ingroup-like). This idea is investigated further in Chapter 5.

Future research could also try to distinguish between evaluative-fit and emotion-fit mechanisms. One way to address this when investigating prototypicality shifts of Black exemplars is to manipulate valence without threat-related emotions. Instead of pairing a Black exemplar with an electrical stimulation that co-inflates negativity and anxiety, a pairing with another stimulus that is negative, but not anxiety provoking (i.e., money loss), would help distinguish between the two mechanisms. Another method to distinguish between the two mechanisms is to pair an exemplar with an emotion that is not applicable to the type of threat that group is associated with. For example, disgust is not associated with the Black outgroup (Cottrell & Neuberg, 2005). If exemplar prototypicality shifts were detected when a Black exemplar was associated with a disgusting emotion, it would rule out an emotion-fit explanation. I attempted to distinguish between the two mechanisms in Chapter 6.

Conclusion

I believe that the prototypicality shifts identified in this research are central to the development and maintenance of negative intergroup relations in society. I demonstrate that prototypicality shifts occur when negativity (or anxiety more specifically) is merely incidental (vs. integral; Bodenhausen, 1993) to the outgroup exemplars. This means that outgroup exemplars do not need to be disliked, angry, or threatening to be treated as representative of outgroups; rather, they increase their outgroup-like qualities by pure virtue of being juxtaposed with a fortuitous negative emotion or event. Importantly, this effect only occurs after the extinction procedure, which demonstrates the long lasting aversive associations have. Exemplar prototypicality is a powerful determinant of category activation (Bruner, 1957; Locke et al., 2005; Medin & Smith, 1981) and a key gatekeeper of generalised changes in intergroup attitudes and stereotypes after discrete learning experiences with the outgroup (Brown & Hewstone, 2005; Rothbart & John, 1985). As such, a critical implication of these prototypicality shifts is that incidental negativity and anxiety not only worsens affective reactions to the outgroup, it also potentially makes individual learners more vulnerable to more and broader (i.e., generalised) negative changes in attitudes and stereotypes further down the track.

To conclude, to best of my knowledge, this research was the first to undergo a systematic analysis of the consequences of aversive conditioning for the categorisation of outgroup exemplars—against the backdrop of more documented affective consequences. As such, because of the present studies, this body of work now demonstrates that incidental negativity not only worsens outgroup exemplar affect, it is also responsible for increasing the perceived fit of these negative and anxiety-provoking exemplars to the outgroup prototype, thus affecting intergroup relations adversely, *once*, through *affect* and, a *second*, through *cognitions*.

Endnotes

1. An a-posteriori power analysis was conducted using the software package, PASS v14. The sample size of 66 was used for the statistical power analysis with one between subject factor (learning type) with 2 levels and two within subject factors (face type and time) with two levels each. The alpha level used for this analysis was $p < .05$. The post hoc analysis revealed the power for detecting a stimulus x time interaction was 1.00. The power for detecting a stimulus x time x learning type interaction was 0.22. Thus, I had adequate power to detect the two way interaction but suboptimal power to detect the three way interaction. A Bayesian analysis (Raferty, 1995; Jarosz & Wiley, 2014) provided positive support for the two factor model (stimulus and time) over the three factor model that involved learning type, inverse of Bayes factor 3.28. Thus, these analyses suggest that there was no difference in the prototypicality shift between direct and vicarious participants.
2. Prototypicality ratings were expected to follow a gradient and the target faces were expected to be rated higher in prototypicality, followed by the 25% variation faces and lastly the 50% variations. I checked that the expected gradient occurred prior to conditioning with a 2 learning type x 2 face type x 3 generalisation gradient (target, 25%, 50%) mixed model ANOVA with face type and generalization gradient as repeated measure on sorting task data. A generalization gradient main effect was observed, F (2,128) = 42.37, $p < .001$, $\eta_p^2 = .40$. The generalization gradient expected order of means was not detected and the 50% face variations were rated higher in prototypicality (M = 4.14, SD = 1.16), followed by the target faces (M = 4.12, SD = 1.39) followed by the 25% face variations (M = 2.80, SD = 1.22). This is in contrast to pilot test data that showed face prototypicality ratings along the expected gradient. Within this cohort of participants something may have caused

the 50% variation face to be rated higher in prototypicality, but the underlying reasons cannot be determined by this data. Despite this unexpected result, the faces that were designed to be closest to target faces showed prototypicality shifts similar to the target variation, whereas more distant faces did not.

3. Participant's pre-test similarity ratings between the two new exemplars and the unsafe and safe face were used to create an index for determining whether the new exemplar was more similar to the unsafe or safe face. Similarity was determined using the formula: New face similarity ([similarity with the unsafe face] – [similarity with the safe face]). A positive number indicates the new exemplar is more similar to the unsafe face and they were categorised as such; a negative number indicates the new exemplar is more similar to the safe face and they were categorised as such; a score of 0 indicates the new exemplar was perceived as equally similar to the unsafe and safe face.
4. For the basic prototypicality effects analyses, the covariate was an index of pre- vs. post-learning change in anxiety to the safe face, relative to the unsafe face. For the generalised prototypicality effects analyses, the covariate was an index of pre- vs. post-learning change in anxiety to the safe face, its 25% and 50% variations, relative to the unsafe face, and its 25% and 50% variations.
5. Nineteen pilot participants rated White and Black faces along perceived prototypicality and anxiety (1 = *not at all*, 7 = *very much*) in order to select comparable White faces to the Black faces previously used. The White four faces (two targets and two new faces) chosen were rated as being statistically comparable along prototypicality and anxiety, all p's > .05. The two configurally-related faces associated with each of the two target faces (25% and 50% variations) always followed a gradient along target, 25%, and 50% (set 1, *M*s [*SDs*] = 4.79[1.36], 4.26

[1.41] and 3.58 [1.26] ; set 2, *M*s [*SDs*] = 4.74 [1.74], 4.58 [1.14], and 3.58 [1.26]); face differences in anxiety were also comparable and followed a gradient (set 1, *M*s [*SDs*] = 2.47 [1.07], 2.26 [.80], and 2.11 [1.05]; set 2, *M*s [*SDs*] = 2.42 [1.12], 2.26 [0.90], and 2.16 [.76]). Hence, selected ingroup faces were suitable to test for basic prototypicality shifts of the two target faces as well as generalised shifts in prototypicality of the face variations and new faces.

6. An a-posteriori power analysis was conducted using the software package, PASS v14. The sample size of 61 was used for the statistical power analysis with two between subject factors (post-test position and learning type) with 2 levels and two within subject factors (face type and time) with two levels each. The alpha level used for this analysis was $p < .05$. The post hoc analysis revealed the power for detecting a stimulus x time x post-test position interaction was .53. The power for detecting a stimulus x time x post-test position x learning type interaction was 0.12. Thus, I had less than adequate power to detect the three and four way interaction. A Bayesian analysis (Raferty, 1995; Jarosz & Wiley, 2014) compared the three factor model (stimulus, time and post-test position) to the four factor model that involved type of learning and a 1.32 inverse of Bayes factor was found. Thus, there is no evidence to support or decline the 4 way interaction and its fit with the data over simpler models.
7. A chi-square test of independence was performed to examine the relationship between contingency awareness and the face stimuli that were counterbalanced. The relationship between these two variables was non-significant, X^2 (1, N = 61) = .770, p = .380. This non-significant result suggests my post-hoc grouping of contingency awareness did not undo face stimuli counterbalancing.

8. For analyses involving contingency awareness, I first ran ANOVAs that included contingency awareness as a factor. However, I always followed up these analyses with lower levels analyses separately for contingent aware and unaware participants irrespective of a significant (vs. non-significant) higher order interaction involving contingency awareness in the top level ANOVA. This approach is in line with Keppel and Wickens' (2004) recommendation to disregard top level analyses when apriori hypotheses are available. While reporting of the top ANOVA is not needed under planned comparisons, I have done so to ensure results are presented in full.
9. The effectiveness of the mask was checked with a series of funnelled questions beginning with questions about noticing anything unusual and ending with direct questions about whether they saw a face. These questions were assessed in the post-test self-reported questionnaire following the extinction procedure.
10. Due to a technical error, direct learning participants had the face presented for 27 milliseconds followed by the mask, whilst vicarious learning participants had the face presented for 17 milliseconds.
11. An a-posteriori power analysis was conducted using the software package, PASS v14. The sample size of 59 was used for the statistical power analysis with two between subject factors (post-test position and learning type) with 2 levels and two within subject factors (face type and time) with two levels each. The alpha level used for this analysis was $p < .05$. The post hoc analysis revealed the power for detecting a stimulus x time x post-test position interaction was .05. The power for detecting a stimulus x time x post-test position x learning type interaction was .05. Thus, I had less than adequate power to detect the three and four way interaction.

References

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661.

Amodio, D. M. & Devine, P. G. (2008). On the interpersonal functions of implicit stereotyping and evaluative race bias: Insights from social neuroscience. In R. E. Petty, R. H. Fazio, H. Russell, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 193-226). New York, NY, US: Psychology Press.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews: Neuroscience, 7*(4), 268–277.

Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459–1469.

Bandura, A. (1977). *Social learning theory.* New York: General Learning Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*(1), 5-25.

Bless, H., & Schwarz, N. (1998). Context effects in political judgement : Assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology*, (28), 159–173.

Bodenhausen, G. V. (1993). Emotions, arousal, and stereotypic judgments: A heuristic model of affect and stereotyping. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 13-37). San Diego: Academic Press.

Bodenhausen, G. V, & Peery, D. (2009). Social categorization and stereotyping in vivo : The VUCA challenge. *Social and Personality Psychology Compass*, *2*(3), 133–151.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(3), 219-231.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory, 11*, 485-494.

Bouton, M. E. (2014). Why behaviour change is difficult to sustain. *Preventative Medicine, 68*, 29-36.

Bouton, M. E., & Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning and Behavior*, *22*(3), 317-324.

Breitmeyer, B. G. (2007). Visual masking: Past accomplishments, present status, future developments. *Advances in Cognitive Psychology*, *3*(1-2), 9–20.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, Vol. 37 (pp. 255-342). San Diego, CA: Elsevier.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152.

Bush, G., Luu, P., & Posner, M. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222.

Cevik, M. O. (2014). Habituation, sensitization, and Pavlovian conditioning. *Frontiers in Integrative Neuroscience*, *8*, 13.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science*, *280*, 77–81.

Coats, S., Latu, I. M., & Haydel, L. A. (2007). The facilitative effect of evaluative fit on social categorization. *Current Research in Social Psychology*, *12*, 54 – 67.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.

Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorisation of multifaceted stimuli. *Journal of Personality and Social Psychology*, *77*(5), 927–941.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Cunningham, W. A, Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: FMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*(10), 1717–1729.

Cunningham, W. A. & Van Bavel, J. J. (2009). A neural analysis of intergroup perception and evaluation. In G.G. Berntson, & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences* (pp. 975-984). Hoboken, NJ: John Wiley & Sons Inc.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice, *Emotion, 9*(4), 585-591.

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes, *Psychological Science*, *15*(5), 319-324.

Dunham, Y. (2011). An angry= Outgroup effect. *Journal of Experimental Social Psychology, 47* (3), 668-671.

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature, 520* (7547), 345-348.

Enns, J., & Di Lollo V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352.

Esteves, F., Parra, C., Dimberg, U., & Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli, *Psychophysiology, 31*, 375-385.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18*(3), 232-239.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology, 65*(6), 1105-1118.

Harris, N., Griffin, A., & Paolini, S. (2016a). Learning about the world from watching others: Vicarious fear learning of outgroups and the moderating role of prior outgroup contact. Manuscript under preparation, University of Newcastle, Australia.

Harris, N., Paolini, S., & Griffin, A. (2016b). Similarity helps: Similarity underpins generalisation of interethnic anxiety during aversive observational learning. Manuscript under preparation, University of Newcastle, Australia.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry*, *60*, 361-368.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421.

Holmes, N. M., & Westbrook, F. R. (2014). ABA renewal is greater when extinction occurs in the same context as cue pre-exposure. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*(3), 369-379.

Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion*, *5*(3), 267–276.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640–643.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization. *Psychological Science*, *15*(5), 342–345.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352-376.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *Journal of problem solving*, *7* (1), 2-9.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological methods, 6* (2), 115.

Kattner, F., Ellermeier, W., & Tavakoli, P. (2012). Both trace and delay conditioning of evaluative responses depend on contingency awareness. *Learning and Motivation*, *43*(1-2), 35–47.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researchers handbook.* United States of America: Pearson Prentice Hall.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, *8*(6), 720–722.

Lipp, O. V. (2006). Human fear learning: Contemporary procedures and measurement. In M. G. Raske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic processes to clinical implication* (pp.37-51). Washington: American Psychological Association.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralisation of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry, 167,* 47-55.

Locke, V., Macrae, C. N., & Eaton, J. L. (2005). Is person categorization modulated by exemplar typicality? *Social Cognition*, *23*(5), 417–428.

Lovibond, P. F., Liu, J. C., Weidemann, G., & Mitchell, C. J. (2011). Awareness is necessary for differential trace and delay eyeblink conditioning in humans. *Biological Psychology, 87,* 393-400.

Lovibond, P. F., Saunders, J. C., Weidemann, G., & Mitchell, C. J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *The Quarterly Journal of Experimental Psychology*, *61*(8), 1199–1216.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3-26.

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47*(2), 391-396.

Maass, A., Castelli, L., & Arcuri, L. (2000). Measuring prejudice: Implicit versus explicit techniques. In D. Capozza, & R. Brown (Eds.), *Social identity processes: Trends in theory and research.* (pp. 96-117). London: SAGE Publications Ltd.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Mallan, K. M., & Lipp, O. V., & Cochrane, B. (2013). Slithering snakes, angry men and out-group members: What and whom are we evolved to fear? *Cognition and Emotion, 27*(7), 1168-1180.

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology*, *45*(6), 1303–1307.

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects: Applying social-cognitive theory to intergroup contact research. *Group Processes & Intergroup Relations, 14*(2), 255-274.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(4), 241–253.

Miller, N., & Dollard, J. (1941). *Social learning and imitation.* New Haven, NJ: Yale University Press.

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between “us” and “them”. *Journal of Personality and Social Psychology*, *99*(1), 62–77.

Mineka, S., & Ohman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry, 52*, 927-937.

Navarrete, C. D., McDonald, M. M., Asher, B.D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an out-group face in a minimal group context. *Evolution and Human Behavior*, *33* (5), 590-593.

Neumann, D. L., Lipp, O. V., & Cory, S. E. (2007). Conducting extinction in multiple contexts does not necessarily attenuate the renewal of shock expectancy in a fear-conditioning procedure with humans. *Behaviour Research and Therapy, 45*(2), 385-394.

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learning and Memory, 13*(6), 681-685.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotypes and social reality.* Oxford: Blackwell.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology, 30,* 125-144.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Ohman, A., & Mineka, S. (2003). The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *Current Directions in Psychological Science*, *12*(1), 5-9.

Ohman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following non-recognized fear-relevant stimuli. *Journal of Experimental Psychology: General, 127,* 69-82.

Olson, M. A, & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality & Social Psychology Bulletin*, *32*(4), 421–433.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*(5735), 785–787.

Olsson, A., Nearing, K. I., Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Social Cognitive and Affect Neuroscience, 2*(1), 3-11.

Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828.

Page, M. M. (1973). On detecting demand awareness by postexperimental questionnaire. *The Journal of Social Psychology*, *91*(2), 305–323.

Paolini, S., Harris, N. C., & Griffin, A. S. (2016). Learning anxiety in interactions with the outgroup: Towards a learning model of anxiety and stress in intergroup contact. *Group Processes and Intergroup Relations*, *19* (3), 275-313.

Park, B., & Judd, C. M. (2005). Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review*, *9*(2), 108–130.

Pashler, H., Baylis, G. C. (1991). Procedural learning: II. Intertrial repetition effects in speeded-choice task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(1), 33-48.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytical test of the intergroup contact theory. *Journal of Personality and Social Psychology, 90, 751-783.*

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis) liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*(1), 130-144.

Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32*(1), 84-99.

Raferty, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.

Ranganath, K. A, & Nosek, B. A. (2008). Implicit attitude generalisation occurs immediately; explicit attitude generalisation takes time. *Psychological Science, 19*(3), 249–254.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist, 43*(3), 151–160

Rescorla, R. A. (2004). Spontaneous Recovery. *Learning and Memory, 11,* 501-509.

Richeson, J. A, & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology, 89*(4), 517–530.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Oxford, England: Lawrence Erlbaum.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effect of intergroup contact. *Journal of Social Issues*, *41*, 81-104.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology*, *32*(4), 309-336.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis)categorization of unattractive brides and attractive prostitutes. *Experimental Psychology*, *55*(3), 182–188.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2*, 307-320.

Shechner, T., Hong, M., Britton, J. C., Pine, D. S., & Fox, N. A. (2014). Fear conditioning and extinction across development: evidence from human studies and animal models. *Biological Psychiatry, 100*, 1-12.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99*(1), 3-21.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404–420.

Steele, J. D., & Lawrie, S. M. (2004). Segregation of cognitive and emotional function in the prefrontal cortex: a stereotactic meta-analysis. *Neuroimage, 21*, 868-875.

Storbeck, J., & Clore, G. L. (2007). On the interdependence of cognition and emotion. *Cognition and Emotion, 21* (6), 1212-1237.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*

(2), 96-102.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-177.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of social conflict. In W. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp.33-47). California: Brooks/Cole.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalisation to the evaluation of novel faces. *Journal of Experimental Social Psychology, 49*, 661-669.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82(6)*, 919-934.

Wardle, S. G., Mitchell, C, J., & Lovibond, P. F. (2007). Flavor evaluative conditioning and contingency awareness. *Learning and Behavior, 35*(4), 233-241.

Weidemann, G., & Kehoe, E. J. (2003). Savings in classical conditioning in the rabbit as a function of extended extinction. *Learning and Behavior, 31*(1), 49-68.

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science, 326*(5960), 1711–1714.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*(5), 815-8127.

Wyer, N. A, Sadler, M. S., & Judd, C. M. (2002). Contrast effects in stereotype formation and change: The role of comparative context. *Journal of Experimental Social Psychology*, *38*(5), 443–458.

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*(3), 424-431.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

Chapter 4: Causal Role of Extinction in Modifying Perceived Prototypicality

Previous research provided evidence that exemplars associated with an aversive stimulus were perceived as being more negative/anxiety provoking following the association (Mallan, Sax, & Lipp, 2009; Olson & Fazio, 2006; Olsson, Ebert, Banaji, & Phelps, 2005; Weisbuch, Pauker, & Ambady, 2009). In Chapter 3 I expanded on this research and investigated the consequences the pairing between an outgroup exemplar and aversive stimulus had on the exemplars cognitive representation through changes in perceived prototypicality. Outgroup exemplars were perceived to be more prototypical of the outgroup after the negative/anxiety provoking association had been acquired and extinguished compared to pre-test. The shift in prototypicality is consistent with evaluative-fit/emotion-fit mechanisms because negativity/anxiety provides a better fit with the Black outgroup (Cottrell & Neuberg, 2005; Tajfel & Turner, 1979, Turner, Hogg, Oakes, Reicher & Wetherell, 1987). Whilst I expect evaluative-fit and emotion-fit to be driving mechanisms involved in shifting exemplar prototypicality, Chapter 3 identified other underlying mechanisms. The extinction process was implicated in prototypicality shifts, and this chapter aims to investigate the role of extinction in more depth to determine the causal role it has in modifying perceived prototypicality.

Studies within Chapter 3 provided evidence that exemplar prototypicality shifts occurred after the extinction procedure and not beforehand. Conditioning consisted of one outgroup exemplar being paired with a negative/anxiety provoking stimulus (unsafe face), whilst another exemplar was never paired with the stimulus (safe face). Perceived exemplar prototypicality was measured prior to, and after extinction. In Study 3.1 post-test prototypicality data was collected post-extinction and the outgroup exemplar was perceived as being more prototypical of the outgroup than at pre-test. Study 3.2

manipulated the time-point at which post-test prototypicality data was collected, providing a better investigation of the role extinction played. In one group post-test prototypicality was again collected post-extinction as in Study 3.1, whilst the other group had post-test prototypicality data immediately after acquisition and before extinction (post-acquisition group). I expected heightened negativity/anxiety in the post-acquisition group to affect the process of categorisation with the unsafe exemplar to cause greater shifts in exemplar prototypicality as a result of stronger links with evaluative-fit/emotion-fit mechanisms. Unexpectedly, no shift of prototypicality was found for the unsafe outgroup face (vs safe face) post-acquisition, and the unsafe face was found to be perceived as more prototypical of the outgroup post-extinction only. These results suggest extinction plays a role in facilitating exemplar prototypicality shifts.

My extinction procedure does not have a set number of trials and extinction occurs when four consecutive trials with no increase in Skin conductance responses occur. Therefore, the number of trials varied for each participant and Study 3.2 took advantage of this by measuring the number of presentations each participant underwent for extinction. I performed a mediational analysis on the effect this inter-individual variation in number of extinction trials had on changes in exemplar prototypicality and found prototypicality shifts were larger as the number of presentations increased. Based on this evidence, I hypothesised that repeated presentations of faces in the absence of any aversive stimulation contributed to shifting an exemplars prototypicality.

Repeated presentations may contribute to shifting exemplar prototypicality by increasing the familiarity of exemplars (Garcia-Marques & Mackie, 2007; Smith, Miller, Maitner, Crump, Garcia-Marques & Mackie, 2006). Repeated exposure with an exemplar is likely to increase familiarity and factors associated with positive evaluation which typically limit stereotypical processing (Fiske & Neuberg, 1990; Zajonc, 1968). However, when exposure is unconfounded with positive evaluation, such as in my experimental paradigm involving negative/anxiety provoking associations, then stereotyping is expected to increase (Smith et al., 2006). For example, Smith et al. (2006) presented participants with a series of individual faces to remember in phase 1 of their experiment. Phase 2 consisted of faces being presented with a brief text description that included an occupation label and information inconsistent with the occupational stereotype. Half of the faces presented during Phase 2 were repeated from Phase 1, whilst the other half of faces were new. Participants rated each face on five trait judgments; three traits were relevant to the group stereotype and the other 2 traits were fillers that acted as a control. Results suggest that text descriptions accompanied by repeated faces were rated more stereotypically than newly presented faces, whilst no differences were found in the filler traits. This result is consistent with the idea that repeated presentations increase familiarity resulting in increased stereotyping.

In Chapter 3, repeated presentations of individual features (i.e., facial features) and group membership cues (i.e., Black skin) were confounded with the simple passage of time, and this could also be a factor in shifting exemplar prototypicality. The passage of time throughout extinction could strengthen memory traces learnt during extinction, which would explain why exemplar prototypicality shifts occur post-extinction. This effect would be consistent with sleep studies that demonstrate sleep strengthens memories, particularly those that have an evaluative/emotional association (Feld & Diekelmann, 2015; Landman et al., 2016; Rauchs et al., 2011). In a mechanism similar to that observed during sleep studies, evaluative-fit/emotion-fit might need time in order to consolidate and strengthen in order to shift prototypicality. This interpretation is consistent with Study 3.2 mediational analysis involving number of extinction trials, because as the number of repeated presentations increased – thereby increasing the passage of time – the larger the prototypicality shift of the unsafe face was. Therefore, the simple passage of time may also contribute to prototypicality shifts.

Study 4.1

Study 4.1 was designed to test whether individuating information or group membership cues were required to be repeatedly presented in order to produce post-extinction shifts of prototypicality. I could have exposed participants to a number of factors such as repeated presentations of an unrelated stimulus (i.e., completely unrelated to individuating information or group membership cues) or a simple passage of time, but I opted to restrict my analysis to repeated presentations of individuating information or group membership cues. I did this because these two types of cues constituted what I considered were the most likely to shift prototypicality within the outgroup context of this study. Testing each and every possibility was logistically impossible due to the large sample sizes required. Individual identity cues contain both discriminative facial features between exemplars and group membership markers. Thus all information can be attended too. If individual identity cues are not needed to shift prototypicality, the next logical choice would be repeated presentations of group membership cues. Therefore, the current study aims to explore the exact mechanism that needs to be repeated in order to determine what component of the extinction process modifies prototypicality.

This study used the same basic experimental procedure described in Chapter 3. One outgroup exemplar was paired with a negative/anxiety provoking stimulus, whilst another outgroup exemplar never received the pairing. Changes in the perceived prototypicality of unambiguous Black computer generated faces with a neutral expression were measured prior to conditioning, and at post-extinction. Type of learning and post-test position manipulations were not investigated in the study. Furthermore, this study focused on exemplar prototypicality shifts and treated acquisition and extinction as manipulation checks rather than exploring the interplay between cognition and affect.

This study differed from the previous chapter as I manipulated type of extinction in a between subjects fashion to determine the component involved in extinction responsible for modifying prototypicality. In particular, I tested whether individuality cues (individuality group; contains both individuating and group membership cues) or group memberships (category membership group; contains only group membership cues and no individuating information) needed to be repeatedly presented in order to shift exemplar prototypicality. The individuality group underwent the standard extinction procedure described in Chapter 3 and viewed the unsafe and safe face in the absence of any aversive stimulus. If both individuality and group membership cues need to be repeatedly presented in order to shift prototypicality, I expect to find exemplar prototypicality shifts in this group only.

The category membership group differed from the standard extinction procedure used in the individuality group because scrambled faces were repeatedly presented during extinction. Two scrambled images were created in a way that they could no longer be individually recognised but retained their original marker of group membership, that is, their Black skin. Each participant in the category membership group was matched to the number of presentations needed for extinction in a similar participant in the individuality group (matched based on age and gender). In this way I ensured that any between subjects differences in prototypicality were solely attributed to the nature of the face stimuli experienced during extinction (individuality vs category membership) and not to any between group differences in the number of extinction trials. If individuality cues are not required to shift prototypicality, I expect to find exemplar prototypicality shifts in the category membership group. This result would suggest individuality cues do not need to be repeatedly presented in order to shift prototypicality. Rather, repeated presentations of group membership cues are sufficient.

Furthermore, as this chapter is an extension of Chapter 3, I continued to explore the same underlying mechanisms. Evidence from Chapter 3 suggests awareness of the contingent relationship between the unsafe face and electrical stimulation contributed to increased shifts of prototypicality of the unsafe face post-extinction. I explore the role contingency awareness has for prototypicality shifts statistically, through the use of a self-reported questionnaire (Clark & Squire, 1998; Page, 1973). In addition, I explored the mediational role that anxiety in general at post- acquisition and number of repeated presentations had for prototypicality shifts.

Method

Participants and Design

Participants were 59 students (19 male, 40 female; M = 23.03, SD = 8.05) from a large regional Australian university. All participants self-identified as being from a White, Anglo-Saxon background. They received monetary compensation (AU$20) or partial course credit for their participation. Participants were randomly assigned to one of two types of extinction condition that varied in the procedure used during extinction (individuality group n = 30, category membership group n = 29).

Procedures and Measures

Response biases associated with repeated measures were minimised by collecting pre-test data between 5 and 28 days (M = 13.55, SD = 4.26) before post-test data collection. During the initial session, participants were seated in front of a computer and completed an online questionnaire and a speeded sorting task as part of pre-test prototypicality data collection. The Black outgroup faces used in this study were the same faces as described in Chapter 3.

The online questionnaire required participants to indicate the extent to which each of the randomly ordered Black outgroup faces was prototypical of Black people in general (*prototypically Black*: 1 = *not at all*, 7 = *very much*). To check that the association between the unsafe face and electrical stimulation became ambiguous, I also asked participants to rate faces according to how anxiety provoking they perceived them to be if they were to meet them (*anxious*: 1 = *not at all*, 7 = *very much*) and how similar each pair of faces was (*similar*: 1 = *not at all*, 7 = *very much*).

During the speeded sorting task participants were presented with individual Black and White target, generalisation and new exemplar faces at the centre of the screen. The face set and instructions replicated that used in Chapter 3 and included two prototypically Black and White exemplars, a 25% and 50% variation of the target faces and two new prototypical Black and White exemplars that were different to the target exemplars. Participants were instructed to sort each face as quickly and as accurately as possible into either the “Black” or “White” category by pressing the green (left handed “S” key) or blue (right handed “L” key) on the keyboard. Category labels were presented in the top left and right corners of the computer screen and corresponded to the location of the relevant key. Each face was presented 14 times and the response keys and category labels were counterbalanced on each side for half of the trials. To increase task difficulty and engagement, faces were inverted for 25% of the presentations (Richeson & Trawalter, 2005).

The second laboratory session began with the work up procedure to select the level of shock to be used during acquisition. The shock was regarded as being uncomfortable, but not painful by participants (Lovibond, Saunders, Weidemann, & Mitchell, 2008). A Powerlab 4/25T (ADInstruments) administered the electrical stimulation through an inbuilt isolated stimulator using a bar electrode attached to the participant's right forearm. To measure participant's physiological arousal, skin conductance electrodes were attached to the distal phalanges of the first and second digits on the participant's left hand and measured via an ADInstruments Model ML116 GSR amplifier using standard MLT116F electrodes. An ADInstruments MLT1132 Piezo respiration belt was attached around the participant's chest to monitor for artefacts such as deep breathing or coughing that would distort the psychophysiological measurements.

During acquisition, two target faces were presented at the centre of the screen six times in a randomised order for 10 s (inter-stimulus interval M = 17.5 s, range 15-20 s). One target face (unsafe face) always co-terminated with a 200 ms electrical stimulation at the level self-selected by the participant during the work-up procedure. The other target face (safe face; counterbalanced) never co-terminated with a stimulation. Pre-test skin conductance data were collected immediately before the acquisition procedure and post-acquisition skin conductance data were collected immediately after acquisition but prior to initiating extinction. To this end, each target face was presented twice for 10 s in the absence of electrical stimulation both before and immediately after acquisition.

The extinction procedure followed post- acquisition skin conductance data collection. Extinction type was manipulated at this stage of the study. Participants randomly assigned to the individuality group underwent the standard extinction procedure used in Chapter 3. The two target faces (unsafe and safe face) were presented an identical number of times in the absence of electrical stimulation until the faces no longer evoked physiological activation over four consecutive presentations (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). The number of extinction trials was set to a minimum of 5 and maximum of 25 presentations per target face. The other half of participants were assigned to the category membership group and viewed scrambled images of the two target faces. The scrambled faces were created using Adobe Photoshop by cutting the image into small pieces and reassembling the pieces randomly in a way which prevented individual identifiable traits from being identified, but kept their marker of category membership (see Figure 10 for the stimuli used in each extinction group). Participants in the category membership group were matched with a participant from the individuality group based on similar age and gender. They underwent the same number of extinction trials as their matched participant, rather than waiting for four consecutive trials of no increases in SCR.

An online questionnaire measured post-extinction self-reported prototypicality and anxiety data using the procedure described for the pre-test. Additionally, I assessed each participant's level of contingency awareness as part of the online questionnaire. Contingency awareness was measured and checked for using the same procedure described in Study 3.2. Briefly, these questions involved asking participants if they reacted the same way to all faces, whether they noticed a pattern in which face was paired with an electrical stimulation, which face was paired with the electrical stimulation and how confident were they in their decision (adapted from Clark & Squire, 1998; Page, 1973; see Appendix O). In order to be considered contingent aware participants needed to accurately select the face paired with shock, with a high degree of confidence (i.e., On a scale 1 -7, a rating of 4 or more on how confident they were the face they selected was paired with the electrical stimulation). Participants also needed to accurately describe the association between the exemplar and electrical stimulation in an open-ended response. Forty participants were classified as contingent non-aware (20 individuality group; 20 category membership group) and 19 as contingent aware (10 individuality group; 9 category membership group). As per ethics requirements, participants in the category membership group underwent the standard extinction procedure to ensure the association developed during acquisition was extinguished.

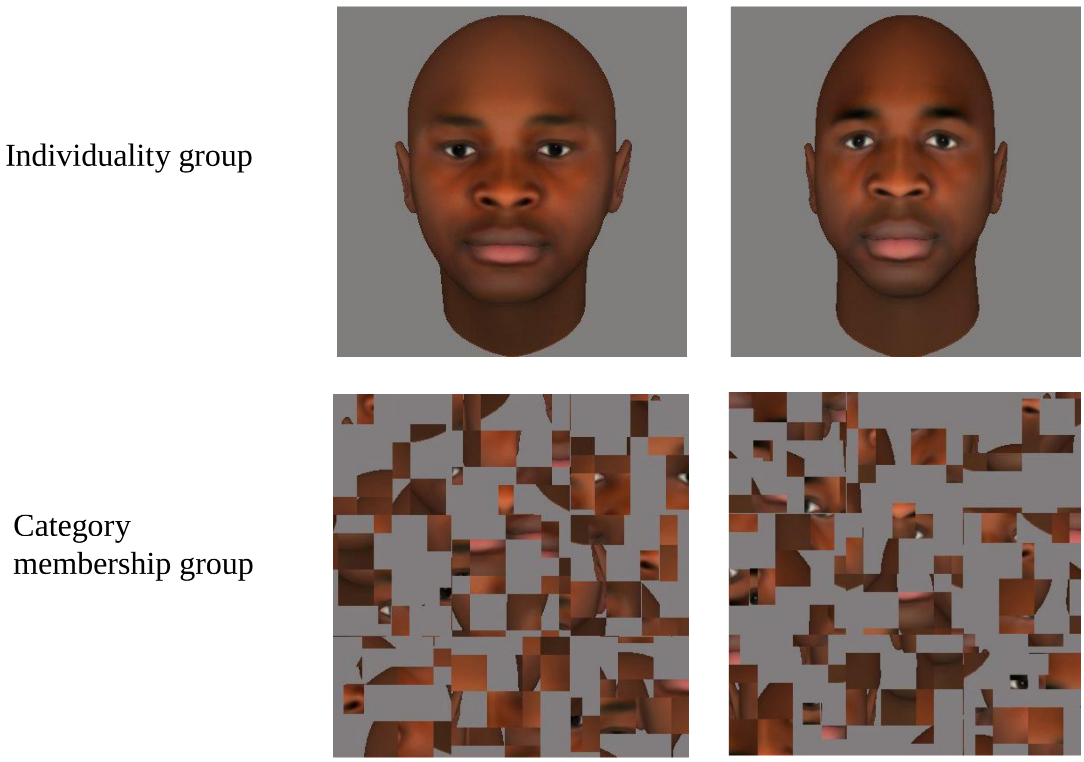


Figure 10. Stimuli used in the type of extinction group manipulation.

Results

Checking Effective Acquisition and Extinction

To determine if effective acquisition and extinction occurred, I first examined participant's skin conductance responses (SCRs). SCRs were recorded and scored following standard guidelines that were described in Study 3.1 (Boucsein et al., 2012; Fowles, Christie, Edelberg, Grings, Lykken & Venables, 1981). Briefly, to measure effective acquisition, first interval SCRs with a minimum amplitude 0.05 μS were calculated and averaged across two presentations for each of the two target faces immediately prior to, and after acquisition. To measure extinction, the same process was used but scores were calculated for the first block of extinction (first two presentations of the unsafe and safe face respectively) and the ten trial block (the last two presentations of the unsafe and safe face respectively up to the tenth trial – minimum number of trials all participants received to satisfy extinction criteria).

Effective acquisition was checked with using a 2 face type (unsafe and safe face) x 2 time (pre- and post- acquisition) repeated measures ANOVA on SCR data. Increases in SCRs were equated to increases in anxiety. A time main effect was detected, F (1, 58) = 8.30, p = .004, ηp^2 = .13. Participants were more anxious at post-acquisition (M = 1.15, SD = .17) than at pre-acquisition (M = 1.08, SD = .12). More importantly, a face type x time interaction was detected, F (1, 57) = 11.13, p = .001, ηp^2 = .16 (Figure 11). The unsafe face has higher SCRs at post- acquisition (M = 1.20, SD = .28) than at pre-acquisition (M = 1.05, SD = .11), t (58) = -3.88, p < .001. In comparison, the safe face showed no change in anxiety from post- acquisition (M = 1.10, SD = .18) to pre-acquisition (M = 1.11, SD = .18), t (58) = .23, p = .817. There is also a difference in post-acquisition scores between the unsafe and safe face, t (58) = 2.33, p = .023.There is also a difference in pre-acquisition scores between the unsafe and safe face, t (58) = -2.23, p = .029. Although there was are differences between the unsafe and safe face at post-test and pre-test, this does not limit my interpretations. My research is interested in *changes* within each face, which is better highlighted by the significant interaction.. These results demonstrate that pairing the unsafe face with an aversive electrical stimulation was effective at causing the unsafe exemplar to be perceived as more negative/anxiety provoking after acquisition relative to before acquisition.

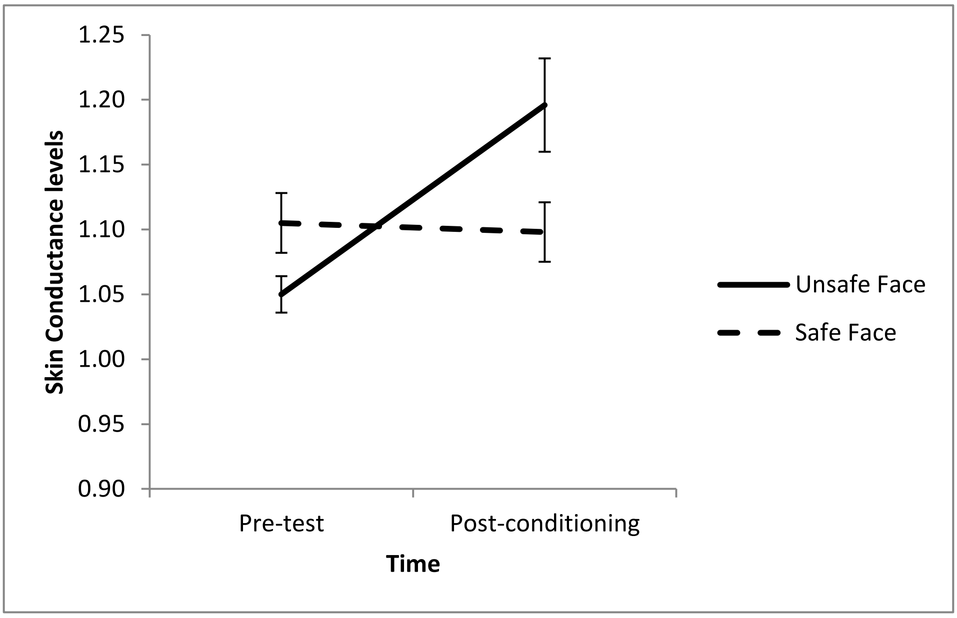


Figure 11. Face type x time interaction to check for effective acquisition.

To check for effective extinction, I ran a 2 type of extinction (individuality group, category membership group) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR data. Category membership group participant's extinction data was calculated from the final extinction procedure at the very end of the experiment, when faces rather than scrambled images were presented. As expected there was no main effect of face type, F (1, 57) = .63, p = .432, η_p^2 = .01 or face type x extinction block interaction F (1, 57) = .10, p = .759, η_p^2 = .00. These two results suggest there is no difference between the unsafe and safe face throughout extinction. Type of extinction did not influence results, observed by a non-significant 2 way interaction involving face type and extinction block, F (1, 57) = .46, p = .500, η_p^2 = .01. Supplemental data analyses described in Study 3.1 were conducted on SCR data and confirmed negativity/anxiety was extinguished and are discussed extensively in Appendix P4.

In the previous extinction analysis, category membership group SCR data from the end of the experiment was used because faces were presented, whereas the same scrambled image was presented during matched extinction in the first instance. To check extinction occurred in both groups I also checked self-reported anxiety data collected at pre-conditioning and post-extinction using a 2 face type (unsafe and safe face) x 2 time (pre-conditioning and post-extinction) x 2 type of extinction (individuality group, category membership group) mixed ANOVA with face type and time as repeated measures. A main effect of face type, F (1, 57) = 6.345, p = .015, ηp^2 = .10, and time F (1, 57) = 55.97, p < .001, ηp^2 = .50 were detected. The time main effect showed a non-associative effect where both the unsafe and safe face were perceived as being more anxiety provoking post-extinction (M = 6.04, SD = 1.90) than at pre-conditioning (M = 5.62, SD = 1.30). Unexpectedly, a face type x time interaction was also observed, F (1, 57) = 10.04, p = .002, ηp^2 = .15 (Figure 12). The unsafe face was regarded as more anxiety provoking post-extinction (M = 3.93, SD = 1.97) than at pre-conditioning (M = 2.31, SD = 1.25), t (58) = -8.32, p < .001. Similarly, the safe face displayed the same effect but of a smaller magnitude. The safe face was regarded more anxiety provoking at post-extinction (M = 3.26, SD = .184) than at pre-conditioning (M = 2.34, SD = 1.75), t (58) = -4.01, p < .001. These two results are presented in Figure 12. This result suggests residual self-reported anxiety persisted towards both target faces post-extinction. The three way interaction involving type of extinction was non-significant, F (1, 63) = 1.63, p = .207, ηp^2 = .03. The non-significant 3 way interaction suggests residual anxiety was present in both types of extinction groups.

Together, the results suggest conditioning was effective and negativity/anxiety became associated with the unsafe exemplar following conditioning. Whist SCR data suggests extinction occurred, self-reported anxiety unexpectedly does not. Rather, self-reported anxiety suggests residual anxiety of the unsafe and safe face persisted throughout extinction.

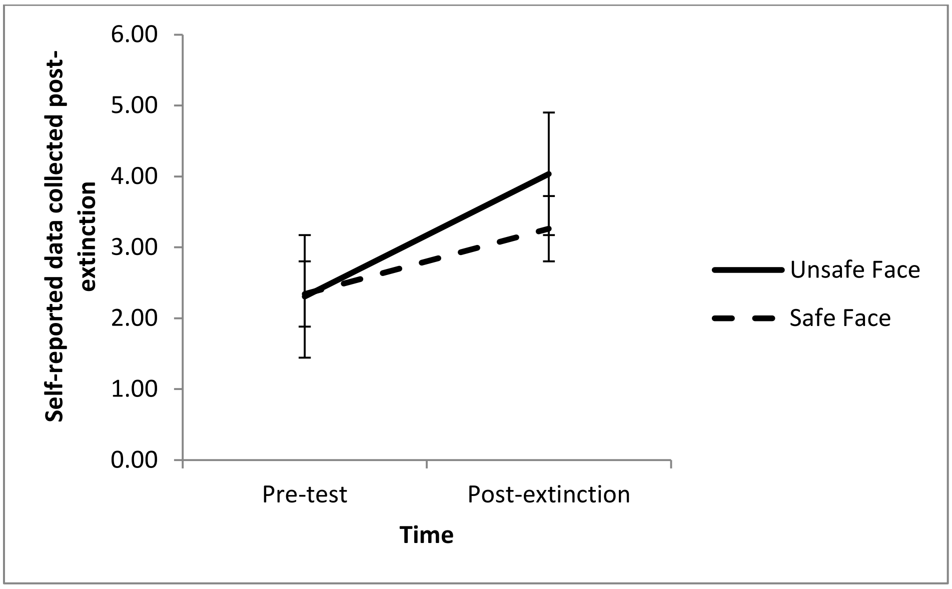


Figure 12. Face type x time interaction to check for effective extinction.

Testing Basic and Generalised Prototypicality Shifts

To test for prototypicality shifts I analysed self-reported prototypicality and sorting task data. Similar to Study 3.2, I predicted the unsafe face would be perceived as being more prototypical post-extinction than at pre-test. A shift of prototypicality in this direction was expected in the individuality group, but not the category membership group, if repeated exemplar presentations with facial features during extinction are a key mechanism underlying the shift. However, if repeated exemplar presentations were not an underlying mechanisms responsible for prototypicality shifts I expected to see exemplar prototypicality shifts in both individuality and category membership groups.

Basic prototypicality shifts were first tested for with self-reported prototypicality data using a 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) x 2 type of extinction (individuality group, category membership group) mixed model ANOVA with face type and time as the repeated measures. A main effect of time was detected, F (1, 57) = 11.21, p = .001, ηp^2 = .16. Both target faces (unsafe and safe) were regarded as being more prototypical of the outgroup post-extinction (M = 6.04, SD = .87) than at pre-test (M = 5.62, SD = 1.21). Unexpectedly, there was no face type x time or face type x time x type of extinction interaction detected, p's > .14.

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A type of extinction x 2 face type x 3 generalisation gradient (target, 25%, 50%) by x time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on the self-reported prototypicality data. The time main effect held and all faces were perceived as being more prototypical at post-extinction (M = 4.00, SD = .95) than at pre-test (M = 3.70, SD = .74), F (1, 57) = 5.71, p = .020, ηp^2 = .09. No other effects were found, p's > .12; hence, no further analyses were carried out on the variation data.

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face (new exemplar 1 N = 25, new exemplar 2 N = 25), safe face (new exemplar 1 N = 18, new exemplar 2 N = 24), or equally similar (new exemplar 1 N = 16, new exemplar 2 N = 10). The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 type of extinction x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on self-reported data.

A time main effect was found for the first new Black face, F (1, 53) = 6.54, p = .013, η_p^2 = .11, and the second new Black face, F (1, 53) = 18.63, p < .001, η_p^2 = .26; Faces were rated quicker at post-extinction (First new Black face: M = 6.36, SD = .74; Second new Black face: M = 6.83, SD = .42) than at pre-test (First new Black face: M = 5.98, SD = 1.10; Second new Black face: M = 6.11, SD = 1.17). No other effects were found, all p's > .510.

Latencies during the speeded sorting task were also used to investigate prototypicality shifts towards the unsafe and safe exemplars. The more prototypical a face is of a category, the faster it should be accurately sorted into its respective category due to its psychological proximity to the group's prototype. Hence, decreases in latency across time indicate that the face has become more prototypical of the outgroup. The same procedure described in Study 3.2 was used to measure and analyse data from the speeded sorting task. Briefly, incorrect categorisation responses were excluded from the latency data analysis and the mean reaction time of the latencies for the correctly categorised unsafe and safe exemplars were log-transformed to normalise the data (Richeson & Trawalter, 2005; Ruys, Dijksterhuis & Corneille., 2008). Responses quicker than 300 ms were excluded from the analyses and extremely long responses (> 3 SD) were rescored to the third standard deviation value for each group. On average, participants incorrectly sorted faces 8.50% of the time (SD = 5.68); a one way ANOVA confirmed there was no systematic difference in errors as a function of type of extinction, p = .224. The average time taken to sort faces were: The unsafe exemplar at pre-test 554 ms (SD = 82 ms) and at post-test 539 ms (SD = 97 ms); the safe exemplar at pre-test 561 ms (SD = 96 ms) and post-test 528 ms (SD = 88 ms).

Basic prototypicality shifts of the target faces were investigated using a 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) x 2 type of extinction (individuality group, category membership group) mixed model ANOVA with face type and time as repeated measures on sorting task data. A main effect of time was detected, F (1, 57) = 11.52, p = .001, ηp^2 = .17. Both target faces (unsafe and safe) were perceived as being more prototypical of the outgroup post-extinction (M = 2.72, SD = .07) than at pre-test (M = 2.74, SD = .06). Unexpectedly, no interactions involving face type x time or face type x time x type of extinction were detected, all p's > .142.

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A type of extinction x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on the sorting task data. The time main effect held and all faces were perceived as being more prototypical as demonstrated by quicker response times to sort faces at post-extinction (M = 2.78, SD = .07) than at pre-test (M = 2.81, SD = .08), F (1, 54) = 24.94, p < .001, ηp^2 = .32. No other effects were found, p's > .131; hence, no further analyses were carried out on the variation data.

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face, safe face, or equally similar. The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 type of extinction x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on sorting task data. No effects were found for either new face, all p's > .185.

Together, the time main effect and non-significant face type x time interaction found in self-reported prototypicality and speeded sorting task suggest a non-associative prototypicality shift of both the unsafe and safe faces irrespective of whether they were paired with shock or not. This is in contrast to Study 3.1 and 3.2 prototypicality data, which found discriminative prototypicality shifts (i.e., a change in prototypicality for one face but not another). The time main effect generalised to exemplars who were similar to faces presented during acquisition and extinction, but did not generalise to new exemplars. The non-significant face type x time x type of extinction interaction means the non-associative prototypicality shift did not differ based on the type of extinction participants underwent.

Exploring the Role of Contingency Awareness in Prototypicality Shifts

I conducted 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) x 2 type of extinction (individuality group, category membership group) x 2 contingency awareness[1] (contingent aware vs contingent unaware) mixed model ANOVAs with face type and time as repeated measures. A marginal face type x time x contingency interaction was detected and subsequently followed up by looking at level of contingency awareness separately, F (1, 55) = 3.02, p = .088, ηp^2 = .05. No effect was found for the higher interaction involving type of extinction, suggesting the effects were of a comparable magnitude between the individuality and category membership groups, p = .428.

Participants in the contingent aware group showed a main effect of time, F (1, 17) = 4.81, p = .042, ηp^2 = .22. Both target faces were regarded as being more prototypical of the outgroup post-extinction (M = 6.30, SD = .88) than at pre-test (M = 5.84, SD = 1.19). Contingent unaware participants also showed a main effect of time, F (1, 38) = 6.35, p = .016, ηp^2 = .14. Both target faces were regarded as being more prototypical post-extinction (M = 5.91, SD = .85) than at pre-test (M = 5.51, SD = 1.17). No other effects were detected, p’s > .05. Unexpectedly, interactions involving face type x time in both contingent aware and unaware participants were non-significant, p’s

> .132. Face type x time effects are presented in Figure 13 for contingent aware participants and Figure 14 for contingent unaware participants. These figures unexpectedly show the marginal 3 way interaction was driven by larger differences in prototypicality for the safe face in contingent aware participants and larger differences for the unsafe face in contingent aware participants. I expanded the model to include face variations and the time main effect in contingent aware participant held, F (1, 17) = 5.10, p = .037, ηp^2 = .23; all other effects were non-significant, all p's > .112.

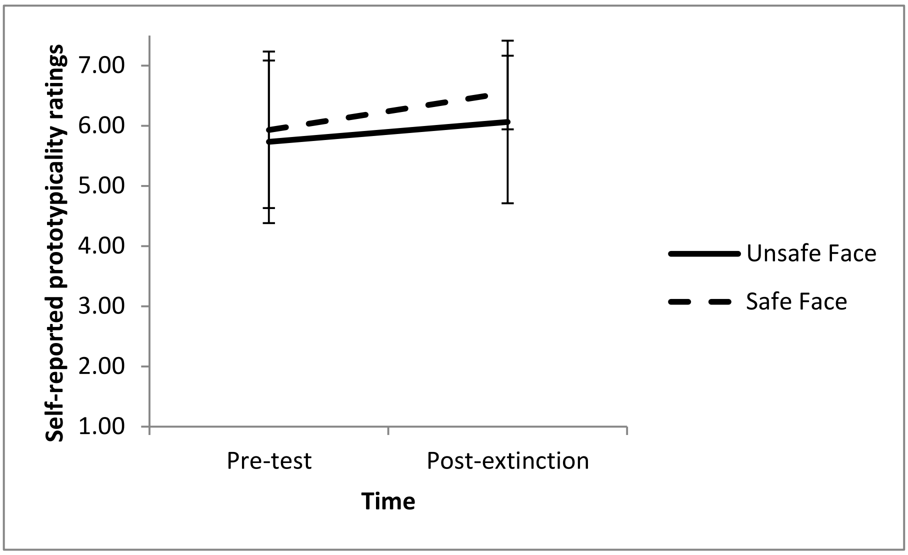


Figure 13. Contingent aware face type x time self-reported prototypicality

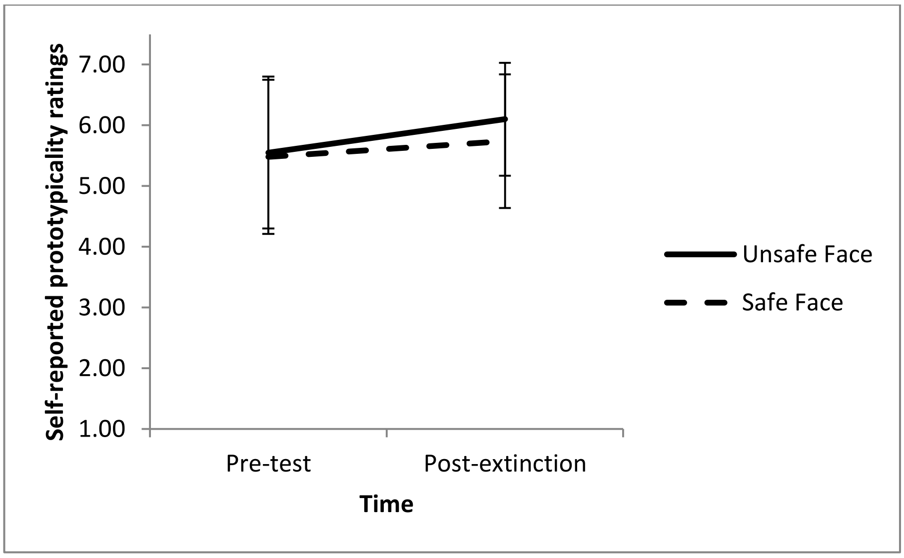


Figure 14. Contingent unaware face type x time self-reported prototypicality

Next I investigated the effects contingency awareness had on prototypicality shifts measured with the speeded sorting task. I conducted two 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) x 2 type of extinction (individuality group, category membership group) x 2 contingency awareness (contingent aware vs contingent unaware) mixed model ANOVAs with face type and time as repeated measures on sorting task data. A significant face type x time x contingency awareness interaction was detected, F (1, 55) = 7.73, p = .007, ηp2 = .12 and subsequently followed up by looking at contingent aware and unaware participants separately.

Contingent aware participants displayed a main effect of time, F (1, 18) = 8.19, p = .01. Both faces were perceived to be more prototypical post-extinction (M = 2.72, SD = .06) than at pre-test (M = 2.74, SD = .06). A face type x time interaction was detected in these participants, F (1, 18) = 12.34, p = .002, ηp2 = .41. This interaction is displayed in Figure 15[2]. As originally predicted, among contingent aware participants, the unsafe face became more prototypical of the outgroup post-extinction (M = 2.71, SD = .05) than at pre-test (M = 2.75, SD = .05), t (18) = 3.86, p = .001. In comparison, the safe face showed no change in prototypicality as sorting latencies for the face did not change from post-extinction (M = 2.72, SD = .06) to pre-test (M = 2.74, SD = .07), t (18) = 1.68, p = .110. The 3 way interaction involving face type, time and type of extinction was non-significant, F (1, 17) = .01, p = .945, ηp2 = .00. I expanded this model to include *face variations*, however no effects were found, all p's > .196. Together the results suggested participants aware of the face-stimulation pairing had a larger change in exemplar prototypicality of the unsafe face and it was perceived as being more prototypical of the outgroup post-extinction among contingent aware participants. The non-significant 3 way interaction suggests that this effect was comparable between groups and category membership cues alone were sufficient to shift prototypicality.

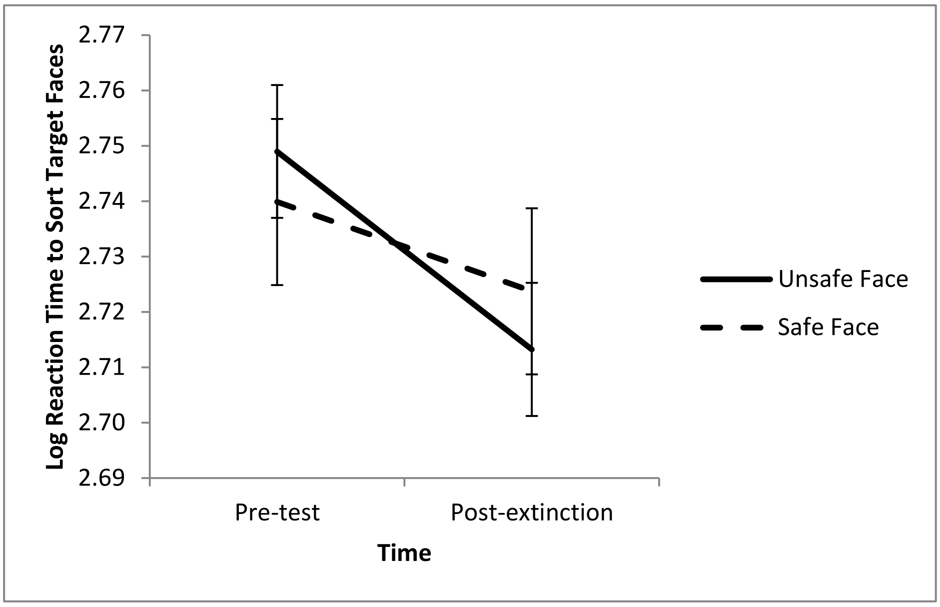


Figure 15. Face type x time interaction on log-transformed reaction times on the sorting task for contingent aware participants.

Participants in the contingent unaware group also displayed a significant face type x time interaction, *F* (1, 38) = 6.05, *p* = .019, $\eta p^2 = .14$. The unsafe face did not change in prototypicality from post-extinction (*M* = 2.73, *SD* = .08) to pre-test (*M* = 2.73, *SD* = .07), *t* (31) = .02, *p* = .988. This interaction is displayed in Figure 16. Unexpectedly, this interaction was driven by the safe face becoming more prototypical as it was sorted more quickly at post-extinction (*M* = 2.71, *SD* = .07) than at pre-test (*M* = 2.74, *SD* = .07), *t* (39) = 2.85, *p* = .007. The 3 way interaction involving type of extinction was non-significant, *p* = .409. I expanded this model to include *face variations,* however no effects were found, all *p*'s > .572. Together, results from participants who were unaware of the face-stimulation suggest an unexpected result and the safe face (vs unsafe face) was perceived as being more prototypical of the outgroup post-extinction. The non-significant 3 way interaction suggests that this effect was comparable between groups and category membership cues alone were sufficient to shift prototypicality.

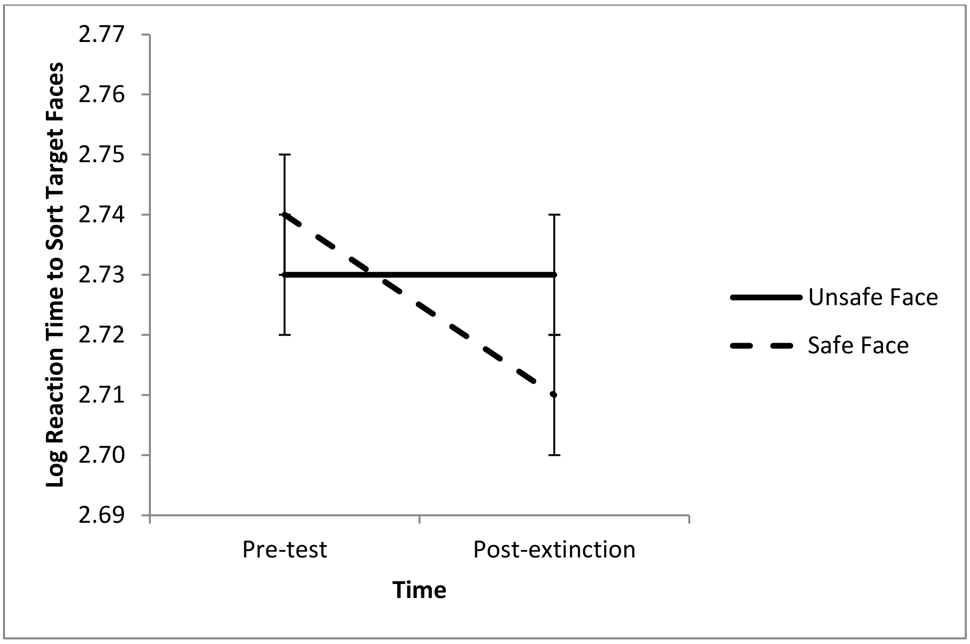


Figure 16. Face type x time interaction on log-transformed reaction times on the sorting task for contingent unaware participants.

Exploring the Role of Mediators in Contingent Aware Prototypicality Shifts

The following analyses are reported on the self-reported prototypicality results for contingent aware participants. Analyses is restricted to this group on the self-reported measure because significant effects were only found for this group on this measure and therefore are the only interactions that can be nullified in mediational tests. To explore the role that anxiety in general has for prototypicality shifts, I performed a mediational analysis with anxiety in general. I focused the mediational analyses on the sorting task data and included contingency awareness because of the prototypicality shifts detected. Anxiety in general was calculated as the mean post-test skin conductance arousal level in response to the unsafe and safe faces collected immediately after acquisition had occurred. To determine the direction of these effects the difference of a difference of prototypicality was correlated with anxiety in general[3]. The difference of a difference score suggests that as prototypicality shifts of the unsafe face increase, the more general anxiety is present. A mediational analysis was performed by entering the average skin conductance response (SCR) at post-acquisition as a covariate into a face type x time x type of extinction x 2 contingency awareness ANCOVA (Baron & Kenny, 1986; Judd, Kenny & McClelland, 2001; Yzerbyt, Muller & Judd, 2004). Anxiety in general increased the significance of the face type x time x contingency awareness interaction, from F (1, 55) = 3.02, p = .088, ηp^2 = .05 to F (1, 54) = 6.66, p = .013, η_p^2 = .11. I followed this effect up by running the same analysis but separately for each level of contingency awareness.

Similar to the previous chapter, I predicted anxiety in general was a factor in the prototypical shift and entering it into the ANCOVA would cancel out or significantly reduce the face type x time interaction among contingent aware and unaware participants. Hence, anxiety could be a factor in changing category representations and controlling this factor among participants would prevent changes in prototypicality. As predicted, general anxiety nullified the face type x time interaction for contingent aware participants in the sorting task, from F (1, 17) = 11.61, p = .003, ηp2 = .41 to F (1, 16) = .05, p = .828, ηp2 = .00. General anxiety also nullified the face type x time interaction for contingent unaware participants in the sorting task, from F (1, 38) = 6.05, p = .019, ηp^2 = .14 to F (1, 37) = .05, p = .830, ηp2 = .00. This finding suggests that heightened anxiety at post-acquisition in general is a factor involved in shifting an outgroup exemplar's prototypicality towards the central tendency of the category in both contingent aware and unaware participants.

The number of extinction trials varied across participants because physiological anxiety took longer to extinguish in some participants than others (M = 13.5, SD = 4.36). I capitalised on the inter-individual variability and examined whether changes in prototypicality were mediated by the number of presentations viewed during extinction. To explore the role that number of extinction trials had for prototypicality shifts, I performed a mediational analysis with this factor. To determine the direction of these effects the difference of a difference of prototypicality was correlated with number of repeated presentations[3]. The calculation of this score suggests that as prototypicality shifts of the unsafe face increase, the greater number of extinction trials witnessed. A mediational analysis was performed by number of extinction trials as a covariate into a face type x time x type of extinction x 2 contingency awareness ANCOVA (Baron & Kenny, 1981; Judd et al., 2001; Yzerbyt et al., 2004). Number of extinction trials increased the significance of the face type x time x contingency awareness interaction, from F (1, 55) = 3.02, p = .088, ηp^2 = .05 to F (1, 54) = 5.62, p = .021, η_p^2 = .09. I followed this effect up by running the same analysis but separately for each level of contingency awareness.

I predicted number of extinction trials was a factor in the prototypical shift and entering it into the ANCOVA would cancel out or significantly reduce the face type x time interaction among contingent aware and unaware participants. As predicted, number of extinction trials nullified the face type x time interaction for contingent aware participants in the sorting task, from F (1, 17) = 11.61, p = .003, ηp2 = .41 to F (1, 16) = .71, p = .413, ηp2 = .04. Number of extinction trials also nullified the face type x time interaction for contingent unaware participants in the sorting task, from F (1, 38) = 6.05, p = .019, ηp^2 = .14 to F (1, 37) = 1.67, p = .204, ηp2 = .04. This finding suggests that number of extinction trials is a factor involved in shifting an outgroup exemplar's prototypicality towards the central tendency of the category in both contingent aware and unaware participants.

Together, mediational analysis suggests that contingent aware and unaware prototypicality shifts were mediated by anxiety in general at post-acquisition and number of extinction trials.

Discussion

Summary and Interpretation of Key Findings

The present research expanded on previous research and investigated the causal role extinction has in modifying prototypicality shifts of outgroup exemplars. Type of extinction was manipulated and one group of participants underwent the standard extinction procedure that repeatedly presented target exemplars. In the individuality group individuality and category membership cues were both present. The category membership group underwent a similar extinction procedure and viewed repeated presentations of scrambled images that contained the group membership cue with individual identifiable cues. Thus, I was in a position to determine whether individual identifiable markers had to be repeatedly presented, or whether repeated presentations of a category membership cues were sufficient to change an outgroup exemplar's fit with their category.

In this study I found a time main effect suggesting a non-associative prototypicality shift of faces involved in conditioning and of similar generalisation faces uninvolved in conditioning. The time main effect was not qualified by type of extinction, suggesting group membership cues could be presented to shift prototypicality and individuality cues were not required. Furthermore, I found differing results between contingent aware and unaware participants on the sorting task.

Participants who were contingent aware perceived the unsafe outgroup exemplar as being more prototypical of the outgroup post-extinction than the safe face. This effect was expected and replicated prototypicality shifts detected in Chapter 3. In contrast, contingent unaware participants perceived the safe outgroup exemplar as being more prototypical of the outgroup than the unsafe face post-extinction. In both contingent aware and unaware participants, non-associative prototypicality shifts did not differ between the individuality and category membership groups. Therefore, one conclusion is repeated presentations of category membership cues in the absence of individual identifiable markers after aversive conditioning was capable of shifting exemplar prototypicality.

The non-associative prototypicality shift observed in my larger sample (i.e., not factoring in contingency awareness) differs from results found in Chapter 3and could be partly due to two reasons. Firstly, in this study there were a larger number of contingent unaware participants (N = 40) than contingent aware participants (N = 19). This means the majority of my participants were not sure which face was associated with the electrical stimulation. Contingent unaware results suggest the safe face (vs unsafe face) was perceived as being more prototypical post-extinction, possibly due to participants associating the negative/anxiety provoking stimulus with the wrong face. In contrast, contingent aware participants correctly identified the unsafe face and it was perceived as being more prototypical of the outgroup post-extinction than the safe face. Thus, results suggest two contrasting prototypicality shifts between contingent aware and unaware participants. When participants are pooled together the two contrasting prototypicality shifts cancel out the difference found against the other face, resulting in a non-associative prototypicality shift.

Secondly, heightened self-reported anxiety post-extinction could have impacted the prototypicality measure. The non-significant face type x time x type of extinction and time x type of extinction interaction suggest my manipulation of extinction did not influence anxiety post-extinction and cannot explain this result. Schaller, Park and Muller (2003) found stereotypical thinking was activated when participants thought they were vulnerable to harm. In the context of my study, higher self-reported anxiety ratings suggest participants believed they were vulnerable to harm. In line with Schaller et al. (2003), participants could have activated stereotypical thought processes towards the two target faces presented during conditioning, resulting in larger prototypicality shifts for both faces. This explanation is consistent with generalisation data, which suggest prototypicality shifts generalised to faces that were similar to the targets used in conditioning, but not to new exemplars. Therefore, the differing prototypicality shifts between contingent aware and unaware participants and greater stereotypical processing as a result of persistent anxiety are two possible explanations of why non-associative prototypicality shifts were found in my larger sample in this study, but not in Chapter 3. Although non-associative prototypicality shifts were found, this effect was not qualified by type of extinction. This suggests repeated presentations of individuating information are not required and repeated presentations of group membership cues are sufficient. This effect is similar to that found in Study 3.3, with results suggesting that negativity/anxiety became associated with the skin colour rather than an exemplar and resulted in non-associative prototypicality shifts.

Contrasting results were found among contingent aware and unaware participants. Participants who were contingent aware perceived the unsafe outgroup exemplar as being more prototypical of the outgroup post-extinction than the safe, which was the expected effect. In contrast, the safe face was perceived as being more prototypical of the outgroup post-extinction than the unsafe face amongst contingent unaware participants. Importantly, contingent aware and unaware prototypicality shifts were not qualified by type of extinction. This means repeated presentations of individuality cues during extinction are not required in order for prototypicality shifts to occur. Rather, repeated presentations of group membership cues were sufficient to shift prototypicality. Therefore, when participants knew which faces were associated with the negative/anxiety provoking stimulus, repeated presentations of group membership cues were enough to change prototypical representations.

Contingent-specific prototypicality shifts found in contingent aware participants are consistent with research on familiarity and stereotyping (Garcia-Marques & Mackie, 2007; Smith et al., 2006). This research suggests that when faces are familiar (through repeated presentations), stereotypical processing occurs. Familiarity is expected to reduce the need to process information at an individual level and increase the reliance on category based processing. In my research contingent aware participants may have attended more to the unsafe face due to the electrical stimulation pairing. Research suggests there is an attentional bias for fear/anxiety related stimuli (van Bockstaele, Verschuere, Tibboel, De Houwer, Crombez & Koster, 2014). This means greater attention is directed to the unsafe face and results in greater familiarity and the ability to distinguish between the unsafe and safe faces. Presentations of the group membership cue during extinction could have reinforced category membership in the familiar-unsafe face, resulting in the contingent-specific prototypicality shift. This could also explain the shift of prototypicality in contingent unaware participants if the safe face was perceived as being more familiar. Therefore, a possible explanation of the effect I detected is certain faces were attended to more, increasing familiarity, and this coupled with reinforcement of category membership resulted in changes of prototypicality.

Limitations and Future Research Ideas

A limitation surrounding my familiarity based interpretation of results is I have no data to suggest familiarity increased. My interpretation is based on research that suggests similar faces were subject to greater stereotypical processing (Garcia-Marques & Mackie, 2007; Smith et al., 2006). Whilst my interpretation is supported by research, it should be tested empirically. To test my interpretation a self-reported measure of familiarity could be included into future research. This measure could be similar to my self-reported prototypicality and anxiety questions and participants asked to rate how familiar each face was perceived to be. This type of measure has been incorporated successfully into previous research (Carbon, Strobach, Langton, Harsanyi, Leder & Kovacs, 2007) and would provide an empirical basis to support my interpretation.

Another limitation of the current study is I am unable to pin point exactly what type of information needs to be repeatedly presented in order to shift exemplar prototypicality. I was able to eliminate repeated presentations of individuating information based on this study's results. Furthermore, results from this study suggest repeated presentations of category membership cues are sufficient to shift prototypicality. However, repeated membership cues are confounded with the simple passage of time. A passage of time might be needed to consolidate and strengthen the evaluative-fit/emotion-fit link between the exemplar and aversive stimulus. A passage of time effect is similar to sleep study research, which found greater learning after participants slept. It is believed that sleep helped to consolidate memory traces (Feld & Diekelmann, 2015; Landman et al., 2016; Rauchs et al., 2011), and a similar effect may be occurring in my study. Furthermore, a passage of time might indicate negativity/anxiety is spontaneously recovered (Bouton, 2002; Rescorla, 2004) and future research should account for this. A greater understanding of the stereotype development and maintenance process can be achieved if future research is able to ascertain what information was required to shift prototypicality.

To pinpoint exactly what occurs during extinction in order to shift prototypicality, the current study could be extended to include all possibilities. For example, in one group Black faces with individuality and group membership could be presented during extinction. Although this was done in the current research, replicating results with a larger designed study would ensure results are reliable. Similarly, another group could be repeatedly presented with category membership cues only, as this is the current minimum requirement identified by this study. Another group could be repeatedly presented with an unrelated stimulus, such as a landscape or faces irrelevant to the category to test the critical role that repeated presentations of any stimulus has. The last group will receive no presentations and time will simply pass to test the idea that a mere passage of time in needed. A study that includes all possibilities will be able to determine the minimum amount of information required to shift prototypicality.

Conclusion

In conclusion, the current research continued to investigate changes in an outgroup exemplar’s category representation after aversive conditioning by exploring the role extinction had in prototypicality shifts. In this chapter I showed repeated presentations of a category membership cue in the absence of individual identity markers were enough to change an exemplar’s fit with their group. Furthermore, we replicated previous results that suggested contingency awareness was a key factor in prototypicality shift. Future research should investigate the exact process behind prototypicality shifts and identify what information has to be presented during extinction. This will provide greater insight into the categorisation process and how stereotypes are developed and maintained.

Endnotes

1. A chi-square test of independence was performed to examine the relationship between contingency awareness and the face stimuli that were counterbalanced. The relationship between these two variables was non-significant, X^2 (1, N = 61) = 2.77, p = .096. This non-significant result suggests my post-hoc grouping of contingency awareness did not undo face stimuli counterbalancing.
2. Observable from the graph is different speeds to sort faces at pre-test, and this was checked to ensure prototypicality were not driven by differences at pre-test. A paired samples t-test suggested there was no difference in pre-test scores between the unsafe and safe face, t (18) = -1.52, p =.146.
3. The difference of a difference of prototypicality was calculated with the formula: (unsafe face at pre-test – unsafe face at post-extinction) – (safe face at pre-test – safe face at post-extinction). As quicker latencies represent greater prototypicality shifts, a positive correlation represents greater prototypicality shifts of the unsafe face (vs safe face) when absolute levels of anxiety were present.

References

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Carbon, C. C., Strobach, T., Langton, S. R. H. Harsanyi, G., Leder, H., & Kovacs, G. (2007). Adaptation effects of highly familiar faces: Immediate and long lasting. *Memory & Cognition*, *35*(8), 1966-1976.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science*, *280*, 77–81.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Feld, G. B., & Diekelmann, S. (2015). Sleep smart – optimizing sleep for declarative learning and memory. *Frontiers in Psychology, 6.* A622.

Fiske, S., & Neuberg, S. L. (1990). A continuum of impression formation form category-based to individuating processes: Influences of information and motivation on attention and interpretation. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol.23, pp. 1–74). San Diego: Academic Press.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18*(3), 232-239.

Garcia-Marques, T., & Mackie, D. M. (2007). Familiarity impacts person perception. *European Journal of Social Psychology, 37,* 839-855.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*(3), 390–421.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological methods, 6* (2), 115-134.

Landman, N., Kuhn, M., Maier, J. G., Feige, B., Spiegelhalder, K., Riemann, D., & Nissen, C. (2016). Sleep strengthen but does not reorganize memory traces in a verbal creativity task. *SLEEP*, *39*(3), 705-713.

Lovibond, P. F., Saunders, J. C., Weidemann, G., & Mitchell, C. J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *The Quarterly Journal of Experimental Psychology*, *61*(8), 1199–1216.

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology*, *45*(6), 1303–1307.

Olson, M. A, & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science, 309*(5735), 785–787.

Page, M. M. (1973). On detecting demand awareness by postexperimental questionnaire. *The Journal of Social Psychology, 91*(2), 305–323.

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *The Journal of Neuroscience, 31*(7), 2563-2568.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning and Memory, 11*, 501-509.

Richeson, J. A, & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology*, *89*(4), 517–530.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis)categorization of unattractive brides and attractive prostitutes. *Experimental Psychology*, *55*(3), 182–188.

Schaller, M., Park, J. H., & Muller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin, 29*(5), 637-649.

Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., & Mackie, D. M. (2006). Familiarity can increase stereotyping. *Journal of Experimental Social Psychology, 42*, 471-478.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of social conflict. In W. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp.33-47). California: Brooks/Cole.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin, 140*(3), 682-721.

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science, 326*(5960), 1711–1714.

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology, 40*(3), 424-431.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27.

Chapter 5: Aversive Learning Causes Perceptual Shifts in Perceived Ingroup Prototypicality

In the previous two chapters, I expanded on the affective consequences the association between an outgroup exemplar and negativity/anxiety had on an outgroup exemplar's categorical representativeness. This was investigated with an aversive conditioning procedure, where one outgroup exemplar was paired with negativity/anxiety and another outgroup exemplar was never implicated in the negativity/anxiety pairing. An extinction procedure followed conditioning and exemplars were repeatedly presented in the absence of negativity/anxiety in order to make the association ambiguous. After extinction, I found exemplars' perceived prototypicality, a well-established marker of the social categorisation process (Corneille & Judd, 1999; Rosch, 1978), shifted and the exemplar associated with negativity/anxiety was perceived as being a better fit with the outgroup.

In this chapter, I continue to investigate shifts in cognitive representations in terms of categorisation judgments and decisions through changes in perceived prototypicality. However, in this chapter I direct my attention from changes in outgroup prototypicality to changes in ingroup prototypicality. Through this research I will contribute to our understanding of intragroup dynamics, with implications for how ingroup stereotypes are formed and maintained. To begin with, I first sought to explain why ingroup-outgroup perceptions might shift through the lenses of different theories.

Theories of Ingroup-Outgroup Perception

Several research traditions discuss the psychological significance of valenced and emotion-loaded ingroup-outgroup perception and suggest possible individual responses to changes in the status quo. Social psychological theories such as self-categorisation theory are focused on ingroup-outgroup evaluations and how different group members are perceived. From this theory's perspective, individuals are motivated to protect the positive valence associated with the ingroup because their self-esteem is partially derived from membership within that group (Oakes, Haslam & Turner, 1994; Oakes, Turner & Haslam, 1991; Rubin & Hewstone, 1998; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). To achieve or maintain high levels of self-esteem, the individual strives to establish the psychological superiority of the ingroup over the outgroup. This effect is in line with an evaluative-fit mechanism, where the ingroup is associated with positive valence and the outgroup typically with negative valence (Coats, Latu & Haydel, 2007; Harwood et al., 2017).

Research into the "black sheep effect" is consistent with the premises of social psychological theories such as self-categorisation theory and an evaluative-fit mechanisms (see Marques & Paez, 1994; Marques, Yzerbyt & Leyens, 1988; Pinto, Marques, Levine & Abrams, 2010). This research area focuses on changes in exemplar evaluations when a negative ingroup member threatens the positivity of the ingroup. Results consistently show that ingroup exemplars associated with negativity are judged and rated more harshly than comparatively unlikeable outgroup exemplars. In contrast, ingroup exemplars associated with positivity are treated more favourably than comparatively likeable outgroup exemplars. The rejection of negative ingroup exemplars and acceptance of likeable ingroup exemplars aligns well with the self-categorisation theory's motivational stance around valenced ingroup-outgroup perceptions that guarantee ingroup's positive distinctiveness in society: Positive ingroup members are accepted as part of the ingroup, whilst negative ingroup members are rejected from the group because their valence compromises ingroup's positive distinctiveness.

Evolutionary theories such as the sociofunctional approach provide an alternate perspective to understanding ingroup-outgroup perceptions (Cottrell & Neuberg, 2005; Dasgupta, DeSteno, Williams & Hunsinger, 2009; Mineka & Ohman, 2002; Seligman, 1971). Evolutionary based psychological theories focus on the role threat and anxiety has in group perceptions. From this theoretical perspective, outgroup members signal a threat to ingroup members because in our species' evolutionary past, outgroup members competed for food and territory, which led to violent clashes and death (Van Vugt & Park, 2010). Through societal and cultural evolution, outgroups became defined by competition, anxiety and threat, whilst ingroups became defined by cooperation and safety.

Evolutionary-based frameworks are in line with research showing that category based judgments increase when outgroup applicable emotions are triggered (i.e., emotion-fit; see Cottrell & Neuberg, 2005; Dasgupta et al., 2009; DeSteno, Dasgupta, Bartlett & Cajdric, 2004). For example, Dasgupta et al. (2009) showed anger and disgust, two intergroup emotions that signal threat, created bias towards the Arabs outgroup and the homosexuals outgroup but only when the emotion applied to, or "fit", pre-existing beliefs about the outgroup. Hence, anger caused bias towards Arabs but not homosexuals, whilst disgust caused bias towards homosexuals, but not Arabs. This effect is in line with an emotion-fit mechanisms, where specific emotions provide a better fit with a specific outgroup based on the type of evolutionary threat that group traditionally posed or are still seen to pose (Cottrell & Neuberg, 2005; Dasgupta et al., 2009).

Consistent with these broad evolutionary theories is research on associative learning (Esteves, Parra, Dimberg & Ohman, 1994; Flykt, Esteves & Ohman, 2007; Olson & Fazio, 2006; Olsson, Ebert, Banaji & Phelps, 2005). This research has repeatedly shown that stimuli from fear relevant categories (prepared stimuli; e.g., snakes/outgroup members) produced larger anxiety learning compared to stimuli from fear irrelevant categories (unprepared stimuli; e.g., butterflies/ingroup members). For example, Olsson et al. (2005) recruited White participant and presented White and Black exemplars on a screen. One White and Black exemplar was paired with an electric shock, whilst another White and Black exemplar never received the pairing. They found the outgroup exemplar associated with the aversive stimulus resisted extinction of anxiety whilst the ingroup exemplar associated with the aversive stimulus fully extinguished. Thus, within an associative learning framework, this result suggests that in the eyes of White participants in contemporary Western societies, there is an intrinsic and inherent link between anxiety and the Black outgroup.

Associative learning research has been extended from fear relevant outgroup members to outgroup members who are arbitrarily defined with no potential for previous contact. For example Navarrete et al. (2012) distinguished groups using a minimal group paradigm where participants completed a bogus perceptual task and randomly assigned to a group, which was distinguished by the colour of a t-shirt. Participants underwent conditioning and one ingroup (same colour t-shirt) and outgroup member (different colour t-shirt) were paired with an electric shock, whilst another ingroup and outgroup exemplar never received the pairing. Unlike Olsson et al. (2005), Navarrete et al. (2012) found preferential learning towards the outgroup exemplar paired with an electrical stimulation compared to the ingroup exemplar during acquisition rather than extinction. Navarrete et al. (2012) explain greater anxiety learning during acquisition (compared to extinction) from an evolutionary framework through the need to learn quickly when limited information is available. This again demonstrates that within an associative learning framework there is an intrinsic and inherent link between anxiety and the outgroup.

Here, I argue that, while social psychological theories focus on the ingroup perception and evolutionary theories focus on outgroup perception, these two theories converge and make similar predictions about how valenced and emotion-loaded ingroup-outgroup perceptions are maintained. Brewer (1999) provided an influential analysis of the link between ingroup-outgroup judgments. An ingroup is viewed positively because one's self-esteem, safety and likelihood of survival are linked to the individual's own group. Brewer noted that attachment to the ingroup does not necessarily lead to outgroup hate. However, attachment to the ingroup can help to foster negativity/fear/avoidance of the outgroup simply because the two groups do not share the same ideals and values or provide the same level of safety for the individual. Thus, mechanisms behind 'ingroup love' would help create a sense of fear towards anyone outside the ingroup—i.e., the outgroup—or at minimum a sense of reduced attraction and psychological engagement.

Research thus far has focused on the evaluation implications of these dynamics for ingroups, but comparatively neglected implications for perceived *ingroup* categorisation shifts. In this research I manipulate the administration of negativity/threat and investigate how categorisation of ingroup exemplars changes. I start by providing a closer look at social categorisation from an *intergroup* (vs. outgroup-only perspective).

Ingroup-Outgroup Dynamics and Categorisation: Theories, Predictions, and Past Research

This work starts from the premise of modal valenced group perceptions. I previously pointed out that the ingroup is typically viewed as psychologically superior to the outgroup because one's self-esteem is boosted from membership within a group that has high standing and survival chances are enhanced. Thus, the ingroup is traditionally viewed through the lenses of positive stereotypes and associated with positive emotions (Fiske, Cuddy, Glick & Xu, 2002; Mackie & Smith, 1998). In contrast, the outgroup is viewed as carrying traditionally negative stereotypes and emotions. Social categorisation activates stereotypes as people are sorted into groups based on similar characteristics and are perceived with the same cognitive representation (Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000; Schneider, 1991). Social categorisation guides group formation and is involved in the accentuation of perceived ingroup similarities and outgroup differences (Bruner, 1957, Corneille & Judd, 1999; Tajfel & Wilkes, 1963). Thus, it serves as a mechanism to maximise the perceived differences between groups to ensure they remain perceptually distinct from each other (Hogg, Abrams, Otten, & Hinkle, 2004). As such, social categorisation is an important process in developing and maintaining group perception.

An important factor in maintaining vs. shifting group perception is the degree to which an ingroup member possesses ingroup like qualities. If an ingroup member behaves in a way that contradicts the ingroup prototype, group perception will shift and that member will be distanced from the group. In this way, categorisation, like evaluations, are an important process in forming and maintaining group perceptions.

Exemplar's inclusion in (or exclusion from) a social group therefore depends on how prototypical an individual is perceived to be of that group (Bruner, 1957; Locke, Macrae & Eaton, 2005; Medin & Smith, 1981; Oakes et al., 1994; Smith & Zarate, 1990). To psychologically include a person into a specific group, they need to have a high level of perceived fit with the group's central tendency. For example, a person would be included in the Black category if they possessed dark skin, a big nose and thick lips; and excluded from the Black category if they possessed lighter skin, a smaller nose and thinner lips.

As highly prototypical group members are closely aligned to the group's central tendency, they are associated with the thoughts and emotions most linked with their group, more than less prototypical members are—i.e., they are stereotyped more (Blair, Judd & Chapleau, 2004; Blair, Judd, Sadler & Jenkins, 2002). For example, Blair et al. (2004) research demonstrated inmates with more Afrocentric features received harsher sentences in the United States compared to inmates with less Afrocentric features for similar crimes. Thus, factors capable of shifting perceived exemplar prototypicality have the potential to shift group perceptions and the application of associated stereotypes.

In this work, the direction of the prototypicality shifts expected to follow aversive conditioning is derived from evaluative-fit and emotion-fit mechanisms and from the premise that the ingroup is associated with positive evaluations/safety emotions and the outgroup with negative evaluations/threat-related emotions. I expect ingroup exemplars paired with an aversive stimulus to be perceived as negative/threatening and, thus, fit less with the ingroup representation. The ingroup's perception will be protected by psychologically or cognitively excluding members that threaten the positivity and safety the ingroup represents. As a consequence, negative/anxiety provoking exemplars will be perceived as less prototypicality of the ingroup after aversive conditioning and shift away from the ingroup prototype and closer to the outgroup prototype. In contrast, I expected ingroup exemplars not associated with negativity/anxiety to be perceived as positive/non-threatening and, thus, to fit more with the ingroup representation. Therefore, positive/safe exemplars should be perceived as more prototypical of the ingroup and shift closer to the ingroup prototype after aversive conditioning and away from the outgroup prototype because negativity/anxiety is not an association normally given to the ingroup.

Consistent with my proposed hypotheses is previous research that investigated the perceived fit of valenced exemplars to a group. Richeson and Trawalter (2005) investigated the effects exemplar valence had on exemplars' group membership. In their studies, White participants viewed images of various White and Black famous people. Famous individuals were chosen because they are known exemplars with pre-existing evaluations. For example, John F Kennedy was a part of the positively valenced White targets because he was well liked President. In contrast, Adolf Hitler was a part of the negatively valenced White targets because of his war crimes. Richeson and Trawalter found that White individuals were faster and more accurate at sorting faces into the "White" category if they were an admired famous White person, as opposed to disliked famous White people. In contrast, disliked Black faces were sorted quicker and more accurately into the "Black" category compared to admired Black faces. Hence, in fitting with social psychological theories (i.e., self-categorisation theory), admired White exemplars were treated as being more prototypical of the ingroup, or more *ingroup-like*, than disliked White exemplars. On the other hand, disliked Black exemplars were treated as being more prototypical of the outgroup, or more *outgroup-like*. By excluding negative/unsafe exemplars from the ingroup representation—via speeded categorisation decisions—this research supports Brewer (1999) idea that the ingroup's positive representation is protected through the categorisation process.

Similar results, with a focus on emotions rather than mere valence, were found in Miller, Maner and Becker's (2010) research. In a series of experiments, threat-related factors (e.g., masculinity, anger expressions etc.) increased the likelihood that an exemplar would be categorised into the outgroup (vs. ingroup) category. For example, in their final study participants were allocated into minimal groups based on a perceptual judgment task. Once allocated to a minimal group participants listened to different voices that were unaltered or altered to sound more masculine. Participants who believed in a dangerous (threatening) world categorised masculine voices into the outgroup category more than the unaltered voices. Thus, when perceived threat was high, ambiguous/unknown exemplars were judged to be a threat and considered as being more outgroup-like among individuals more inclined to associate dangers with 'otherness'. In line with the evolutionary theories of emotion functionality, and Brewer's (1999) analysis, these results suggest group perceptions and categorisation decisions can be used to exclude and protect the ingroup from threatening exemplars.

While theoretically sound, data from the previous research failed to investigate ingroup categorisations, changes in perceived fit across time and contingency-related effects—these are the focus of my research.

Research Gaps and the Present Research's Methods

The research I conducted and reported in the early chapters has the merit of investigating the temporal dimensions of outgroup categorisation but failed to look at *in*group categorisation and *in*group stereotype formation. In line with mechanisms of valence- and emotion-fit, I typically found that an outgroup exemplar paired with an aversive stimulus was perceived to be closer to the outgroup central tendency after the aversive pairing compared to beforehand. In this research I expanded my focus to investigate temporal dimensions in the categorisation of ingroup exemplars as a function of superimposed negativity/anxiety.

Investigations into how perceived prototypicality shifts across time are important because they can help us understand how stereotypes evolve and change. Exemplar valence in the previous research was operationalised by internal factors, such as individual targets' reputation, history and previous action. For example, Richeson and Trawalter (2005) chose well known exemplars that were liked or disliked. A problem with evaluative responses generated by this type of exemplar valence is that they cannot change easily and are not suitable to assess changes over time. The research methodology used in the current research is similar to that used in the previous chapters: I experimentally *superimposed* valence/anxiety onto to group exemplars through a conditioning procedure. An aversive stimulus was paired with one exemplar, but not another. I measured for exemplars' perceived prototypicality before and after conditioning to assess changes over time. I also measured the perceived prototypicality of configurally related and new exemplars to determine whether prototypicality shifts generalise to other exemplars.

The conditioning procedure used to superimpose valence allowed contingency effects to be investigated. I can superimpose negativity/anxiety on one exemplar and leave another exemplar free from the superimposed valence/anxiety. This approach can determine with precision whether prototypicality changes occur *selectively* in the specific paired exemplar, but *exempt* the unpaired exemplar by virtue of the controlled superimposed valence/threat (aka. an associative process) or rather whether the changes indiscriminately affect all exemplars (aka. a non-associative process). Investigations into contingency-related effects are important because they can clarify the breadth of the psychological consequences of individuals' experiences with a limited number of group exemplars on new exemplars and group-level responses (for an extensive discussion, see Paolini, Harris, & Griffin, 2016). Understanding how exemplar perceptions develop and change—along contingency-related or non-associative-related effects and within a design that includes both the ingroup and the outgroup—can ultimately shed light on when and how negative stereotypes are applied and might be differentially applied to certain group members more than others.

Study 5.1

In this study, the perceived prototypicality of unambiguous ingroup faces of neutral expression was measured prior to, and after, the faces were paired with an aversive stimulus. Outgroup exemplars, as part of a within-subject design, were also incorporated into this research design as a way of benchmarking new findings against our earlier results. This means there is an unsafe and safe exemplar for both the ingroup and outgroup. Generalisation stimuli were also included to determine the extent to which prototypicality shifts generalise to other exemplars.

This study implemented an aversive conditioning procedure in the context of a minimal group paradigm (Tajfel, Billig, Bundy & Flament, 1971). In minimal group studies, laboratory created groups that do not exist in real life are used and found to mimic the psychological processes activated by real groups (Deffenbacher, Park, Judd & Correll, 2009; Dobbs & Crano, 2001). I began this research with minimal groups for the potential offered by this method to control for prior familiarity with the groups and associated stereotypes and, thus, to investigate valence associated with minimal groups away from intrinsic ingroup-outgroup contexts. Hence, through this method, a neater test of exemplar's ingroup-outgroup categorisation was pursued and clearer test of the effect an aversive pairing has on perceived goodness of fit between the exemplar and their group was made possible.

Given the ingroup-outgroup nature of the design, the prototypicality data in this study was recoded in a relational way: lower values on the prototypicality index indicated the face being perceived as being more prototypical of the ingroup (and less prototypical of the outgroup); higher values indicated the face being perceived as being more prototypical of the outgroup (and less prototypicality of the ingroup). Therefore, I expected both the ingroup unsafe face and outgroup unsafe face to have higher values at post-test because this represented a shift *away* from the ingroup prototype and *towards* the outgroup prototype. In contrast, I expected both the ingroup safe face and outgroup safe face to have lower values at post-test because this represented a shift towards the ingroup prototype and away from the outgroup prototype. This pattern was expected to result in a significant face type x time interaction that is not qualified by target group.

Although our main hypotheses do not suggest an ingroup-outgroup asymmetry in prototypicality shifts, different asymmetries are potentially implicated by explanations that invoke an evaluative-fit vs. emotion-fit mechanism. Evaluative-fit explanations in social psychological theories (i.e., self-categorisation theory) have a focus on ingroup dynamics and thus might imply larger shifts in prototypicality towards ingroup stimuli, than outgroup stimuli. Emotion-fit explanations in evolutionary theory-inspired frameworks (i.e., functional emotion theories) have a focus on outgroup perception and thus might imply larger shifts in prototypicality towards outgroup stimuli, than ingroup stimuli. I will check these possibilities by inspecting the higher order interaction of face type, time, and group type.

Method

Participants and Design

Participants were 57 students (21 male, 36 female; *M* = 22.98 years, *SD* = 6.83) from a large regional Australian university. All participants reported being from a White, Anglo-Saxon background. Three participants were excluded from the study because they did not correctly identify the colour of the group that they were assigned to, which left 54 valid participants. Participants received a small monetary compensation ($25) or partial course credit for their participation. Participants were randomly assigned to one of the two levels of the participant group membership factor; they were assigned to either belong to a green under-estimator group or to a blue over-estimator group, as part of a 2 post-test position (after acquisition, after extinction) x 2 target group (ingroup/outgroup) x 2 face type (unsafe/safe face) x 2 time (pre-/post-test) mixed model design with target group, face type and time as the repeated measures. Twenty nine participants were allocated to the after acquisition condition (14 green under-estimator, 15 blue over-estimator) and 25 participants to the after extinction condition (14 green under-estimator, 11 blue over-estimator). There were 28 participants randomly allocated as belonging to the green under-estimator group and 26 participants to the blue over-estimator group[1].

Materials

A subset of the faces and the variations (and the colour palette used) used in this study are observable in Figure 17. Four faces of Caucasian male adults of neutral expression and frontal orientation were chosen from the Radboud Faces Database as stimulus material for all participants (Langner, Dotsch, Bijlstra, Wigboldus, Hawk & Knippenberg, 2010). Four *target faces* were presented with either a blue or green background so they could be identified as belonging to the ‘blue over-estimator’ or 'green under-estimator' group. One blue over-estimator and one green under-estimator face were paired with an aversive stimulus; a different blue over-estimator and green under-estimator face were never paired with the aversive stimulus. The four target faces were counterbalanced across participants so that each face was presented as an over or under-estimator (i.e. either an ingroup or outgroup member) and was either presented with or without the aversive stimulus.

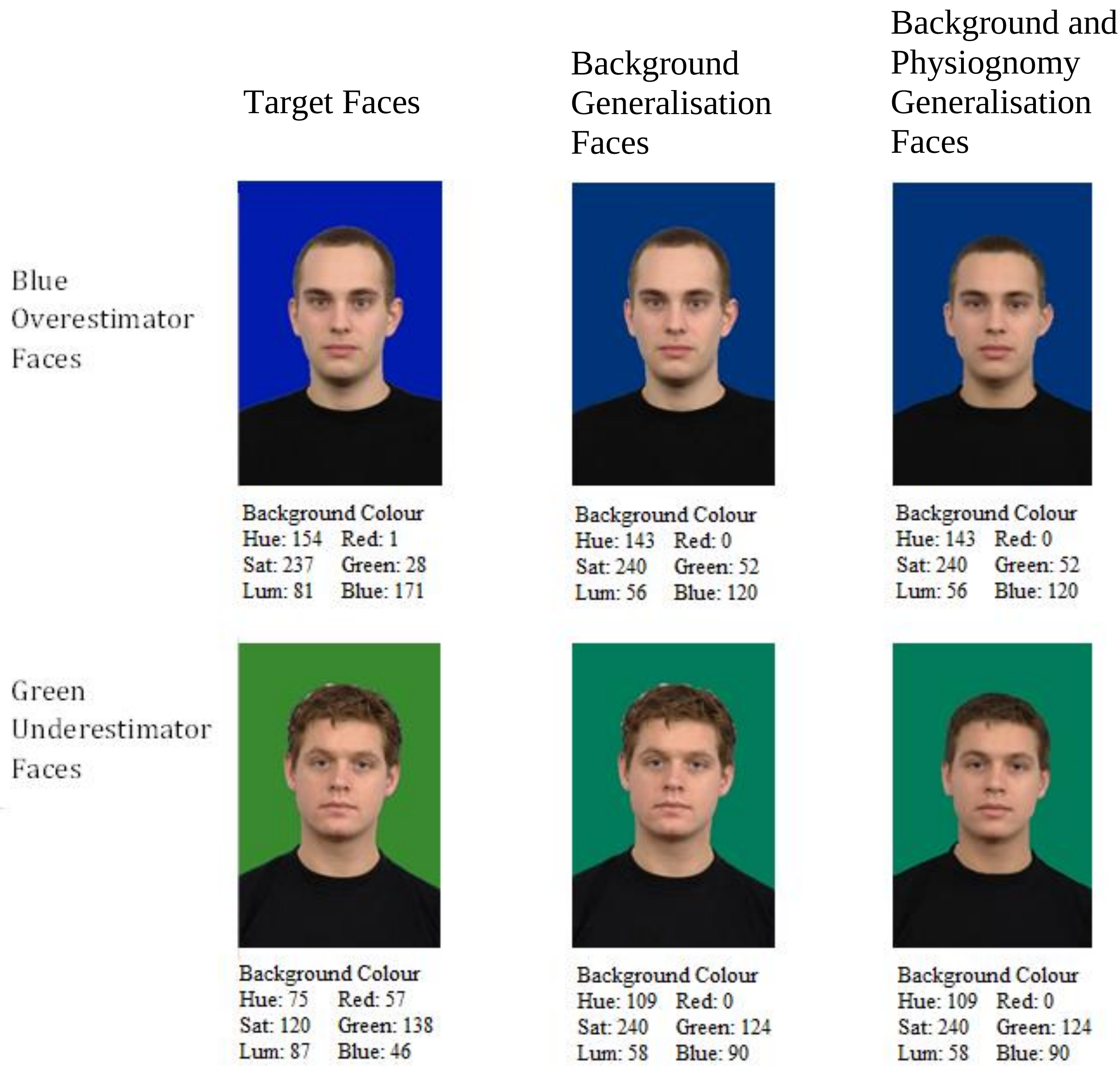


Figure 17. Subset of the target, background generalisation and physiognomy and background generalisation faces used in Study 5.1.

The four target faces were altered to create an additional four faces, referred to as *background generalisation faces*. In order to investigate the influence that the group membership identifier – the background colour – had on prototypicality shifts, I manipulated the background colour presented with the background generalisation faces. Under-estimators' background generalisation faces had their background colour changed to appear less green (and bluer). Over-estimators' background generalisation faces had their background colour changed to appear less blue (and greener). Hence, four new background generalisation faces were created that maintained the same target exemplar, but were presented with a changed background colour. The background colour was changed to fall in between the blue-green extreme (see background colour settings in Figure 17).

The four background generalisation faces were altered to create an additional four faces, referred to as *physiognomy-and-background generalisation faces*. In order to investigate the influence that the facial physiognomy had on generalised prototypicality shifts, background generalisation faces were also morphed using Facegen, in a way that varied their facial features to appear less Eurocentric and towards an average of all Facegen faces (this included Caucasian, Black-African, Asian and Middle Eastern faces). Hence, four new generalisation faces were created that maintained the background colour used for generalisation faces, but changed the facial appearance of the target stimuli. Therefore each of the four *target faces* had two types of generalisations faces — *background generalisation faces* and *physiognomy-and-background generalisation faces* — making a total of 12 faces.

Procedure and Measures

To begin the study participants learnt about bogus research and group differences in estimation patterns. They learnt that approximately half the population were under-estimators of stimuli in the environment, whilst the other half were over-estimators. Participants were informed that in this study group membership would be represented with two different colours. Over-estimators were represented with a blue background and under-estimators with a green background. An example of the faces and their group membership was displayed on the computer screen. The two estimation groups were presented as the extreme points at the polar ends of a bipolar continuum and individuals could fall anywhere along the two extremes of the continuum. No other information about the estimation groups were provided in order to prevent any unexpected valenced perceptions of the two groups.

A bogus dot estimation task followed, where participants had to guess the amount of dots on ten subsequent screens. A set of five slides with 60, 70, 80, 90 and 100 dots were presented in a random order twice each for three seconds on the computer screen. After each slide was presented, the participant was asked “how many dots did you see in the slide”, with response options ranging between 40 and 120. Participants were led to believe that the computer analysed their responses and a feedback screen identified them as either a blue over-estimator or a green under-estimator. Instead, the task results were bogus and participants were randomly allocated into either the green under-estimator group or the blue over-estimator group. Participants were fitted with coloured wrist bands during the remainder of the study to reinforce and support the group membership manipulation. The group they were assigned to was their (lab-created) ingroup and the group not assigned to became their (lab-created) outgroup.

Data was collected in a single laboratory session. Pre-test self-reported data was collected as part of an online questionnaire after the dot estimation task. Participants indicated how prototypical they perceived the 12 faces (target and generalisation faces rated; randomly ordered) to be of the estimation groups using a 7 point Likert scale, with 1 = *prototypical of "Green Under-estimator* and 7 = *prototypical of "Blue" Over-estimator*. Because of the relational nature of this scale, lower ratings on the scale indicated the face was perceived to be more prototypical of the under-estimator group (than the over-estimator group) and higher ratings indicated the face was perceived to be more prototypical of the over-estimator group (than the under-estimator group). Self-reported anxiety was collected as part of a manipulation check to ensure acquisition and extinction were effective. The same feeling thermometer measures described in Study 3.3 were administered to ensure the standard minimal group bias developed. Feeling thermometer data was collected for blue over-estimators first and then for green under-estimators[2].

Following self-reported data collection, participants underwent a work up procedure. The work up procedure was similar to the methods used in the previous chapters, where participants selected the intensity of electrical stimulation that they regarded as uncomfortable but not painful to be used during acquisition (Lovibond, Saunders, Weidemann, & Mitchell, 2008). The electrical stimulation was administered via a Powerlab 4/25T inbuilt isolated stimulator (ADInstruments) using a bar electrode attached to the participant's right forearm. Skin conductance electrodes were attached to the distal phalanges of the first and second digits on the participant's left hand and measured via an ADInstruments Model ML116 GSR amplifier using standard MLT116F electrodes. Respiration was monitored to check for artefacts using an ADInstruments MLT1132 Piezo respiration belt attached around the chest.

Pre-test skin conductance responses (SCRs) were measured prior to acquisition and post-test SCRs immediately following acquisition and before extinction. During acquisition, target faces were presented in the centre of the screen for 10 s (inter-stimulus interval M = 17.5 s, range 15-20 s) in a randomised order. Two targets, one ingroup and one outgroup face (unsafe faces) always co-terminated with a 200 ms electrical stimulation at the level selected during the work up procedure. Two different target faces, one ingroup and one outgroup face (safe faces; counterbalanced), were never paired with a stimulation.

Similar to Study 3.2, post-test position was manipulated in this study and self-reported post-test measures were collected after acquisition (but before extinction) or after extinction to continue exploring the role of extinction. Extinction consisted of the four target faces (unsafe and safe faces) being presented (always an identical number of times) in the absence of any electrical stimulation until no further reduction in physiological activation was observed. Post-test self-reported data collection used the same measures described at pre-test with addition of contingency awareness questions. Contingency awareness was measured and checked for using the same procedure described in Study 3.2. Briefly, these questions involved asking participants if they reacted the same way to all faces, whether they noticed a pattern in which a face was paired with an electrical stimulation, which face was paired with the electrical stimulation and how confident were they in their decision (adapted from Clark & Squire, 1998; Page, 1973; see Appendix O). In order to be considered contingent aware participants needed to accurately select the face paired with shock, with a high degree of confidence (i.e., On a scale 1 -7, a rating of 4 or more on how confident they were the face they selected was paired with the electrical stimulation). Participants also needed to accurately describe the association between the exemplar and electrical stimulation in an open-ended response. Eighteen participants were classified as contingent non-aware (9 after acquisition group; 9 after extinction group) and 36 as contingent aware (20 after acquisition group; 16 after extinction group).

Results and Discussion

Checking for Standard Minimal Group Bias

Minimal group paradigms are used readily in intergroup research because the same ingroup favouritism effects that occur in real groups are typically produced in contexts of trivial and relatively contrived group categorisations and away from the confounds of group familiarity and stereotypes. I tested for a basic intergroup bias effect with a 2 target group (ingroup and outgroup) x 2 post-test position (before or after extinction) x 2 participant's group membership (under-estimator or over-estimator) mixed model ANOVA using participants' pre-test feeling thermometer ratings for the ingroup and outgroup, with target group as the repeated measure. A main effect of target group was detected, F (1, 50) = 5.45, p = .024, ηp^2 = .10. A paired samples *t*-test confirmed the ingroup (M = 60.00, SD = 20.47) was liked more than the outgroup (M = 50.56, SD = 18.06), t (53) = 2.33, p = .023. All other effects were non-significant and ingroup favouritism did not vary between groups, all *p's* > .134. Thus, as expected in standard minimal group paradigms, valence about the groups was not provided through the minimal group cover story; it was instead subjectively superimposed by the participants as a result of mere ingroup-outgroup categorisation dynamics.

Checking Effective Acquisition and Extinction

To determine if effective acquisition and extinction occurred, I first examined participant's skin conductance responses (SCRs). SCRs were recorded and scored following standard guidelines that were described in Study 3.1 (Boucsein et al., 2012; Fowles, Christie, Edelberg, Grings, Lykken & Venables, 1981). Briefly, to measure effective acquisition, first interval SCRs with a minimum amplitude 0.05 μS were calculated and averaged across two presentations for each of the two target faces immediately prior to, and after acquisition. To measure extinction, the same process was used but scores were calculated for the first block of extinction (first two presentations of the unsafe and safe face respectively) and the ten trial block (the last two presentations of the unsafe and safe face respectively up to the tenth trial – minimum number of trials all participants received to satisfy extinction criteria).

To check for effective acquisition, I performed a 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition vs after extinction) x 2 target group (ingroup and outgroup) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with target group, face type and time as repeated measures on SCR data. A significant face type x time interaction was found, F (1, 49) = 7.87, p = .007, η_p^2 = .14. A paired samples t-test indicated the unsafe face had higher levels of SCR activity at post-test (M = 1.10, SD = .19) compared to pre-test (M = 1.04, SD = .07), t (52) = -2.94, p = .005. In contrast, the safe face had similar levels of SCR activity at post-test (M = 1.02, SD = .07) and pre-test (M = 1.04, SD = .09), t (52) = 1.09, p = .281. This suggests conditioned anxiety was higher for the unsafe faces than the safe faces.

The 2 way interaction was further qualified by a 3 way interaction involving target group, face type and time, F (1,49) = 16.51, p < .001, η_p^2 = .25. When following up the 3 way interaction effects were driven by the outgroup unsafe face. The outgroup unsafe face had higher levels of SCR activity at post-test (M = 1.17, SD = .28) than at pre-test (M = 1.03, SD = .10), t (52) = -4.17, p = <.001. This result suggests greater anxiety learning towards the outgroup unsafe face than the ingroup unsafe face.

To check for effective extinction of the association, I ran a 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition vs after extinction) x 2 target group (ingroup and outgroup) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with target group, face type and extinction block as repeated measures on SCR data. SCR data during the first block of extinction (first two presentations of the unsafe and safe face) and the ten trial block (the last two presentations of the unsafe and safe face up to the tenth trial) was chosen as all participants were exposed to a minimum of ten extinction trials. As expected the face type x extinction block interaction F (1, 47) = 3.31, p = .075, η_p^2 = .06, and target group x face type x time interaction were non-significant, F (1, 47) = .003, p = .995, η_p^2 = .00. These results provide evidence for effective extinction of anxiety. Supplemental data analyses conducted on SCR suggests negativity/anxiety extinguished and are discussed extensively in Appendix P5. I also checked for effective acquisition and extinction with self-reported anxiety data. These data suggests that whilst effective acquisition occurred, self-reported anxiety persisted beyond extinction. Self-reported anxiety data are reported in Appendix P5 for brevity.

Testing Basic and Generalised Prototypicality Shifts

In this study, the self-reported measure was anchored with prototypicality ratings for over-estimators and under-estimators, resulting in ingroup and outgroup responses as a function of participants' group membership assignment. To account for this design, I reverse scored the over-estimator data. Therefore, for all participants—irrespective of which group (blue vs. green) the participant belonged to—higher values indicate the face was more prototypical of the outgroup, and lower values indicate the face was more prototypical of the ingroup. I expected the ingroup and outgroup unsafe face to have higher prototypicality ratings at post-test, indicative of a shift *away* from the ingroup prototype and *towards* the outgroup prototype. In contrast, I expected the safe ingroup and outgroup face to have lower values at post-test than pre-test, indicative of a shift *towards* the ingroup prototype and *away* from the outgroup prototype.

Basic Prototypicality shifts were tested using a 2 face type (unsafe and safe) x 2 time (pre- and post-test) x 2 target group (ingroup and outgroup) x 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition, after extinction) mixed model ANOVA, with target group, face type and time as the repeated measures. As a result of the recoded data and the relational nature of our prototypicality measure, an interaction between face type and time was expected that was not qualified by target group. I expected both the ingroup and outgroup *unsafe* face ratings to move in the same direction but this to represent different kinds of representation shifts (i.e., a shift *away* from the ingroup prototype and a shift *towards* the outgroup prototype). Similarly, the ingroup and outgroup *safe* faces were expected to move in the same direction but represent different types of shifts (i.e., a shift *towards* the ingroup prototype and a shift *away* from the outgroup prototype).

The manipulation of ingroup and outgroup face assignment was successful, as a main effect of target group showed ingroup faces were perceived more ingroup-like (M = 3.35, SD = 1.24) and outgroup faces perceived more outgroup-like (M = 4.72, SD = 1.22), F (1, 50) = 19.14, p < .001, ηp^2 = .28. More importantly, a face type x time interaction was detected and can be observed in Figure 18, F (1, 50) = 6.23, p = .016, ηp^2 = .11. As expected, the safe faces mean was rated with a lower value at post-test (M = 3.85, SD = 1.43), compared to pre-test (M = 4.26, SD = 1.51), t (53) = 2.79, p = .007. This result showed the safe faces, irrespective of whether they belong to the ingroup or outgroup, were perceived to be more ingroup-like and less outgroup-like after conditioning than beforehand. Unexpectedly, prototypicality shifts of the unsafe face were not detected and ratings were similar at post-test (M = 4.06, SD = 1.23) and pre-test (M = 3.98, SD = 1.32), t (53) = -.80, p = .429. These results suggest safe faces (vs unsafe faces) are driving the interaction and were perceived to be more ingroup-like and less outgroup-like after conditioning than beforehand.

Unlike the earlier studies, this time the face type x time x post-test position interaction was non-significant, F (1, 50) = .55, p = .460, ηp^2 = .01. In Study 3.1 and 3.2 prototypicality shifts were detected following the extinction procedure. Hence, contrary to the earlier studies, this result suggests that prototypicality shifts for lab-created groups were of the same magnitude when collected before or after anxiety extinction. All other higher order effects involving face type and time (including the face type x time x target group interaction) were non-significant, p's > .122. Thus ingroup-outgroup asymmetries were not found.

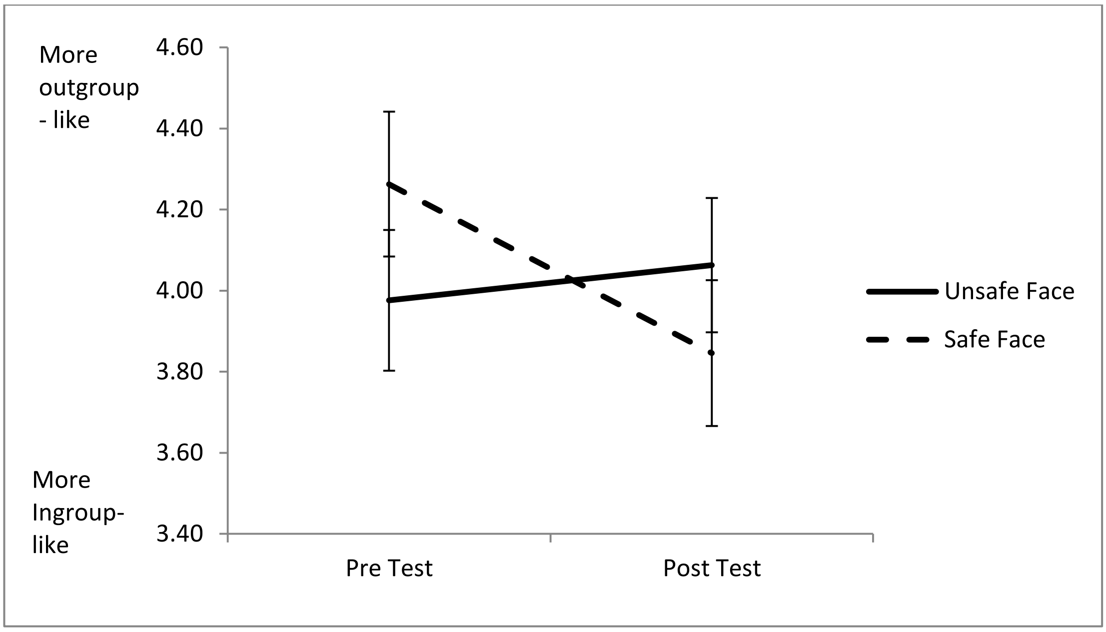


Figure 18. Face type x time interaction on perceived prototypicality for the unsafe and safe faces.

I expanded the previous model and included the *background generalisation faces* and *physiognomy-and-background generalisation faces* to test for generalisation effects. I conducted a 2 face type (unsafe and safe) x 3 generalisation (target, background generalisation, physiognomy-and-background generalisation) x 2 time (pre-and post-test) x 2 target group (ingroup and outgroup) x 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition, after extinction) mixed model ANOVA, with target group, face type, generalisation and time as the repeated measures. A face type x time interaction was found in the generalisation data, F (1, 50) = 5.35, p = .025, ηp^2 = .10. The 3 way interaction of face type, generalisation and time was however non-significant, F (2,100) = .46, p = .633, ηp^2 = .01, suggesting no differences between the target and two forms of generalisation faces. Hence, prototypicality shifts extended in a homogenous fashion from the target faces to the generalisation faces. Safe faces (collapsed across ingroup/outgroup and the three levels of the generalisation factor) were perceived as being more prototypical of the ingroup and less prototypical of the outgroup at post-test (M = 3.94, SD = 1.28) than at pre-test (M = 4.25, SD = 1.31), t (53) = 2.38, p = .021. This result indicated that the safe ingroup and outgroup face—across the target and generalisation faces—became more ingroup-like and less outgroup-like after aversive conditioning and, thus, replicated the effects found for the target faces. No effects were again found for the unsafe face, suggesting perceived prototypicality did not vary at post-test (M = 4.20, SD = 1.12) and pre-test (M = 4.04, SD = 1.17), t (53) = -1.35, p = .182.

In addition to the two way interaction, the only other significant higher order interaction from the generalisation data was a 5 way interaction involving face type x time x target group x generalisation x post-test position, F (2,100) = 7.07, p = .001, ηp^2 = .12. This interaction was investigated further by looking at the post-test position groups separately.

In the after extinction condition a significant face type x time interaction was detected, F (1, 23) = 5.57, p = .027. The three way interaction of face type x time x generalisation was again non-significant, F (2, 46) = .43, p = .650, ηp^2 = .02. These two results together indicate that among the after extinction participants, prototypicality shifts towards the generalisation faces was the same as shifts towards the target faces. There was no difference between pre- and post-test for the unsafe faces, t (24) = -.43, p = .675. However, the safe faces (target and generalisation faces) were perceived to be more ingroup-like and less outgroup-like at post-test (M = 3.75 , SD = 1.36) than pre-test (M = 4.30, SD = 1.35). A paired samples t-test confirmed this difference, t (24) = 3.30, p = .003. This is consistent with research conducted in the previous chapters, which found generalised prototypicality shifts after the extinction of anxiety. All higher order interactions were non-significant, p's > .144, which meant ingroup-outgroup asymmetries were not identified.

In the after acquisition group I found a marginally significant face type x time x target group interaction, F (1, 27) = 3.42, p = .076, ηp^2 = .11 and a significant face type x time x generalisation x target group interaction, F (2, 54) = 5.83, p = .005, ηp^2 = .18. Lower level analyses revealed that the interactions were due to two faces in particular (all other effects non-significant, p's > .090: The ingroup unsafe *background generalisation* face was perceived as being more prototypical of the outgroup and less prototypical of the ingroup at post-test (M = 3.37, SD = 2.19) than at pre-test (M = 2.77, SD = 1.77), t (29) = 1.99, p = .056; The outgroup safe *physiognomy-and-background generalisation* face was perceived as being more prototypical of the ingroup and less prototypical of the outgroup at pre-test (M = 4.43, SD = 1.90) than at post-test (M = 4.96, SD = 1.32), t (29) = 1.96, p = .058. Altogether, the results in the after acquisition group do not provide interpretable effects.

Together, results in this section suggest that safe target faces were perceived as being more ingroup-like and less outgroup-like. Unexpectedly, no changes in prototypicality of the unsafe target faces were found. The non-significant interaction involving post-test position suggests that the prototypicality shifts found were of a comparable magnitude between after acquisition and after extinction participants, which was also unexpected. The two unexpected results do not align with prototypicality shifts found in Chapter 3 and 4, which found prototypicality of the unsafe face at post-extinction only. Generalisation results suggest similar shifts of prototypicality generalised from the target faces to generalisation stimuli and the safe face was also responsible for generalised prototypicality shifts.

Exploring the Role of Contingency Awareness in Prototypicality Shifts

To determine the role that contingency awareness[3] played in *target faces* prototypicality shifts, I ran the same ANOVA previously for *target faces* and included contingency awareness as a between subject factor. When expanding the model to include contingency awareness a face type x time x contingency awareness interaction was found, F (1, 46) = 4.67, p = .036, ηp^2 = .09. All other effects involving contingency awareness were non-significant. I followed up the significant interaction by analysing results for contingent aware and unaware participants separately.

First I examined the effects for contingent *aware* participants. A main effect of target group verified ingroup faces were perceived as more ingroup-like (M = 3.32, SD = 1.00) and outgroup faces were perceived as more outgroup-like (M = 4.62, SD = 1.23), F (1, 32) = 13.71, p = .001, ηp^2 = .30. A face type x time interaction was detected, F (1, 32) = 8.96, p = .005, ηp^2 = .22. Similar to the pooled data, no change in prototypicality were found for the unsafe faces at post-test (M = 4.17, SD = 1.25) and pre-test (M = 3.90, SD = 1.48), t (35) = - 1.57, p = .126. In contrast, the safe faces became more ingroup-like and less outgroup-like at post-test (M = 3.64, SD = 1.53) than at pre-test (M = 4.16, SD = 1.61), t (35) = 2.71, p = .01.

A significant face type x time x target group x post-test position interaction was also detected among contingent aware participants and followed up by looking at post-test position groups (20 after acquisition participants, 16 after extinction participants) separately, F (1, 32) = 4.19, p = .049, ηp^2 = .12. Participants who completed post-test measures after acquisition produced no significant interactions, all *p*'s > .084. Contingent aware participants who completed post-test measures of prototypicality after extinction produced a significant face type x time interaction, F (1, 14) = 5.81, p = .030, ηp^2 = .29. Similar to the basic target exemplar shifts of prototypicality described above, safe faces—irrespective of whether they were ingroup or outgroup faces—were rated as more prototypical of the ingroup and less prototypical of the outgroup at post-test (M = 3.35, SD = 1.52) than at pre-test (M = 3.98, SD = 1.60), t (15) = 2.51, p = .024. No change in prototypicality was found for the unsafe face and similar prototypicality ratings were found at post-test (M = 4.50, SD = 1.34) and at pre-test (M = 4.21, SD = 1.59), t (15) = -1.06, p = .307.

Results for contingent *unaware* participants were analysed next. A significant face type x time x post-test position x participant's group membership was detected, F (1, 14) = 7.75, p = .015, ηp^2 = .36. This interaction was followed up by looking at post-test position and participants group membership conditions separately, but no significant results were detected. All other interaction involving face type and time were non-significant, *p*'s > .239.

I also expanded the contingency awareness analysis to include generalisation faces. I entered generalisation faces into the top level model analysed, and also separately for contingent aware and unaware participants. No effects were found in these analyses, suggesting contingency awareness had no effect on generalised prototypicality shifts.

Therefore, as expected and in line with previous chapters, entering contingency awareness into my analysis resulted in prototypicality shifts for contingent aware participants who completed post-test measures post-extinction. In this group the safe faces were perceived as being more prototypical of the ingroup and less prototypical of the outgroup. No effects were found for the unsafe faces or when contingent aware participants completed post-test measure after acquisition. Prototypicality shifts were not found for contingent unaware participants, which aligns with the idea that contingency awareness is implicated in prototypicality shifts. Unexpectedly, target face prototypicality shifts did not extend to generalisation faces when contingency awareness was entered as a factor into the analysis.

Summary of Results

The results from Study 5.1 provided novel evidence for prototypicality shifts of ingroup exemplars and replicated earlier findings of prototypicality shifts of outgroup exemplars, as a function of aversive conditioning among minimal groups: Across all participants the safe faces were perceived as being more prototypical of the ingroup and less prototypical of the outgroup. Unexpectedly, prototypicality shifts of the unsafe faces were not detected. Presumably, this effect was found because safe ingroup and outgroup exemplars promoted the positive distinctiveness that is expected of the ingroup and did not threaten the individual's safety. As a result safe faces were perceived closer to the ingroup prototype and further away from the outgroup prototype. This study included both ingroup and outgroup exemplars, and prototypicality shifts of the safe face might be a result of participants' psychological focus on the ingroup (positivity/safety) rather than the outgroup (negative/threatening). Furthermore, safe face prototypicality shifts were similar between the after acquisition and after extinction group. The results for post-test position were unexpected because exemplar prototypicality shifts were only detected after extinction in previous studies. It is possible that the processes that operate during extinction that facilitate outgroup exemplar prototypicality shifts detected in previous chapters are not needed when dealing with groups for which there are no pre-existing group cognitions (see Navarrete et al. 2012 for a similar discussion about the dissociation between learning effects and resistance to extinction in social vs minimal groups). Safe face prototypicality shifts also generalised to exemplars whose group membership cue and facial physiognomy was manipulated, providing evidence for generalisation effects.

Similar to Chapter 3 and 4, an exploratory analysis into contingency awareness found it played a role in driving ingroup *and* outgroup prototypicality shifts of safe faces. Prototypicality shifts were found in contingent aware participants, whilst no interpretable effects were found among contingent unaware participants. Furthermore, contingent aware participants' results were statistically reliable when participants completed post-test measures after the extinction procedure, but not beforehand. This is in line with the results found in Chapter 3 and 4 that found prototypicality shifts post-extinction only. Unlike previous studies, the effect did not generalise to similar faces amongst contingent aware participants. Contingency awareness results should be interpreted with caution and can only be considered preliminary due to the low power and small participant numbers as a resulting of exploring this factor.

While I obtained the standard ingroup favouritism effect pre-conditioning found in minimal group paradigm studies, there was no effect of target group on the prototypicality data. Hence, the ingroup-outgroup status of the targets did not qualify my basic prototypicality shift effects. As a result, I am not in a position to conclude which mechanism fits the results better (i.e., evaluative-fit explanations' emphasis on ingroup-focused effects vs. emotion-fit explanations' emphasis on outgroup-focused effects).

Study 5.2

In Study 5.2 I continued to investigate ingroup (vs outgroup) exemplar prototypicality shifts, but used established social groups rather than minimal groups. White Anglo Saxon participants underwent the standard acquisition and extinction procedures described in Chapters 3 and 4 with computer generated faces. However, a key change from Study 5.1 was a between-subjects design where participants saw *either* ingroup *or* outgroup exemplars (as opposed to both ingroup and outgroup exemplars). I moved to a between-subjects design for two reasons. Firstly, a within-subjects design with established social groups would most likely trigger demand characteristics, corrective processes, and floor/ceiling effects on prototypicality ratings (i.e., meta contrast – see Corneille & Judd, 1999). Secondly, a move to a between-subjects design meant a relational prototypicality measure was not used; in Study 5.1 exemplars were rated as being more ingroup like/less outgroup like (or vice versa). Instead of the relational measure, I will use a similar procedure to that used in Chapters 3 and 4 and prototypicality will be measured on one dimension only – either how prototypically White (ingroup condition) or Black (outgroup condition) exemplars are perceived to be. Similar to Chapters 3 and 4, exemplar prototypicality was measured with self-reported prototypicality and via a speeded sorting task, prior to conditioning, and post-extinction.

Generalisation data was again collected as part of this study. A 25% and 50% variation of the White and Black target faces were used to test for *face variation* generalisation effects. In addition, two new White and Black exemplars were used to test generalisation effects to *new exemplars*. In Study 5.1 generalisation effects were tested for by manipulating the group membership cue, background colour, and facial physiognomy. *Face variations* in Study 5.2 manipulated the group membership cue also, but this manipulation involved faces becoming progressively less White or Black and moving towards a computational average of all Facegen White, Black, Middle Eastern and Asian faces.

My hypotheses are based on evaluative-fit and emotion-fit mechanisms that focus on *ingroup* positivity and *outgroup* threat respectively. I hypothesise that the unsafe ingroup exemplar will be perceived as being less prototypical of the ingroup post-extinction than prior to conditioning. The safe ingroup exemplar is expected to shift in the opposite direction to the unsafe face and be perceived as being more prototypical of the ingroup. The unsafe outgroup exemplar was hypothesised to be perceived as more prototypical of the outgroup post-extinction than prior to conditioning. The safe outgroup exemplar is expected to shift in the opposite direction to the unsafe face and be perceived as being more prototypical of the outgroup. Unlike, Study 5.1 I expect target ethnicity to interact with face type and time because this will indicate a prototypicality shifts in different directions for ingroup and outgroup faces.

Method

Participants and Design

Sixty-six participants were initially recruited from a large regional Australian university. Due to technical errors three participants were excluded, which left 63 valid participants (21 male, 42 female; M = 23.33 years, SD = 8.48). All participants were White and reported an Anglo Saxon background. They received monetary compensation (AU$20) or partial course credit for their participation. Participants were randomly assigned to one of the two conditions in a 2 target ethnicity group design (White target exemplars *n* = 32, Black target exemplars *n* = 31).

Materials

Faces were developed using the face morphing software FaceGen v3.3.1. Two target faces from each ethnicity group were used during conditioning and acted as the unsafe and the safe face (counterbalanced). Target faces were either prototypically White or Black males, aged approximately 25 years, clean shaven and of frontal orientation with neutral expressions. Generalisation effects were tested for using a 25 and 50 percent configurally related variations of each target face that moved progressively away from the White or Black prototype (towards a computational average of all FaceGen White, Black, Middle-eastern and Asian faces) and two new prototypical White or Black faces. The Black faces used in this study were the same set of faces described in Study 3.1. White faces were selected based on comparable prototypicality and anxiety ratings to the Black faces through pilot testing. Sixteen pilot participants of Anglo-Saxon ethnicity rated the target faces along prototypicality with respect to their target group (*Prototypically White/Black*: 1 = *not at all*, 7 = *very much*) and anxiety (1 = *not at all*, 7 = *very much*). Target Faces were found to be high in perceived prototypicality (ingroup *Grand M* = 5.16, *SD* = 1.64, outgroup *Grand M* = 5.09, *SD* = 1.60) and low in anxiety (ingroup *Grand M* = 3.25, *SD* = 1.20, outgroup *Grand M* = 2.78, *SD* = 1.18). Individual face means along prototypicality and anxiety can be observed in Table 1. I conducted paired samples *t*-tests by comparing each Black face with each White face, and by comparing the two White/Black faces with each other. Pairwise comparisons confirmed no differences between the paired faces on prototypicality and anxiety, all *p*'s > .135. Hence, the White and Black faces used in this study were selected from the target faces and are suitable for this study.

Table 1. Pilot Study Results

Target Face	Prototypicality		Anxiety	
	M	SD	*M*	*SD*
White target face 1	4.69	1.97	3.63	1.75
White target face 2	5.63	1.57	2.88	1.09
Black target face 1	4.72	1.93	2.81	1.38
Black target face 2	5.47	1.48	2.75	1.13

Procedure and Measures

To minimise response biases caused by repeated measurements, participants completed two testing sessions separated by a minimum of five days and a maximum of 28 days (*M* = 10.44, *SD* = 8.98). During the initial session, participants were seated in front of the computer screen and completed an on-line questionnaire and the speeded sorting task. Target ethnicity group was manipulated *between-subjects* from the onset and participants made face judgments towards exemplars of the ethnic group they were randomly allocated to (either White ingroup faces or Black outgroup faces). Participants indicated the extent to which each of the White or Black eight (randomly ordered) faces were prototypical of their target ethnicity group (*prototypically White/Black*: 1 = *not at all*, 7 = *very much*). To check whether self-reported extinction occurred, we also asked our participants to rate faces along anxiety (*anxious*: 1 = *not at all*, 7 = *very much*). To check how similar faces were, similarity was measured between each pair of faces presented throughout the study (1 = *Not at all similar*, 7 = *Very similar*). A feeling thermometer was also included to check for the standard intergroup bias - participants rated their overall feeling of the ingroup and outgroup between 0 (*very cold*) and 100 (*very warm*) degrees, with the scale increasing in 10 degree increments.

The speeded sorting task followed the online questionnaire. Participants were presented with a set of 16 faces that included the individual White and Black target, generalisation and new exemplar faces individually at the centre of the screen. Participants were instructed to sort each individual face as quickly and as accurately as possible into the "Black" or "White" categories by pressing the green (left handed "S" key) or blue (right handed "L" key) on the keyboard. The category labels were presented in the top left and right corners of the screen and corresponded to the location of the relevant key. Faces were presented in a series of continuing blocks, with one block consisting of all 16 faces. There were 14 blocks in total (i.e., each face presented 14 times) and response keys and category labels were counterbalanced after seven blocks. Faces were inverted for 25% of the presentations to increase task difficulty and, thus, task engagement (Richeson & Trawalter, 2005).

The second laboratory session began with the work-up procedure described in Study 5.1 to select a level of shock regarded as uncomfortable but not painful by participants (Lovibond et al., 2008). The same equipment and method described in Study 5.1 were used to administer the electrical stimulation and measure skin conductance responses and respiration. As part of the acquisition task, two target faces (either Black or White depending on condition allocation) were presented at the centre of the screen six times for 10 s (inter-stimulus interval M = 17.5 s, range 15-20 s), each in a randomised order. One target face (the unsafe face) always co-terminated with a 200 ms electrical stimulation at the level selected during the work up procedure. The other target face (the safe face; counterbalanced) was never paired with a stimulation. To test for acquisition, the pre- and post- acquisition data was collected continuously either side of the acquisition task using the same procedure described in Study 5.1.

The extinction phase followed post-acquisition skin conductance data collection and used the same standard extinction procedure described in Chapter 3 and Study 5.1. Briefly, the two target faces were presented (an identical number of times) in the absence of the electrical stimulation until no increases in SCR could be detected. The number of extinction trials was set to a minimum of 5 and a maximum of 25 presentations per target face. Similar to Study 3.1 and post-extinction participants in Study 3.2, all post-test data in this study was collected post-extinction. The same measures described at pre-test (self-reported prototypicality and anxiety, and the speeded sorting task) were again collected. In addition to these measures, I also tested for contingency awareness using the same procedure described in Study 5.1. Twenty participants were classified as contingent non-aware (10 ingroup target ethnicity; 10 outgroup target ethnicity) and 43 as contingent aware (22 ingroup target ethnicity; 21 outgroup target ethnicity).

Results

Checking for Standard Intergroup Bias

The standard intergroup bias effect was checked for in a 2 target group (ingroup and outgroup) x 2 target ethnicity (ingroup or outgroup) mixed model ANOVA on participants' feeling thermometer ratings for the ingroup and outgroup, with target group as the repeated measure. Target group refers to the judgements made about the ingroup and outgroup on the feeling thermometers, whilst target ethnicity refers to participant's random group allocation to which ethnicity type they will be conditioned with. A main effect of target group was found and the ingroup was rated more positively (M = 81.11, SD = 14.53) than the outgroup (M = 71.59, SD = 19.61) at pre-test, F (1, 61) = 19.59, p <.001, ηp^2 = .24. Thus, the standard intergroup bias was found.

There was no difference in intergroup bias between the target ethnicity group allocation as the two way interaction between target group and target ethnicity was non-significant, p = .696.

Checking Effective Acquisition and Extinction

To determine if effective acquisition and extinction occurred, I first examined participant's SCRs. SCRs were recorded and scored the same way described in Study 5.1. A standard acquisition effect was checked for by analysing SCRs with a 2 target ethnicity (ingroup or outgroup) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed ANOVA with face type and time as the repeated measures[4]. Main effects of face type, F (1, 60) = 8.74, p =.004, ηp^2 = .13, and time were detected, F (1, 60) = 11.88, p =.001, ηp^2 = .17. More importantly, a marginal face type x time x target ethnicity interaction was observed, F (1, 60) = 3.44, p = .069, ηp^2 = .05. This interaction was followed up by looking at each target ethnicity group separately.

The ingroup condition produced a main effect of face type, F (1, 31) = 9.10, p = .005, ηp^2 = .28, and time, F (1, 31) =5.23, p = .029, ηp^2 =.14. Importantly, a face type x time interaction was also observed, F (1, 31) = 5.16, p = .03, ηp^2 =.14. Paired samples t-tests indicated that the ingroup unsafe face had higher SCRs at post-test (M = 1.18, SD = .27), compared to pre-test (M = 1.07, SD = .15), t (31) = 2.44, p = .021; the ingroup safe face did not change in SCRs from post-test (M = 1.06, SD = 1.47) to pre-test (M = 1.05, SD = .13), t (31) = .52, p = .608. This result indicated that the ingroup unsafe face had higher SCRs at post-test, which is indicative of more anxiety after acquisition. It suggests conditioning successfully occurred in the ingroup condition.

The outgroup condition produced a main effect of time, F (1, 29) = 6.62, p = .015, ηp^2 = .19. Both faces became more anxiety provoking at post-test (Grand total: M = 1.16, SD = .23; unsafe face: M = 1.17, SD = .25; safe face: M = 1.15, SD = .21) than at pre-test (Grand total: *M* = 1.06, *SD* = .11; unsafe face: *M* = 1.08, *SD* = .12; safe face: *M* = 1.04, *SD* = .09). Unexpectedly, the 2 way interaction involving face type and time was non-significant, *F* (1, 29) = .13, *p* = .719, ηp^2 = .01. Together the results suggested participants did not distinguish between the unsafe and safe outgroup faces and displayed higher SCR's towards both the unsafe and safe outgroup faces.

To check for effective extinction of the association, I ran a 2 target ethnicity (ingroup vs outgroup) x 2 face type (unsafe and safe face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR data. SCR data during the first block of extinction (first two presentations of the unsafe and safe face) and the ten trial block (the last two presentations of the unsafe and safe face up to the tenth trial) was chosen as all participants were exposed to a minimum of ten extinction trials. As expected there was no main effect of face type, *F* (1, 60) = .84, *p* = .351, η_p^2 = .02, or face type x extinction block interaction, *F* (1, 60) = 3.39, *p* = .07, η_p^2 = .05. These two results suggest there is no difference between the unsafe and safe face throughout extinction. I supplemented these data analyses with additional analysis to confirm SCRs were brought back to baseline. Supplemental data analyses conducted on SCR suggests negativity/anxiety extinguished and are discussed extensively in Appendix P6.

The physiological measure of anxiety provided evidence that anxiety was extinguished. However, to confirm effective extinction of anxiety, *self-reported* anxiety collected after the extinction procedure was also analysed with a 2 target ethnicity (ingroup or outgroup) x 2 face type (unsafe and safe face) x 2 time (pre- and post-extinction) mixed ANOVA with face type and time as the repeated measures. A main effect of time was detected and anxiety was higher at post-extinction (*M* = 3.26, *SD* = .87) than at pre-test (*M* = 2.89, *SD* = 1.07), *F* (1, 61) = 4.59, *p* = .038, ηp^2 = .069. All other effects were non-significant, all *p*'s > .193. This result suggests both target faces were considered more anxiety provoking post-extinction than at pre-test and residual anxiety in general was present at the time in which we collected prototypical ratings. Residual anxiety may have affected prototypicality judgments and perceptions, relating to the issues described through Chapter 4 that investigated extinction in more depth.

Testing Basic and Generalised Prototypicality Shifts

To test for prototypicality shifts I analysed self-reported prototypicality and sorting task data. I expected the unsafe ingroup exemplar to be perceived as being less prototypical of the ingroup and the unsafe outgroup exemplar to be perceived as being more prototypical of the outgroup. In contrast and based on Study 5.1, safe faces are expected to be perceived as being more prototypical of the ingroup and less prototypical of the outgroup. The expected shifts are based on evaluative-/emotion-fit mechanisms because positivity/safety is typically associated with the ingroup and negativity/anxiety with the outgroup.

Basic prototypicality shifts were first tested for with self-reported prototypicality data using a 2 target ethnicity (ingroup or outgroup) x 2 face type (unsafe and safe) x 2 time (pre-test and post-extinction) mixed model ANOVA with face type and time as the repeated measures. The expected three way interaction between target ethnicity, face type and time was non-significant, $p = .904$. However a significant target ethnicity x time interaction was detected and subsequently followed up, $F\ (1, 61) = 4.58$, $p = .036$, $\eta p^2 = .07$. I combined the scores for the unsafe face and safe face at pre-test and post-test and carried out a paired samples *t-test* in each target ethnicity. No difference was observed in the *ingroup* condition between pre-test ($M = 5.75$, $SD = .20$) and post-test ($M = 5.70$, $SD = 1.70$), $t\ (32) = .29$, $p = .776$). In the outgroup condition both outgroup faces, irrespective of whether they were paired with the aversive stimulus or not, became more prototypical of the outgroup at post-extinction (*M* = 5.98, *SD* = .76) compared to pre-test (*M* = 5.44, *SD* = .99), *t* (30) = -2.85, *p* = .008. Participants allocated to the outgroup target ethnicity did not discriminate between the unsafe and safe face, and instead perceived a shift towards the outgroup prototype for both faces.

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A 2 target ethnicity x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on the self-reported prototypicality data. A generalisation gradient main effect was found, *F* (2,122) = 522.27, *p* < .001, ηp^2 = .90, but more importantly a generalisation gradient x time interaction was detected, *F* (2,122) = 7.20, *p* = .001, ηp^2 = .11. The faces followed the expected generalisation gradient pattern with target faces being rated more prototypical of the group, followed by 25% faces and then 50% faces. The generalisation gradient x time interaction was driven by two contrasting trends: Target faces were rated as being more prototypical of the ingroup/outgroup at post-extinction (*M* = 5.84, *SD* = 1.09) than at pre-test (*M* = 5.60, *SD* = 1.17) whilst the opposite effect was found for generalisation faces and rated higher in ingroup/outgroup prototypicality at pre-test (25% faces *M* = 3.43, *SD* = 1.17; 50% faces *M* = 2.37, *SD* = 1.02) than post-extinction (25% faces *M* = 3.03, *SD* = 1.22; 50% faces *M* = 2.18, *SD* = 1.06). This suggests that irrespective of target ethnicity, target faces increased in prototypicality whilst generalisation faces decreased in prototypicality.

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face, safe face, or equally similar. The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new face, using a 2 target ethnicity x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on sorting task data. The first new exemplar was rated as being more prototypical at post-extinction (M = 6.17, SD = .83) than at pre-test (M = 5.78, SD = 1.16) as evident by a time main effect, F (1, 57) = 6.14, p = .016, ηp^2 = .10. A time main effect was also detected for the second new exemplar and it was rated as being more prototypical at post-extinction (M = 6.50, SD = .66) than at pre-test (M = 5.93, SD = 1.30), F (1, 57) = 11.80, p = .001, ηp^2 = .17. All other effects for the first and second new exemplar were non-significant, all p's > .094. These two results suggest that both new exemplars, irrespective of whether they were ingroup or outgroup faces, became more prototypical of their respective group. This effect was not influenced by whether the face was more similar to the unsafe or safe face as evident by non-significant 2 and 3 way interactions.

Latencies during the speeded sorting task were also used to investigate prototypicality shifts towards the unsafe and safe exemplars. The same procedure described in Study 3.2 was used to measure and analyse data from the speeded sorting task. Briefly, incorrect categorisation responses were excluded from the latency data analysis and the mean reaction time of the latencies for the correctly categorised unsafe and safe exemplars were log-transformed to normalise the data (Richeson & Trawalter, 2005; Ruys, Dijksterhuis & Corneille, 2008). Responses quicker than 300 ms were excluded from the analyses and extremely long responses (> 3 SD) were rescored to the third standard deviation value for each group. On average, participants incorrectly sorted faces 9.74% of the time (SD = 5.02); a one way ANOVA confirmed there was no systematic difference in errors as a function of type of extinction, p = .224. The average time taken to sort faces were: The unsafe exemplar at pre-test 546 ms (SD = 78 ms) and at post-extinction 554 ms (*SD* = 89 ms); the safe exemplar at pre-test 544 ms (*SD* = 72 ms) and post- extinction 550 ms (*SD* = 83 ms).

Basic prototypicality shifts of the target faces were investigated using a 2 target ethnicity (ingroup or outgroup) x 2 face type (unsafe and safe) x 2 time (pre-test and post- extinction) mixed ANOVA with face type and time as the repeated measures on the sorting task data. Results showed a marginal time x target ethnicity interaction, F (1, 61) = 3.70, p = .059, ηp^2 =.06. The time x target ethnicity interaction was subsequently explored further and followed up by looking at each target ethnicity group separately. Despite this interaction only being marginally significant, I explored the data further to provide an insight into what direction exemplar prototypicality was shifting – albeit a minor shift of prototypicality. The differences across pre- and post-extinction for each target ethnicity group can be observed in Figure 19. The ingroup condition produced a marginal main effect of time, F (1, 31) = 3.66, p = .065, ηp^2 = .11. There was a tendency for all ingroup faces to be sorted more slowly at post-extinction (*M* = 2.76, *SD* = .07), compared to pre-test (*M* = 2.74, *SD* = .07). Hence, there was a trend for both the unsafe and safe ingroup faces to be perceived as less ingroup-like post-extinction. The outgroup condition failed to produce any significant effects, all *p*'s > .411.

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the unsafe and safe face in order to test for generalisation effects. A 2 target ethnicity x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on the sorting task data[5]. The time main effect held and all faces were perceived as being more prototypical as demonstrated by quicker response times to sort faces at post-extinction (*M* = 2.79, *SD* = .07) than at pre-test (*M* = 2.81, *SD* = .07), F (1, 56) = 5.03, p = .029, ηp^2 = .08. The main effect was further qualified by a generalisation gradient x time interaction, *F* (2,112) = 9.14, $p < .001$, $\eta p^2 = .14$. This interaction was largely driven by the 50% target variation face being sorted quicker at post-extinction (*M* = 2.84, *SD* = .09) than at pre-test (*M* = 2.87, *SD* = .11), *t* (60) = 4.02, *p* = <.001. The other two faces did not change in the time from being sorted at pre-test and post-extinction, both *p*'s > .278. No other effects were found involving the generalisation gradient, *p'*s > .193; hence, non-significant interactions involving target ethnicity suggest results did not vary between the White and Black target exemplars and no further analyses were carried out on the variation data.

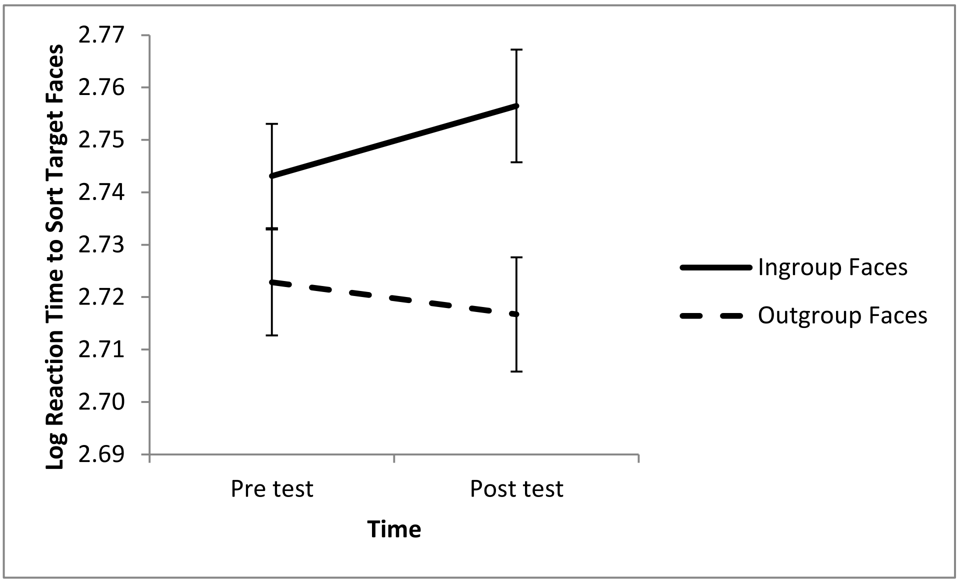


Figure 19. Log Reaction times to sort ingroup and outgroup Faces

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the unsafe face, safe face, or equally similar. The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 type of extinction x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on sorting task data. No effects were found for either new face, all *p*’s > .323.

In summary, self-reported data found a non-associative prototypicality shift in the *outgroup* condition and both the unsafe and safe outgroup faces were perceived as being more prototypical of the outgroup post-extinction. No effect was found for ingroup faces on self-reported data. Similarly, no generalisation effect was found for face variations with self-reported data, but both new exemplars shifted in prototypicality. Target ethnicity was not involved in the new exemplar prototypicality shift meaning new ingroup exemplars were perceived as being more prototypical of the ingroup and outgroup exemplars more prototypical of the outgroup post-extinction. The speeded sorting task detected a non-associative shift also, but the shift was found in the *ingroup* condition. Both the unsafe and safe ingroup faces tended to be perceived as being less prototypical of the ingroup at post-extinction. No effect was found for outgroup faces on the speeded sorting task. No interpretable generalisation effects were found to either face variations or new exemplars on the speeded sorting task.

Exploring the Role of Contingency Awareness in Prototypicality Shifts

To analyse the effects contingency awareness[6] had a 2 target ethnicity (ingroup vs outgroup) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) mixed model ANOVA with face type and time as repeated measures was run on the self-reported prototypicality data. In this top level ANOVA no significant interactions were detected, all *p*’s > .399. Despite no significant interactions being found, planned comparisons were carried out in line with Study 3.2 and I investigated for prototypicality shifts in contingent aware and unaware participants separately (Keppel & Wickens, 2004).

I repeated the earlier analysis separately for contingent aware (n = 43) and contingent unaware (n = 20) participants. As expected, no significant results were observed for contingent *unaware* participants, all p's > .320. Similar to the previous results, a marginally significant time x target ethnicity interaction was detected for contingent *aware* participants, F (1, 41) = 4.07, p = .050, ηp^2 = .09. I followed up this interaction by looking at the ingroup and outgroup conditions separately. There were no significant results observed in the *ingroup* condition, with only a face type main effect approaching significance, F (1, 9) = 2.65, p = .068, ηp^2 = .15. All other effects were non-significant in the ingroup condition, all p's > .669. A significant time main effect was detected in the *outgroup* condition, F (1, 20) = 5.31, p = .032, ηp^2 = .21. The previous effect in the outgroup condition for all participants was replicated and both outgroup exemplars were perceived more prototypical of the outgroup at post-extinction (grand total M = 5.80, SD = .79; unsafe face M = 5.95, SD = .86; safe face M = 5.64, SD = 1.20) than at pre-test (grand total M = 5.21, SD = 1.00; unsafe face M = 5.33, SD = 1.09; safe face M = 5.10, SD = 1.14). I expanded this investigation to include the *face variations*. The generalisation gradient x time interaction detected in the main analysis was found again amongst *contingent aware* participants. All other results in the *contingent aware* and *unaware* group were non-significant, all p's > .152. In summary, amongst contingent aware participants both the unsafe and safe face were perceived as being more prototypical of the respective ethnicity post-extinction than at pre-test. As the interaction involving target ethnicity was non-significant, it suggests similar trends between the ingroup and outgroup exemplars. Caution should be taken when interpreting these results as the main interaction was non-significant and smaller sample sizes as a result of further breaking target ethnicity groups down result in limited power.

Next I investigated the effects contingency awareness had on prototypicality shifts measured with the speeded sorting task. I ran a 2 target ethnicity (ingroup vs outgroup) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (unsafe and safe face) x 2 time (pre-test and post-extinction) mixed model ANOVA with face type and time as repeated measures on the sorting task data. Unexpectedly, no significant interactions involving contingency awareness were observed, all *p*'s > .215. Despite no significant interactions, similar to Chapter 3, I investigated contingent aware and contingent unaware participants separately in line with Keppel and Wickens (2004) planned comparison procedure.

Unexpectedly, contingent *unaware* participants produced a significant time x target ethnicity interaction that was subsequently followed up, F (1, 18) = 4.87, p = .041, ηp^2 = .21. The ingroup condition produced a marginal time main effect, F (1, 9) = 4.32, p = .067, ηp^2 = .33. Both ingroup faces were perceived as less prototypical of the ingroup at post-extinction (grand total *M* = 2.77, *SD* = .04; unsafe face *M* = 2.76, *SD* = .04; safe face *M* = 2.77, *SD* = .05) than at pre-test (grand total *M* = 2.74, *SD* = .02; unsafe face *M* = 2.74, *SD* = .03; safe face *M* = 2.74, *SD* = .03). There was no discrimination between the ingroup unsafe and safe face and both were perceived as being less prototypical of the ingroup as a function of conditioning. No effects were detected in the outgroup condition among contingent unaware participants, all *p'*s > .366. Unexpectedly, no significant effects were found in the contingent *aware* group, all *p'*s > .319. This pattern of results did not replicate previous findings which found contingency awareness was involved in prototypicality shifts. When the model expanded to include *variation* faces, no effects were detected, all *p*'s > .075. In summary, the time main effect for ingroup faces detected among all participants was replicated in contingent unaware. No effects were found in the contingent aware group or for any generalisation data involving contingency awareness.

Summary of Results

In summary, my manipulation checks for acquisition and extinction suggest acquisition occurred in the ingroup condition, as SCR data showed the unsafe face evoked higher SCRs at post-test than at pre-test; contingent-specific conditioning did not occur in the outgroup condition, as both the unsafe and safe face became more anxiety provoking at post-test than at pre-test. Physiological extinction of anxiety occurred and participants showed no increases in skin conductance. However, self-reported data showed anxiety was higher in both target ethnicity groups after extinction, although the anxiety was not face specific. Prototypicality results from this study showed a trend for ingroup exemplars to be perceived as less prototypical of the ingroup and outgroup exemplars to be perceived as more prototypical of the outgroup after aversive conditioning. Unexpectedly, non-associative prototypicality shifts were found and both the unsafe and safe target faces were perceived as being less prototypical of the ingroup and more prototypical of the outgroup. This result is inconsistent with the pattern of results found in Study 5.1 and the previous experimental chapters because face specific prototypicality shifts were detected. In particular, the sorting task produced a marginal trend showing both safe and unsafe ingroup faces were perceived as being less prototypical of the ingroup – despite SCR data suggesting discriminative conditioning. This trend was more apparent in contingent unaware participants than contingent aware participants, which is not in line with previous findings for contingency awareness. The pattern of results is reversed in the outgroup condition and the unsafe and safe faces were perceived as more prototypical of the outgroup in the self-reported data. In the outgroup condition prototypicality shifts were larger in participants who were aware of the face-shock pairing.

General Discussion

Summary and Interpretation of Key Findings

The present research expanded on previous research and investigated how an *ingroup* exemplar perceived prototypicality shifted after being associated with a negative/anxiety provoking stimulus. Perceived prototypicality was measured prior to, and after, aversive conditioning with a negative/anxiety provoking stimulus. Study 5.1 tested for prototypicality shifts with minimal groups in a within-subjects design (ingroup *and* outgroup exemplars presented), whilst Study 5.2 tested for shifts with social groups in a between-subjects design (ingroup *or* outgroup exemplars presented). I expected to find prototypicality shifts that were consistent with evaluative-fit and emotion-fit mechanisms and expected unsafe exemplars to be perceived as less ingroup-like and more outgroup-like after conditioning. In contrast, I expected safe exemplars to be perceived as more ingroup-like and less outgroup-like after conditioning.

The expected prototypicality shifts were partially found in Study 5.1 using minimal groups. Safe faces were perceived as being more ingroup-like and less outgroup-like post-extinction than at pre-test. However, no shift of prototypicality was detected in the unsafe face. Similar to the third and fourth chapters, contingency awareness was an important process in the development of prototypicality shifts. Shifts of prototypicality of the safe faces in Study 5.1 were stronger for participants who were aware of the contingent relationship between the face-shock pairing. Furthermore, contingent aware prototypicality shifts occurred only when post-test measures were collected after extinction and not after acquisition. This is similar to the studies reported in the first two chapters where extinction appeared to play a role in changing exemplar prototypicality. In contrast, prototypicality shifts were not observed in contingent unaware participants.

This set of results mapped onto the expectations derived from social psychological (i.e., self-categorisation theory; Oakes et al., 1994; Oakes, et al., 1991; Turner et al., 1987) and evolutionary theories (i.e., sociofunctional approach; Cottrell & Neuberg, 2005; Dasgupta et al., 2009; Mineka & Ohman, 2002; Seligman, 1971). These results are consistent with the idea that individuals were motivated to belong to a positively valenced group and a group associated with (cooperation and) safety (Brewer, 1999). Presumably, the psychological inclusion into the ingroup of safe exemplars in the form of perceptions of heightened ingroup prototypicality/lowered outgroup prototypicality enhanced the positive valence and integrity of the ingroup and/or the physical integrity of the individual within this group. Thus, the perception of safe exemplars shifted and were perceived as being more ingroup like after conditioning, which is consistent with the positive/safety mechanisms of evaluative and emotion-fit.

A notable difference between previous experimental chapters and Study 5.1 was which faces shifted in their representation. My previous studies found the unsafe faces shifted in prototypicality more than the safe faces, whereas in this study safe faces shifted in prototypicality more than the unsafe faces. I believe this difference was due to the comparative context in which participants saw Study 5.1's faces. In my previous experiments, participants made judgments towards outgroup exemplars only, and ingroup exemplars were not available. The inclusion of the ingroup in comparison with the outgroup may have changed the way participants made judgments. Previous research has demonstrated that, in intergroup settings, judgments and perceptions of the ingroup are more severe compared to those of outgroup (Marques & Paez, 1994; Pinto et al., 2010). Individuals are more concerned with the ingroup and how it is perceived and judged, with the outgroup perception being psychologically less important (Brewer, 1999; Ryan, Robinson, & Housmann, 2004). In this study safe faces were more important to the ingroup perception because participants were psychological including these exemplars into the group, with less attention being given to exemplars who were psychologically excluded from the group. For example, when a threat is perceived a group will unite and ingroup identification becomes more critical. Hence, a comparative context in which ingroup and outgroup exemplars are perceived results in group polarisation, in which the ingroup perception becomes significant (Doosje, Haslam, Spears, Oakes & Koomen, 1998; Doosje, Spears, Ellemers & Koomen, 1999).

Safe face prototypicality shifts detected in Study 5.1 generalised to similar faces. The generalisation faces in this study differed from face variations in previous studies because the group membership cue that was varied was unrelated to the face. Rather, this study manipulated the background colour to appear less blue and more green (or less green and more blue). This means the group membership moved away from the ingroup and closer to the outgroup (or away from the outgroup and closer to the ingroup). Another set of generalisation faces was used, which consisted of the varied group membership cue in conjunction with the variation of target faces facial physiognomy. Results revealed that the safe face prototypicality shifts found for target faces generalised to the two types of variation faces. These results suggest that group membership cues are important in order for effects to generalise and facial physiognomy not as important. Whilst it is possible that conditioning occurred to the background colour and the face was irrelevant, my interpretation aligns with previous studies that used face variations that changed both group membership cues and facial physiognomy together. Results from this study suggest the previously detected generalisation effects in previous studies were due to changes in group membership cues.

Study 5.2 results differed from what was found in Study 5.1 and participants failed to distinguish between the unsafe and safe face. In Study 5.2 both the ingroup unsafe and safe face were perceived as being less prototypical of the ingroup after they were paired with an aversive stimulus on the sorting task. Hence, the prototypicality shift generalised to both target faces and non-associative shifts of prototypicality were found. A non-associative specific shift of prototypicality was also detected for outgroup faces on the self-reported measure. Awareness of the contingent relationship did not improve results: When participants were contingent aware the expected shifts in prototypicality were expected to be enhanced. However, participants did not distinguish between the two target faces and a shift in prototypicality for the unsafe and safe face in the same direction was found again. Despite finding no evidence for contingent-specific prototypicality shifts, the non-associative prototypicality shift was in the expected direction as negativity/anxiety became associated with both exemplars. Consistent with evaluative-fit and emotion-fit mechanisms, ingroup faces were perceived as being less ingroup like and outgroup faces as more outgroup like following aversive conditioning.

Limitations and Future Research Ideas

I believe that the disparity between Study 5.1 and 5.2 occurred for a variety of reasons. Firstly, the two studies reported in this chapter used different designs. Study 5.1 used a within subject design and Study 5.2 used a between subject design. In Study 5.1, the within subject design facilitated comparisons between the ingroup and outgroup and this could have led to the different trends in the prototypicality shifts for the unsafe and safe faces. Comparative contexts between ingroups and outgroups (vs ingroup or outgroup only) vary the way in which participants perceive a group and these perceptions are not fixed (Doosje et al., 1998; Doosje et al., 1999). For example, research conducted by Corneille and Judd (1999) into the meta-contrast principle found intergroup differences were enhanced when participants made judgments in the presence of another group. Furthermore, prototypicality is not a static property and is malleable depending on a host of contextual factors (Bless & Schwarz, 1998; Corneille & Judd, 1999). On key dimensions that were important to each group's identity, the difference in these dimensions were more pronounced when impressions were formed about multifaceted stimuli in the context of another group than by when the stimuli was presented in the absence of another group. I believe a similar effect occurred in Study 5.1 and the safe faces were placed subjectively more in the ingroup category. Study 5.2 used a between subjects design and the perception of ingroup faces was not formed in the presence of the outgroup. Rather than the accentuation of two groups' similarities and differences being formed, an over exclusion bias could have occurred. Any ingroup member that could possibly harm the positive/non-threatening image of the group was excluded (Leyens & Yzerbyt, 1992).

A second factor in the disparity between studies' results could be the lack of discriminative conditioning in Study 5.2. In the previous experimental chapters I found evidence for exemplar prototypicality shifts following discriminative conditioning – only the unsafe exemplar was perceived as more negative/anxiety provoking after conditioning. In Study 5.1, contingent-specific conditioning occurred and the ingroup and outgroup unsafe faces were perceived as more negative/anxiety provoking. Contingent specific conditioning resulted in prototypicality shifts of the safe face. In Study 5.2, contingent specific conditioning did not occur in the outgroup condition and both the unsafe and safe face were perceived as being more negative/anxiety provoking. In the absence of discriminative conditioning, a non-associative prototypicality shift of the unsafe and safe face was found. The conditioning procedure used in Study 5.2 was similar to the other studies reported in previous chapters and previous research (Mallan, Sax & Lipp, 2009; Olson & Fazio, 2006; Olsson et al., 2005; Weisbuch, Pauker & Ambady, 2009), and contingent specific conditioning should have occurred. Further research is needed to determine why contingent-specific conditioning did not occur, and to better understand the effect this had for prototypicality shifts.

Another unexpected difference between the two studies was the role that contingency awareness had in the results. In Study 5.1 prototypicality shifts were more pronounced in the contingent aware group, which was expected. In Study 5.2 however, a non-associative ingroup exemplar prototypicality shift was detected in the contingent unaware group, with no shifts in prototypicality found in the contingent aware group. This result is in direct contrast to Study 5.1 and the previous two chapters, which found prototypicality shifts were greater amongst contingent aware participants. This unexpected finding should be interpreted with caution, as the exploratory nature of this analysis resulted in small sample sizes and reduced power. Contingency awareness did not inform decisions around required sample sizes, which led to power issues and possible false-positive/negative results (Cohen, 1992). Future research should provide a more stringent test of my exploratory analyses by taking these factors into account when planning the study design and required sample sizes.

The data obtained in this chapter and in my previous chapters cannot distinguish between evaluative-fit and emotion-fit mechanisms as asymmetries in the direction of prototypicality shifts between ingroup and outgroup exemplars could not be detected. Target group was not involved in any interactions found in Study 5.1 and there was no obvious ingroup-outgroup asymmetry found on prototypicality measures in Study 5.2. Thus, at an empirical level I cannot distinguish between the two mechanisms.

Moreover, I used a conditioning procedure that conflates negativity (valence) and anxiety (threat). By using a procedure that co-inflates valence and threat, I cannot determine whether changes in prototypicality were driven by valence (in line with the evaluative -fit) or anxiety (emotion-fit). Understanding which mechanism explains prototypicality shifts will help to identify necessary processes to improve stereotypes and relations and society.

To distinguish between evaluative-fit and emotion-fit mechanisms a procedure that does not co-inflate valence and emotion should be used. To test the role that valence alone has in driving prototypicality shifts, a conditioning procedure with a purely negative stimulus that is not anxiety provoking should be used. As such, an evaluative-fit mechanism into prototypicality shifts can be tested in a way that controls for emotion-fit. For example, Delgado, Labouliere and Phelps (2006) showed pairing money loss with a stimulus was sufficient in causing the stimulus to be perceived more negatively. A similar conditioning procedure could be used and money loss could be paired with exemplars. Perceived prototypicality can be tested prior to and after conditioning with money loss, as this will provide a test for evaluative-fit that is independent of emotion-fit. Furthermore, a reversal of this effect can be investigated by pairing faces with money gain (appetitive conditioning). Other possibilities for future experiments include manipulating the source of the threat posed to the individual around self-esteem vs physical threat.

Further research should also try to provide clarity over whether the exemplar's prototypicality is shifting, or whether the group prototype is shifting. Based on exemplar models of categorisation (Kruschke, 2011; Nosofsky, Kruschke & Mckinley, 1992; Zaki, Nosofsky, Stanton & Cohen, 2003), I interpret my results as the faces shifting towards or away from the group prototype. However, another possibility is the group prototype is shifting closer to the unsafe/safe exemplar. From this research I cannot distinguish whether it is the exemplar shifting towards the group prototype, or vice versa. To distinguish between the two processes, future research should include a measure that determines what is shifting. For example, questions that measure stereotypical features of a group can be asked in conjunction with exemplar prototypicality, to determine whether the exemplar or group perception shifts. This type of measure will provide a greater understanding of the processes that operate in exemplar and group perception.

Conclusion

To conclude, to the best of our knowledge, this is the first research that has investigated the consequences that aversive conditioning has for shifting the goodness of fit ingroup exemplars had with the ingroup prototype. In our view, the prototypicality shifts identified in this research have implications for how group stereotypes are formed and maintained. In this chapter I showed the safe ingroup exemplar that was never paired with the aversive stimulus was perceived as more ingroup-like after conditioning, at least in Study 5.1. In Study 5.2, a non-associative ingroup exemplar prototypicality shift was detected on the sorting task and both the unsafe and safe face were perceived as less ingroup-like after conditioning. I explained these results as a reflection of the negative/threatening members being fenced off and distanced from the ingroup prototype and non-negative/safe members being closely aligned with the ideal representation or ingroup prototype. Thus, overall, I showed incidental negativity increased the perceived fit of non-negative/safe exemplars to the ingroup prototype. Exemplar prototypicality is a key component of category activation (Bruner, 1957; Medin & Smith, 1981; Locke et al., 2005) and a key gatekeeper to generalised changes towards groups (Brown & Hewstone, 2005; Rothbart & John, 1985). As such, prototypicality shifts as a result of incidental negativity have critical implications for what stereotypes are applied to an exemplar and more broadly for group attitudes and stereotypes.

Endnotes

1. I was not interested in participants own allocation to group membership, apart from the perspective of the within subject target group factor. Hence participant's group membership did not enter power considerations when designing this study. I included participant's group membership (under-estimator/over-estimator) in the analysis to ensure the random allocation of participants into groups was successful and did not result in any stimulus/group specific effects.

2. Similar to the previous chapters, a speeded sorting task was included as part of this study, but due to a programming error these data could not be analysed.

3. A chi-square test of independence was performed to examine the relationship between contingency awareness and the face stimuli that were counterbalanced. Face stimuli were counterbalanced using four different groups that varied in the face paired with the electrical stimulation and the background colour. The relationship between these two variables was non-significant, X^2 (3, N = 54) = 2.67, p = .446. This non-significant result suggests my post-hoc grouping of contingency awareness did not undo face stimuli counterbalancing.

4. One participant was excluded from this analysis as SCR data was not available due to equipment failure.

5. Three participants in the ingroup target ethnicity condition and 2 participants in the outgroup target ethnicity condition were excluded from the analysis due to missing data.

6. A chi-square test of independence was performed to examine the relationship between contingency awareness and the face stimuli that were counterbalanced. As the target ethnicity group used different facial stimuli, I checked the relationship separately for each group. The ingroup target ethnicity group produced a significant relationship indicating the post-hoc grouping may have undone my counterbalancing X^2 (1, N = 32)

= 5.24, p = .022. The outgroup target ethnicity group relationship between these two variables was non-significant, X^2 (1, N = 31) = .015, p = .901. This non-significant result suggests my post-hoc grouping of contingency awareness in the outgroup target ethnicity did not undo face stimuli counterbalancing.

References

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*(1), 5-25.

Bless, H., & Schwarz, N. (1998). Context effects in political judgement : Assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology*, (28), 159–173.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429-444.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 37 (pp. 255-342). San Diego, CA: Elsevier.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*(2), 123–152.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science, 280*, 77–81.

Coats, S., Latu, I. M., & Haydel, L. A. (2007). The facilitative effect of evaluative fit on social categorization. *Current Research in Social Psychology, 12,* 54 – 67.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorisation of multifaceted stimuli. *Journal of Personality and Social Psychology*, *77*(5), 927–941.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice, *Emotion, 9*(4), 585-591.

Deffenbacher, D. M., Park, B., Judd, C. M., & Correll, J. (2009). Category boundaries can be accentuated without increasing intergroup bias. *Group Processes and Intergroup Relations*, *12*(2), 175-193.

Delgado, M. R., Labouliere, C. D., & Phelps, E. A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Social Cognitive and Affective Neuroscience, 1*(3), 250-259.

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes, *Psychological Science, 15*(5), 319-324.

Dobbs, M., & Crano, W. D. (2001). Outgroup accountability in the minimal group paradigm: Implications for aversive discrimination and social identity theory. *Personality and Social Psychology Bulletin, 27*(3), 355-364.

Doosje, B., Haslam, S. A., Spears, R., Oakes, P. J. Koomen, W. (1998). The effect of comparative context on central tendency and variability judgements and the evaluation of group characteristics.

Doosje, B., Spears, R., Ellemers, N., & Koomen, W. (1999). Perceived group variability in intergroup relations: The distinctive role of social identity. *European Review of Social Psychology, 10*(1), 41-74.

Esteves, F., Parra, C., Dimberg, U., & Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli, *Psychophysiology, 31*, 375-385.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category–based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1-74.

Flykt, A., Esteves, F., Ohman, A. (2007). Skin conductance responses to masked conditioned stimuli: Phylogenetic/ontogenetic factors versus direction of threat?, *Biological Psychology, 74*, 328-336.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18*(3), 232-239.

Harwood, J., Joyce, N., Chen, C-Y., Paolini, S., Xiang, J., & Rubin M. (2017). Effects of past and present intergroup communication on perceived fit of an outgroup member and desire for future intergroup contact. *Communication Research, 44* (4), 530-555.

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The social identity perspective: Intergroup relations, self-conception, and small groups. *Small Group Research, 35*, 246-276.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researchers handbook.* United States of America: Pearson Prentice Hall.

Kruschke, J. K. (2011). Models of attention learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 120-152). New York: Cambridge University Press.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377-1388.

Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, *22,* 549–569.

Locke, V., Macrae, C. N., & Eaton, J. L. (2005). Is person categorization modulated by exemplar typicality? *Social Cognition*, *23*(5), 417–428.

Lovibond, P. F., Saunders, J. C., Weidemann, G., & Mitchell, C. J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *The Quarterly Journal of Experimental Psychology, 61*(8), 1199–1216.

Mackie, D. M., & Smith, E. R. (1998). Integrative relations: Insights from a theoretically integrative approach. *Psychological Review, 105*, 499-529.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology, 45*(6), 1303–1307.

Marques, J. M., & Paez, D. (1994). The 'black sheep effect': Social categorization, rejection of ingroup deviates, and perception of group variability. In W. Strobe, & M. Hewstone (Eds), *European Review of Social Psychology* (Vol 5, pp. 37-68). Chichester, UK: Wiley.

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The "black sheep effect": Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*, 1–16.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(4), 241–253.

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them". *Journal of Personality and Social Psychology, 99*(1), 62-77.

Mineka, S., Ohman, A. (2002). Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry, 52*, 927-937.

Navarrete, C. D., McDonald, M. M., Asher, B.D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an out-group face in a minimal group context. *Evolution and Human Behavior, 33* (5), 590-593.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 211-233.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotypes and social reality.* Oxford: Blackwell.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30,* 125-144.

Olson, M. A, & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality & Social Psychology Bulletin*, *32*(4), 421–433.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*(5735), 785–787.

Paolini, S., Harris, N. C., & Griffin, A. S. (2016). Learning anxiety in interactions with the outgroup: Towards a learning model of anxiety and stress in intergroup contact. *Group Processes and Intergroup Relations*, *19* (3), 275-313.

Page, M. M. (1973). On detecting demand awareness by postexperimental questionnaire. *The Journal of Social Psychology*, *91*(2), 305–323.

Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: who triggers the black sheep effect? *Journal of Personality and Social Psychology*, *99*(1), 107–119.

Richeson, J. A, & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of personality and social psychology*, *89*(4), 517–30.

Rosch, E. (1978). Principles of categorization. In E. Rosch & Lloyd, B. B. (Eds.), *Cognition and categorization* (pp. 27-48). Oxford, England: Lawrence Erlbaum; England.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effect of intergroup contact. *Journal of Social Issues*, *41*, 81-104.

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review, 2*, 40-62.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis)categorization of unattractive brides and attractive prostitutes. *Experimental Psychology, 55*(3), 182–188.

Ryan, C. S., Robinson, D. R., & Housmann, L. R. M. (2004). Group socialization, uncertainty reduction, and the development of new members' perception of group variability. In V. Yzerbyt, C.M. Judd & O. Corneille (Eds), *The psychology of group perception: Perceived variability, entitativity, and essentialism* (pp. 206-219). New York: Taylor and Francis.

Schneider, D. J. (1991). Social cognition. *Annual Review of Psychology, 42*, 577-561.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2,* 307-320.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition, 8*, 243-262.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-177.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology, 54*(2), 101-114.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Van Vugt, M., & Park, J. H. (2010). The tribal instinct hypothesis: Evolution and the social psychology of intergroup relations. In S. Sturmer, & M. Snyder (Eds.), *The* *psychology of prosocial behaviour: Group processes, intergroup relations, and helping* (pp.13-32). Malaysia: Wiley-Blackwell.

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science*, *326*(5960), 1711–1714.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1160-1173.

Chapter 6: Comparative Analysis of Appetitive and Aversive Learning

Previous experimental chapters quantified the effects of aversive conditioning on perceived prototypicality towards ingroup and outgroup exemplars. These studies established that pairing an outgroup face with a mild electro-tactile stimulation leads the exemplar being perceived as *more* outgroup-like post-extinction compared to pre-conditioning. In contrast, safe ingroup faces (never paired with the electrical stimulation) were perceived as being *more* ingroup like post-conditioning compared to pre-conditioning. The direction of the prototypicality shifts aligns with evaluative-fit and emotion-fit explanations because outgroups are typically perceived as being negative and anxiety provoking and ingroups positive and non-anxiety provoking (Cottrell & Neuberg, 2005; Dasgupta, DeSteno, Williams & Hunsinger, 2009; Tajfel, Billig, Bundy & Flament, 1971; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). This study aimed to distinguish between evaluative-fit and emotion-fit mechanisms, as well as show changes in outgroup exemplar prototypicality in the opposite direction to that previously found when paired with a positively valenced stimulus.

In this chapter, I expanded on the previous research in two ways. Firstly, I replaced the mild electro-tactile stimulation with a different aversive unconditioned stimulus, money loss. With this manipulation, I attempted to retain the aversive nature of the unconditioned stimulus whilst removing the anxiety-driven/threatening component. Removing the anxiety-driven/threatening component from the reinforcer provided a means of exploring the nature of the processes underpinning changes in perceived exemplar prototypicality. In previous chapters the aversive unconditioned stimulus used conflated negative valence and anxiety. As a result, I was unable to determine if an evaluative-fit mechanism (driven by negative valence) or emotion-fit mechanism (driven by anxiety) were driving exemplar prototypicality shifts. By largely (but not completely) removing anxiety from my unconditioned stimulus and leaving negative valence intact, I was in a position to more accurately test the effect responsible for driving shifts of exemplar prototypicality.

Secondly, I tested the effects that a positive unconditioned stimulus, money gain, had on exemplar prototypicality. This manipulation enabled me to test the prediction that an association with positivity, rather than negativity, causes outgroup exemplars to become *less* outgroup like. Support for this prediction would indicate that an evaluative-fit mechanism is driving exemplar prototypicality shifts, advancing theories on dissolving intergroup categorisation. Practically, associating an outgroup member with a positively valenced stimulus would help to reduce intergroup categorisation and reliance on stereotypes.

Extensions of Previous Research

An evaluative-fit interpretation of my previous results aligns with social psychological theories (Tajfel & Turner, 1979; Turner et al., 1987). According to these theories self-esteem is partially derived from membership within a social group (Rubin & Hewstone, 1998). These theories are valence-focused in that the ingroup is associated with positive valence because it reflects on an individual's self-perception. The alliance to the ingroup results in the outgroup being typically associated with negative valence (Reynolds, Turner, & Haslam, 2000). Essentially, assigning different valences to ingroup and outgroup members helps boost an individual's self-esteem. My previous chapters demonstrated an outgroup exemplar paired with negative valence was perceived as being more prototypical of the outgroup after conditioning because the exemplar aligned more closely with the negative status of the outgroup. Similarly, ingroup members who were not paired with the aversive stimulus were perceived as being more ingroup-like because they aligned more closely with the positive status of the ingroup. Thus, the direction of exemplar prototypicality shifts in my previous chapters aligns with evaluative-fit mechanisms.

An emotion-fit interpretation can also be applied to my previous results. An emotion-fit interpretation aligns with an evolutionary account that suggests previous results are driven by group-relevant anxiety and threat rather than valence. It is argued that humans fear outgroup members because of the threat they pose to the ingroup (Van Vugt & Park, 2010). For example, threats could be in the form of competition for food, territory or sexual partners, and these threats are what led to the outgroup association with anxiety and threat-related emotions. Cottrell and Neuberg (2005) distinguish further between threats and suggest different groups elicit different threat based emotions based on the type of threat an outgroup posed. For example, in the US context African Americans and Mexican Americans are associated with physical threat such as anger and anxiety, whilst gay men are associated with a different kind of threat through disease and elicit disgust based emotions (Cottrell & Neuberg, 2005; Dasgupta et al., 2009; DeSteno, Dasgupta, Bartlett & Cajdric, 2004). Thus, outgroups are associated with a specific type of threat based and activation of these emotions will result in category based judgements that align more closely with these groups (i.e., Black exemplars are associated with anxiety; Cottrell & Neuberg, 2005). From this perspective, my previous chapters demonstrated outgroup exemplars paired with an anxiety provoking stimulus were perceived as more outgroup-like because the exemplar aligned more closely with the group’s perceived threat. Similarly, ingroup exemplars who were not paired with the anxiety provoking stimulus were perceived as being more ingroup-like as they aligned more closely with the safety status of the ingroup. Thus, the direction of exemplar prototypicality shifts in my previous chapters could equally align with emotion-fit mechanisms.

Better discrimination between psychological processes driving shifting prototypical representations will contribute to our understanding of the key determinants of outgroup stereotypes and might provide important applied benefits. For example, understanding the exact mechanism involved in shifting exemplar prototypicality will lead to better developed intervention strategies that reduce stereotypical categorisation and improve the quality of stereotypical responses towards different groups. Therefore, an aim of this chapter is to gain a better understanding of the mechanism that drives exemplar prototypicality shifts.

A logical extension of my previous research is to determine whether an exemplar’s perceived prototypicality will shift in the opposite direction after an exemplar is paired with an appetitive stimulus. If I can show an exemplar’s perceived prototypicality shifted in the opposite direction following appetitive conditioning, it could lead to new intervention strategies that reduce negative stereotypes towards certain groups. Therefore, another aim of this study was to determine if an outgroup exemplar’s perceived prototypicality would shift and be perceived as less outgroup-like following appetitive conditioning.

In summary, Chapter 6 aims to: (1) Experimentally distinguish between evaluative-fit and emotion-fit mechanisms as the underlying factor involved in exemplar prototypicality shifts; (2) Determine if exemplar prototypicality can be shifted in the opposite direction to that previously found following an association with an appetitive stimulus.

Selecting a Suitable Experimental Paradigm

To address my two research aims an alternative unconditioned stimulus to the previously used mild electro-tactile stimulation was used. I looked more broadly into evaluative conditioning research, which investigates the conditions under which the valence of an unconditioned stimulus becomes associated with a neutral stimulus after repeated presentations (De Houwer, 2009; Walther, Nagengast & Trasselli, 2005; Walther, Well & Dusing, 2011). Possible alternative unconditioned stimuli included odour (Gottfried, Doherty & Dolan, 2002), food (Baeyens, Eelen, Van Den Bergh & Crombez, 1990), sex (Both, Laan, Spiering, Nilsson, Oomens & Everaerd, 2008), money (Breiter & Rosen, 1999; Delgado, Labouliere & Phelps, 2006; Elliot, Friston & Dolan, 2000) and valenced pictures (Walther & Nagengast, 2006). To select a negatively valenced unconditioned stimulus I considered how difficult and practical it would be to implement in the laboratory (Franken, Huijding, Nijs & van Strien, 2011). Furthermore, to address the second research aim the positively valenced unconditioned stimulus reinforcer needed to be of similar magnitude to the negatively valenced unconditioned stimulus. Matching the magnitude of effect between a negative and positive valenced reinforcer as closely as possible was necessary to ensure both unconditioned stimuli had similar psychological significance (Hermann, Ziegler, Birbaumer, & Flor, 2000; Martin-Soelch, Linthicum, & Ernst, 2007). To this end I chose money loss/gain as the unconditioned stimulus.

My research paradigm was based on Delgado et al. (2006) research, which investigated the effectiveness money had as an unconditioned stimulus during conditioning. In Delgado and colleagues' study participants accrued money in a computerised card guessing game. The objective of the card guessing game was to guess whether the value of a card was higher or lower than the number five. Following the card guessing game, conditioning occurred. During conditioning participants lost money to a blue square and maintained their balance to a yellow square. Results revealed higher skin conductance responses towards the blue square paired with money loss, as opposed to the yellow square that resulted in no balance change. Therefore, this study indicated money can be used as an unconditioned stimulus/reinforcer during conditioning studies.

The current study adapted the procedure used in Delgado et al. (2006) research. The card guessing game was incorporated into conditioning and framed as a study on gambling. Pictures of outgroup exemplars replaced geometric shapes used in Delgado et al. research because a social context was investigated in this research. I expected valence transference observed with geometric shapes to be extended to outgroup exemplars. Although money gain was not investigated in Delgado et al., I expected winning money against an outgroup exemplar to be associated with positive valence similar to losing money causing negatively valenced associations.

During conditioning, participants randomly allocated into the win (appetitive - positive) conditioning treatment always won money against a paired outgroup exemplar. Participants allocated into the lose (aversive - negative) conditioning treatment always lost money against a paired outgroup exemplar. In both conditions, another outgroup exemplar, the unpaired exemplar, always resulted in a draw (neutral outcome) and there was no change in participants' monetary balance. Although the unpaired exemplar was paired with tying, it will be referred to as the 'unpaired exemplar' throughout this study because I assumed the exemplar was not paired with a valenced reinforcer. Self-reported prototypicality and a speeded sorting task (Richeson & Trawalter, 2005) were used to measure the perceived prototypicality of the outgroup face at pre-test and at post-extinction.

The use of an anxiety free reinforcer addressed the first aim of the study. I hypothesised that an association with negative valence alone, rather than anxiety coupled with negative valence, would be sufficient to change the way participants perceived the outgroup stimuli in line with evaluative-fit. I predicted that outgroup exemplars paired with money loss would be perceived as more outgroup-like post-extinction than pre-test because negative valence is typically associated with outgroup perceptions (Aberson, Healy & Romero, 2000; Tajfel et al., 1971). Hence, exemplars paired with losing money were expected to be rated higher in prototypicality and sorted more quickly at post-extinction than at pre-test compared to the unpaired exemplar because the paired outgroup exemplar are perceived as more typical of the outgroup (Richeson & Trawalter, 2005; Ruys, Dijksterhuis & Corneille, 2008).

The second aim of this study required the use a positively valenced reinforcer in order to investigate whether an outgroup exemplar could be perceived as less outgroup-like following an association with a positively valenced stimulus. I predicted that the outgroup exemplar paired with winning money would be perceived as less outgroup like because positivity is atypical of the outgroup in line with evaluative-fit (Aberson et al., 2000; Locke, Macrae & Eaton, 2005; Tajfel et al., 1971). A shift in perceived prototypicality away from the group prototype would be indicated by lower self-reported prototypicality ratings and slower reaction times to sort the face at post-extinction relative to at pre-conditioning.

In changing the previous methodology and using money as an unconditioned stimulus coupled with a gambling cover story, additional complexities to the conditioning procedure were introduced that did not exist in previous experimental chapters and Delgado et al.'s (2006) research. I created a competitive situation in which participants won or lost money against outgroup opponents. My gambling game is considered a zero sum competition because one player received money and another player lost money. Added complexities not present in previous chapters could influence conditioning; therefore, additional measures were included to account for this possibility. Evidence suggests that males and females behave differently in competitive situations and participants demographics and competitiveness was measured to possibly account for these differences (McDonald, Navarrete, & Van Vugt, 2012; Van Vugt, De Cremer, & Janssen, 2007; Van Vugt & Park, 2010).

In summary, I adapted Delgado et al. (2006) research paradigm for my research purposes involving aversive and appetitive conditioning. Money was used as my unconditioned stimulus because it allowed me to address my research aims in a practical manner: (1) Money loss in my experimental context is expected to be a negative but non-anxiety provoking reinforcer; (2) Money can be used in an aversive and appetitive context; (3) Money can be easily adapted into our current methodology.

Study 6.1

Method

Participants and Design

Participants were 92 students (56 male, 36 female; *M* = 22.14 years, *SD* = 5.01) from a large regional Australian university[1]. All participants reported being from a White, Anglo-Saxon background. Participants received partial course credit and a small monetary compensation (AUS $5) for completing the study. Participants were randomly assigned to one of two conditioning treatments (win treatment *N* = 46 [27 male, 19 female], lose treatment *N* = 46 [29 male, 17 female]). The conditioning treatment assignment refers to Task 2, Game 2, where participants were paired with either winning or losing money against an outgroup opponent.

Procedures and Measures

Participants were informed at the beginning of the study that the aim of this research was “to measure bodily activation during various gambling scenarios”. This cover story was used as a form of deception that provided to participants a believable cover story whilst limiting their knowledge about the true intentions of the study – I believed participants would respond differently if they were aware the gambling outcome was fixed and exemplar prototypicality was the true focus of research. To minimise response biases caused by repeated measures, participants provided pre-test data in the first laboratory session that was completed between five and 28 days (M = 7.25, SD = 2.76) prior to the second laboratory session. During the first laboratory session, all participants were seated in front of a computer and completed an online questionnaire and speeded sorting task. Black outgroup faces described in Chapter 3 were used in this study. This face set consisted of two prototypically Black exemplars, a 25% and 50% variation of these faces shifting progressively away from the Black prototype (towards an average of all Facegen faces), and two new prototypically Black faces.

As part of a larger on-line survey, participants indicated the extent to which each of the randomly ordered Black outgroup faces was prototypical of Black people in general (*prototypically Black*: 1 = *not at all*, 7 = *very much*). Next, participants rated how likeable they regarded the eight randomly ordered Black faces (1 = *very dislikeable*, 7 = *very likeable*). I included likeability measures in this paradigm to check whether a positive or negative association was acquired by the target faces (i.e., acquisition). Similarity was measured between each pair of faces presented throughout the study (1 = *Not at all similar*, 7 = *Very similar*). A General Evaluative Scale (Wright, Aron, McLaughlin-Volpe & Ropp, 1997) was included in this study, which consisted of seven opposite valenced terms, such as warm-cold and negative-positive, measured along a seven point Likert scale with the valence terms as the anchors. Items measuring individuals' competitiveness[2] (Houston, Harris, McIntire & Francs, 2002) and gender followed. A stereotyping measure was incorporated into this study following all self-reported measures during the second batch of testing; details of this measure and results for this measure are reported as a footnote due to the exploratory nature of this analysis[3].

Next, a speeded sorting task required participants to sort individual Black and White faces into two opposing categories ('Black' vs. 'White'), as quickly and as accurately as possible (Richeson & Trawalter, 2005; Ruys et al., 2008). The Black and White faces used for this task were the same as those described in the second experiment of Chapter 3 (8 Black faces: 2 targets, 4 generalisation faces, 2 new exemplars; 8 White faces: 2 targets, 4 generalisation faces, 2 new exemplars). During the sorting task, participants were required to sort these faces by pressing the green (left handed "S" key) or blue buttons (right handed "L" key) on the keyboard. Category labels were presented in the top corner of the screen and corresponded to the relevant key location. Each face was presented 14 times each and response keys and category labels were counterbalanced for half of the trials. The faces were inverted 25% of the time to increase task difficulty and engagement (Richeson & Trawalter, 2005).

The second laboratory session began with participants being seated in front of a computer screen and connected to the psychophysiological equipment. To measure arousal towards the faces, MLT116F (ADInstruments) skin conductance electrodes were attached to the distal phalanges of the first and second digits on the participant's left hand and connected to an ADInstruments Model ML116 GSR amplifier. An ADInstruments MLT1132 Piezo respiration belt was attached around the participant's chest to help monitor for any potential artefacts, such as deep breathing or coughing. Facial expressions were measured via muscle activity as a marker for positive or negative valence perception of the faces (i.e., electromyography (EMG)), but we had technical problems and as a result the data is of limited interpretability; see Appendix P7 for details of EMG method and results.

Verbal instructions pertaining to each of the gambling tasks were given next, which also appeared in more detail on the screen. To gamble, participants played an adapted version of the card guessing game used by Delgado et al. (2006), which required participants to guess if the value of a card was higher or lower than five (lower options 1-4, higher options 6-9). To guess the card was higher than five, participants were instructed to press the upwards arrow key on the keyboard and to guess the card was lower than five to press the downwards arrow key. A correct guess resulted in winning $10, whilst an incorrect guess resulted in losing $10. Participants were told $10 of “game money” equated to $1 of “real money” and their final balance at the end of the experiment would be paid to them. We expected this monetary incentive to increase task engagement.

Each participant underwent several tasks and played the card guessing under different conditions; see summary in Table 2. Participants began with Task 1, which was designed to give participants the opportunity to practice playing the card guessing game and accrue money for the remaining tasks. Task 2, Games 1 and 3, were played to collect pre-test and post-conditioning physiological data, while Game 2 served to present participants with pairings of outgroup stimuli and the experience of winning or losing and thereby creating conditioned responses to the outgroup faces. Finally, Task 3 aimed to extinguish conditioned responses acquired during Task 2, and Task 4 served to equalise participants' final balance between the winning and losing conditioning treatments as per ethics requirement.

Table 2. Summary of Task Aims and Design of Card Guessing Games

Task	Game	Aim	Face description	Feedback
Task 1		Practice and accrue money	No faces presented	At each trial
Task 2	Game 1	Pre-learning physiological responses	Faces were the audience to the participant.	Summary only presented at end of game
	Game 2	Conditioning	Faces were the opponent to the participant	At each trial
	Game 3	Post-learning physiological responses	Faces were the opponent to the participant	Summary only presented at end of game
Task 3		Extinction phase	Faces were the player and participant the audience	N/A
Task 4		Equalise participant's balance	No faces presented	At each trial

Task 1 consisted of 25 trials in which all participants won on 15 trials and lost on 10 trials (order randomised). Thus, all participants had a balance of $50 after Task 1. In this task participants were shown a card with a question mark located in the centre of the card on the computer screen and given three second to guess whether the card was higher or lower than five. Written feedback saying "You won" or "You lost" was presented for two seconds and the participants progressive balance was displayed in the top right of the screen. The inter-stimulus interval was two seconds. If participants failed to make a guess after three seconds, a prompt appeared asking participants to make their guess but to make it quicker next time. No faces were presented during Task 1.

Task 2, Game 1 was designed to collect pre-test physiological data and required participants to play the same card guessing game as in Task 1. However, participants were instructed that feedback would not occur until the end of the game and that they would be playing in front of an audience that they were to attend to – the audience was used as part of the cover story to allow faces to be presented in order to collect pre-conditioning data similar to my previous conditioning paradigms. After participants made their guess, a face (the 'audience') appeared for six seconds (Figure 20). This game consisted of 10 trials (ISI range 15-20 s, M = 17s) with two Black target faces presented twice and the 25% and 50% generalisation faces and two new Black faces presented once each. The feedback as to whether participants guessed correctly or incorrectly did not occur until the end of the game, so to prevent an association developing between the faces and gambling outcomes. At this point in time, all participants were informed they had guessed correctly on five trials and incorrectly on five trials and their balance remained at $50.

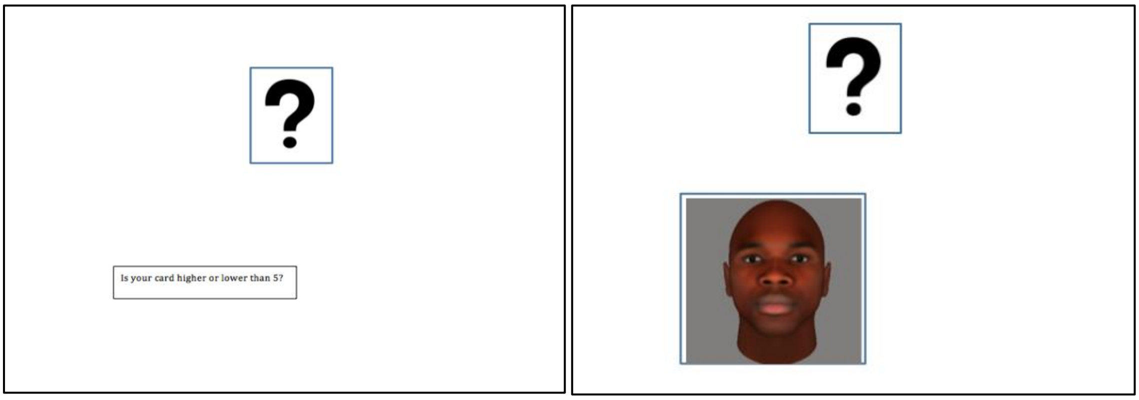


Figure 20. Example view of guess and exemplar presentation screens in Task 2, Game 1.

During Task 2, Games 2 and 3, participants were informed they would be playing against other fictional players. Conditioning treatment was manipulated during Task 2, Game 2. Participants were informed that if they guessed correctly and their opponent guessed incorrectly they would win $10; if they guessed incorrectly and their opponent guessed correctly they would lose $10; and if both guessed correctly or incorrectly a tie would occur and no money would be won or lost. Game 2 consisted of 10 trials (ISI range 15-20 s, *M* = 17s) that involved the presentation of two Black target faces. One target face was presented five times paired with winning (appetitive condition) or losing money (aversive condition) and the other target face was presented five times and always resulted in a tie. The conditioning treatment participants underwent (appetitive or aversive) was randomly allocated. The faces were presented for six seconds in a randomised order and the outcome pairings counterbalanced across faces. Outcome feedback was presented simultaneously beside the face during the last two seconds of the period of time the face was visible. Face and feedback then disappeared simultaneously. A progressive balance was presented in the top right of the screen to reinforce the salience of the pairings in Task 2 Game 2, and can be seen in Figure 21. By the end of Game 2, participants in the winning treatment had increased their monetary balance from $50 to $100 while participants in the losing treatment had decreased their monetary balance from $50 to $0.

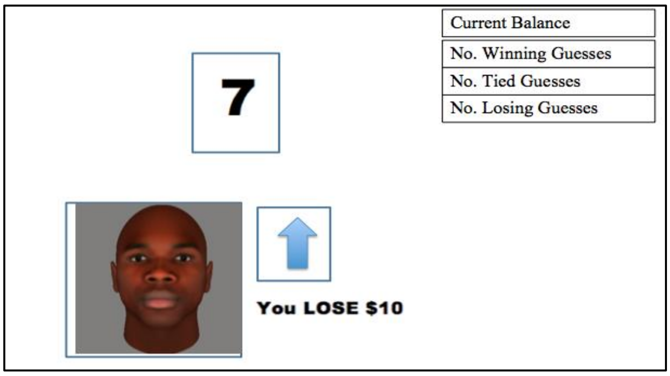


Figure 21. Example view of feedback screens in Task 2, Game 2.

Game 3 was designed to enable collection of post-test physiological data. As in Game 2, participants continued to play against the fictional opponent. However, as in Game 1, feedback occurred at the very end of the game and not after each guess. Instead, at the end of Game 3, participants were told they had won twice, lost twice and tied six times against their opponent. Thus, participants' monetary balance at the end of Game 3 was the same balance that participants had at the end of Game 2.

Evidence from the previous chapters indicated that changes in exemplar perceived prototypicality as a result of conditioning were at least in part attributable to repeated exposure in the absence of any aversive outcome, as experienced during extinction. Therefore, the next game, namely Task 3, was designed to provide such repeated exposure to face stimuli. Task 3 was framed as a game in which the participant acted as the audience to the fictional player. This time, however, participants were told that they would not be able to see the other player's guess or outcome. Each of the two target faces was presented repeatedly in the absence of any gambling outcome until neither face elicited a SCR across four consecutive presentations occurred (i.e. physiological arousal decreases had bottomed out). The number of presentations was set to a minimum of 5 and a maximum of 25 presentations per target face. Self-reported post-extinction prototypicality ratings and the speeded sorting task were collected once Task 3 had been completed, using the same online questionnaire and speeded sorting task described for the first laboratory session. In addition to these measures, I also tested for contingency awareness using the same procedure described in Study 3.2. Twenty nine participants were classified as contingent aware (15 win treatment; 14 lose treatment) and 63 as contingent unaware (31 win treatment; 32 lose treatment).

Finally, all participants played Task 4, which was identical to Task 1 in order to ensure participants left with the same money balance (as per ethics requirement). To achieve this, participants in the winning conditioning treatment were made to lose money and participants in the losing conditioning treatment were made to win money. All participants were debriefed and left the laboratory session with AUS$5.

Results

Checking Effective Acquisition and Extinction

To determine if effective acquisition and extinction occurred, I first examined participant's skin conductance responses (SCRs). SCRs were recorded and scored following standard guidelines that were described in Study 3.1 (Boucsein, Fowles, Grimnes, Ben-Shakhar, Roth, Dawson, et al., 2012; Fowles, Christie, Edelberg, Grings, Lykken & Venables, 1981). Briefly, acquisition was checked for using first interval SCRs with a minimum amplitude 0.05 μS calculated and averaged across two presentations for each of the two target faces immediately prior to, and after acquisition. To measure extinction, the same process was used but scores were calculated for the first block of extinction (first two presentations of the paired and unpaired face respectively) and the ten trial block (the last two presentations of the paired and unpaired face respectively up to the tenth trial – minimum number of trials all participants received to satisfy extinction criteria).

To check for effective acquisition, I performed a 2 conditioning treatment (win vs lose) x 2 face type (paired/unpaired face) x 2 time (pre- acquisition /post- acquisition) mixed model ANOVA with face type and time as the repeated measures on skin conductance responses. Analysis was restricted to 83 participants due to a technical error in SCR data collection. The analysis revealed a marginally significant face type x time x conditioning treatment interaction, which was followed up by analysing each conditioning treatment group separately, F (1, 81) = 3.74, p = .057, η_p^2 = .04.

In the lose conditioning treatment a significant face type x time effect was observed, F (1, 41) = 4.33, p = .044, η_p^2 =.10. Participants who lost against a Black outgroup exemplar exhibited higher skin conductance activity towards the paired face at post-acquisition (M = 1.23, SD = .38) than at pre-acquisition (M = 1.11, SD = .19), t (41) = -2.12, p = .04. No changes in skin conductance were observed towards the unpaired face at post-acquisition (M = 1.13, SD = .20) relative to pre-acquisition (M = 1.14, SD = .26), t (41) = -.20, p = .843. Together, these results suggest participants in the lose conditioning treatment were successfully conditioned and the Black outgroup face associated with losing was perceived with more negativity compared to the unpaired face. No significant effects were observed in the win conditioning treatment, all p's > .131.

To check conditioned effects were extinguished, I ran a 2 conditioning treatment (win vs lose) x 2 face type (paired and unpaired face) x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR. SCR data during the first block of extinction (first two presentations of the paired and unpaired face) and the ten trial block (the last two presentations of the paired and unpaired face up to the tenth trial) was chosen as all participants were exposed to a minimum of ten extinction trials. As expected there was no main effect of face type, F (1, 52) = .59, p = .445, η_p^2 = .01 or face type x extinction block interaction F (1, 52) = .61, p = .440, η_p^2 = .01. These two results suggest there is no difference between the paired and unpaired face and conditioned effects were extinguished. Supplemental data analyses conducted on SCR suggests negativity was extinguished and are discussed extensively in Appendix P7.

I also checked for extinction of any conditioned responses with self-reported liking data collected pre-conditioning and post-extinction. A 2 conditioning treatment (win vs lose) x 2 face type (paired and unpaired face) x 2 time (pre-conditioning and post-extinction) mixed model ANOVA was conducted with face type and time as the repeated measures on the self-reported liking data. As expected, no significant interactions involving face type and time were observed, *p'*s > .227. This result suggests shifts in likeability of the exemplars did not change from pre-conditioning to post-extinction.

In addition to checking for acquisition and extinction effects, the inclusion of a General Evaluative Scale allowed me to assess group-level changes of evaluations from pre-conditioning to post-extinction (as opposed to face-specific changes). The General Evaluative Scale items were combined and averaged to form a reliable Black general evaluative index at pre-conditioning and post-extinction (Cronbach's alpha pre-test = .91, post-extinction = .90). I ran a 2 conditioning treatment (win vs lose) x 2 time (pre-conditioning and post-extinction) mixed model ANOVA with time as the repeated measure on participants general evaluative index of the Black group. I found a significant time main effect, F (1, 83) = 13.26, $p < .001$, $\eta_p^2 = .14$. The time main effect indicated that across both the win and lose conditioning treatments, the Black outgroup was perceived with a worse evaluation post-extinction (*M* = 5.01, *SD* = 1.01) than at pre-conditioning (*M* = 5.37, *SD* = .96). Thus, whilst no difference was found for face-specific evaluations post-extinction, this result suggests the group was perceived with worse evaluations.

In summary, SCR data suggests acquisition was successful in the lose conditioning treatment and the face paired with a losing result was perceived as being more negative compared to the face paired with a neutral result. In contrast, SCR data suggests acquisition was not successful in the win conditioning treatment. SCR data confirmed that any association developed throughout acquisition was extinguished.

Self-reported likability rating results of each face also confirmed face-specific evaluations were at similar levels to pre-conditioning. In contrast, the General Evaluative Scale results suggest that group-level evaluations were worsened following conditioning.

Testing Basic and Generalised Prototypicality Shifts

To test for prototypicality shifts I analysed self-reported prototypicality and sorting task data. These data were used to test two hypotheses. Firstly, an association between an outgroup exemplar and a negatively evaluative loaded stimulus (vs emotion loaded stimulus) was sufficient to change an outgroup exemplar's perceived prototypicality in line with evaluative-fit mechanisms and the exemplar would be perceived as more prototypical of the outgroup. Secondly, different conditioning treatments, namely win (positive association) and lose (negative association), would change the goodness of fit between opponent Black faces and their outgroup category in opposite directions in line with evaluative-fit mechanisms. I expected an association between an outgroup exemplar and positively evaluative loaded stimulus would shift outgroup exemplar prototypicality and they would be perceived as being less outgroup-like.

Basic prototypicality shifts were first tested for with self-reported prototypicality using a 2 conditioning treatment (win vs lose) x 2 face type (paired/unpaired face) x 2 time (pre-test/post-extinction) mixed ANOVA with face type and time as the repeated measures. A time main effect was observed, F (1, 90) = 10.17, p = .002, η_p^2 = .01. The time main effect indicates a non-associative prototypicality shift and both the paired and unpaired face, in both the win and lose conditioning treatment, were perceived as being more prototypical of the outgroup at post-extinction (M = 5.81, SD = .79) than at pre-test (M = 5.43, SD = 1.08). Unexpectedly, the three way interaction involving conditioning treatment, face type and time was non-significant, F (1, 90) = .01, p = .928, η_p^2 = .00. The non-significant interaction suggests that pre-test to-post-extinction changes in perceived prototypicality of Black opponent faces did not differ across participants that lost or won against their opponent. No other effects were significant, all p's > .127.[4]

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the paired and unpaired face in order to test for generalisation effects. A 2 conditioning treatment (win vs lose) x 2 face type (paired/unpaired face) x 3 generalisation gradient (target, 25%, 50%) x 2 time (pre-test/post-extinction) mixed model ANOVA with face type, generalisation gradient and time as repeated measures was conducted on the self-reported prototypicality data. A generalisation gradient main effect was found that showed target faces being rated as more prototypical (M = 5.61) than the 25% (M = 3.32) and 50% variation faces (M = 2.25), F (2,166) = 679.34, p < .001, ηp^2 = .89. The time main effect found for basic prototypicality shifts remained when generalisation faces were entered into the analysis, F (1, 83) = 11.63, p = .001, ηp^2 = .12. More importantly, a marginally significant face type x time x conditioning treatment interaction was detected and subsequently followed up by looking at each level of the conditioning treatment separately, F (1, 83) = 3.86, p = .053, ηp^2 = .04.

In the *win conditioning treatment* the generalisation gradient main effect found previously was again detected, F (2, 86) = 434.73, p < .001, ηp^2 = .91. A face type x generalisation gradient x time interaction was also detected, F (2, 86) = 3.29, p = .042, ηp^2 = .07. This effect was driven by the paired target face being rated as more prototypical of the outgroup post-extinction post-test (M = 5.74, SD = 1.05) than at pre-test (M = 5.36, SD = 1.41), t (45) = 1.97, p = .055; and the unpaired face 50% variation being rated as more prototypical of the outgroup post-extinction post-test (M = 2.31, SD

= 1.03) than at pre-test (*M* = 2.03, *SD* = .96), *t* (43) = 1.87, *p* = .068. All other lower order interactions were non-significant, all *p*'s > .129. In the *lose conditioning treatment* the generalisation gradient, *F* (2, 80) = 279.47, *p* < .001, ηp^2 = .88, and time main effect found in the larger model held, *F* (1, 40) = 10.66, *p* = .002, ηp^2 = .21. A face type x time interaction was also detected, *F* (1, 40) = 5.10, *p* = .029, ηp^2 = .11. The paired faces (target, 25% and 50% variation faces) were rated as being more prototypical post-extinction (*M* = 4.11, *SD* = .94) than at pre-test (*M* = 3.52, *SD* = .93), *t* (45) = 4.01, *p* < .001. The unpaired faces (target, 25% and 50% variation faces) were also rated as being more prototypical post-extinction (*M* = 4.00, *SD* = 1.07) than at pre-test (*M* = 3.65, *SD* = 1.00) but the difference was not as large, *t* (45) = 3.56, *p* = .001. Face variation results suggest whilst differences were found in the win conditioning treatment between two sets of faces, no meaningful pattern of results occurred. In the lose conditioning treatment the paired and unpaired faces (including variations) were perceived as being more prototypical, but larger differences were found for the paired faces.

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the paired face, unpaired face, or equally similar. The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 conditioning treatment x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on self-reported prototypicality. The first new exemplar was rated as being more prototypical at post-extinction (*M* = 6.24, *SD* = .79) than at pre-test (*M* = 5.73, *SD* = 1.18) as evident by a time main effect, *F* (1, 79) = 12.42, *p* = .001, ηp^2 = .13. A time x new face similarity interaction was also detected and followed up by looking at each level of face similarity separately, *F* (2, 79) = 3.27, *p* = .043, ηp^2 = .08. This effect was more prominent when the new exemplar was similar to the paired face, *F* (1, 29) = 11.41, *p* = .002, ηp^2 = .28. No other effects were detected in any of the face similarity groups, all *p*'s > .097. A time main effect was also detected for the second new exemplar and it was rated as being more prototypical at post-extinction (*M* = 6.55, *SD* = .54) than at pre-test (*M* = 5.97, *SD* = 1.21), *F* (1, 79) = 12.54, *p* = .001, ηp^2 = .14. All other effects for the first and second new exemplar were non-significant, all *p*'s > .597.

Reaction times obtained during the speeded sorting task were also analysed to investigate changes in exemplar perceived prototypicality. The procedure for analysing these data was the same as described in Study 3.2. Briefly, incorrect responses were excluded from the analysis and the mean reaction times for the correctly categorised paired and unpaired faces were log transformed to normalize the data (Richeson & Trawalter, 2005; Ruys et al., 2008). Responses quicker than 300 ms were excluded from the analyses and extremely long responses (> 3 SD) were rescored to the upper limit for each group. On average, participants incorrectly sorted faces 11.00% of the time (*SD* = 7.16); a one way ANOVA confirmed there was no systematic difference in errors as a function of the manipulations, all *p*'s > .324. The average time taken to sort faces were: The paired exemplar at pre-test 547 ms (*SD* = 92 ms) and at post-test 545 ms (*SD* = 91 ms); the unpaired exemplar at pre-test 558 ms (*SD* = 91 ms) and post-test 545 ms (*SD* = 88 ms).

Basic prototypicality shifts of the target faces were investigated for using a 2 conditioning treatment (win vs lose) x 2 face type (paired/unpaired face) x 2 time (pre-test/post-extinction) mixed model ANOVA with face type and time as the repeated measures on the sorting task data. Unexpectedly, the three way interaction involving conditioning treatment, face type and time was non-significant, *F* (1, 90) = .06, *p* =

.805, $\eta_p^2 = .01$. This indicates that pre-test to-post-extinction changes in perceived prototypicality of outgroup faces did not differ across participants that lost or won against their opponent. All other effects were also non-significant, all *p'*s > .140.[5]

I expanded this mixed model ANOVA used to test basic prototypicality shifts and included *variations* of the paired and unpaired face in order to test for generalisation effects. A 2 conditioning treatment (win vs lose) x 2 face type (paired/unpaired face) x 3 generalisation gradient (target, 25%, 50%) x 2 time (pre-test/post-extinction) mixed model ANOVA with face type, generalisation gradient and time as repeated measures was conducted on the sorting task data. A generalisation gradient main effect was found that showed target faces being rated as more prototypical via quicker reaction times (M = 2.73) than the 25% (M = 2.80) and 50% variation faces (M = 2.86), F (2,162) = 296.81, $p < .001$, $\eta p^2 = .79$. A marginally significant face type x time interaction was also observed, F (1, 81) = 3.83, $p = .054$, $\eta p^2 = .05$. Lower level analysis using paired samples *t*-tests failed to detect any effects, all *p*'s > .073.

To test whether prototypicality shifts generalised to *new exemplars*, participants similarity ratings were used to determine whether participants perceived the new exemplar as more similar to the paired face, unpaired face, or equally similar. The process used for computing similarity ratings was the same as described in Study 3.1. Individual analyses were carried out for each new Black face, using a 2 conditioning treatment x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on sorting task data. No effects were found for either new face, all *p*'s > .076.

In summary, a non-associative prototypicality shift was detected on self-reported prototypicality data and both the paired and unpaired face were perceived as being more prototypical of the outgroup post-extinction than at pre-test. The effect extended to face variations and I found participants in the lose conditioning treatment perceived the paired and unpaired faces (including variations) as being more prototypical of the outgroup. The first new exemplar was perceived as being more prototypical of the outgroup at post-extinction than at pre-test, with the effect being more prominent when the new exemplar was more similar to the paired face. The second new exemplar was also perceived as being more prototypical at post-extinction on self-reported data, irrespective of which face it was more similar too. Unexpectedly, no effects were found on the sorting task data. The non-significant interaction involving conditioning treatment means the prototypicality shifts detected on the self-reported measure were similar between the win and lose conditioning treatment and expected differences between positive and negative associations did not occur.

Exploring the Role of Contingency Awareness in Prototypicality Shifts

To analyse the effects contingency awareness[6] had a 2 conditioning treatment (win vs lose) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (paired/unpaired face) x 2 time (pre-test/post-extinction) mixed model ANOVA with face type and time as repeated measures was run on the self-reported prototypicality data. In this top level ANOVA no significant interactions were detected, all *p*'s > .107. Despite no significant interactions being detected, planned comparisons were carried out in line with Study 3.2 and I investigated for prototypicality shifts in contingent aware and unaware participants separately (Keppel & Wickens; 2004).

I repeated the earlier analysis separately for contingent aware (n = 29) and contingent unaware (n = 63) participants. *Contingent aware* participants were analysed first and a time main effect was detected, F (1, 27) = 7.41, p = .011, ηp^2 = .22. The time main effect was further qualified by conditioning treatment and I followed up this 2 way interaction by looking at each conditioning treatment group separately, F (1, 27) = 4.54, *p* = .042, ηp^2 = .14. There was no significant difference in exemplar perceived prototypicality observed in the *win conditioning* treatment, *p* = .641. In the *lose conditioning* treatment the non- associative prototypicality shift found both faces were more prototypical post-extinction (*M* = 6.26, *SD* = .63) than at pre-test (*M* = 5.44, *SD* = 1.01), *t* (13) = 3.05, *p* = .009. All other effects for contingent aware participants were non-significant, all *p*'s > .162. A time main effect was detected for *contingent unaware* participants and suggests a non-associative prototypicality shift where both the paired and unpaired faces were perceived as being more prototypical of the outgroup post-extinction (*M* = 5.83, *SD* = .98) than at pre-test (*M* = 5.41, *SD* = 1.37), *F* (1, 61) = 5.91, *p* = .018, ηp^2 = .08. All other effects for contingent unaware participants were non-significant, all *p*'s > .568.

I expanded this investigation to include the *face variations*. A generalisation gradient x time interaction was detected in the *win, F* (2, 28) = 4.42, *p* =.021, ηp^2 = .24 and *lose* conditioning treatments, *F* (2, 26) = 3.98, *p* =.031, ηp^2 = .23 among *contingent aware* participants. In the *win* conditioning treatment the effect was driven by the 25% and 50% variations being perceived as more prototypical post-extinction (25%: *M* = 3.51, *SD* = 1.05; 50% *M* = 2.40, *SD* = .85) than at pre-test (25%: *M* = 2.77, *SD* = .76; 50% *M* = 1.77, *SD* = .61), 25% *t* (14) – 2.70, *p* = .017; 50% *t* (14) = 2.94, *p* = .017. In the *lose* conditioning treatment the effect was driven by the target being perceived as more prototypical post-extinction (*M* = 6.26, *SD* = .63) than at pre-test (*M* = 5.44, *SD* = 1.02), *t* (13) = 3.05, *p* = .009. All other effects in the contingent aware analyses were non-significant, all *p*'s > .071. Next I investigated effects for *contingent unaware* participants. Importantly, a face type x time x conditioning treatment interaction was found in contingent unaware participants. I followed this interaction up by looking at each level of conditioning treatment separately. A generalisation gradient x time interaction was detected in the *win* conditioning treatment, F (2, 56) = 3.55, p = .035, ηp^2 = .112. However, lower level analyses revealed no significant differences, all p's > .126. In the *lose conditioning treatment* a face type x time interaction was detected, F (1, 26) = 6.37, p = .018, ηp^2 = .19. The paired face and its face variations were perceived as being more prototypical post-extinction (M = 4.36, SD = .98) than at pre-test (M = 3.60, SD = .93), t (31) = 4.61, p < .001. Although the difference was not as large as the paired face, the unpaired face and its face variations were also perceived as being more prototypical post-extinction (M = 4.37, SD = 1.16) than at pre-test (M = 3.72, SD = 1.07), t (31) = 3.40, p = .002.

Next I investigated the effects contingency awareness had on prototypicality shifts measured with the speeded sorting task. I ran a 2 conditioning treatment (win vs lose) x 2 contingency awareness (contingent aware vs contingent unaware) x 2 face type (paired/unpaired face) x 2 time (pre-test/post-extinction) mixed model ANOVA with face type and time as repeated measures on the sorting task data. Unexpectedly, no significant interactions involving contingency awareness were observed, all p's > .148. Despite no significant interactions being detected, similar to Chapter 3, I investigated contingent aware and contingent unaware participants separately in line with Keppel and Wickens (2004) planned comparison procedure. Unexpectedly, no effects were detected amongst contingent *unaware* and *aware* participants, all p's > .063. When the model expanded to include *variation* faces, no effects were detected either, all p's > .078.

In summary, contingent aware participants in the lose conditioning treatment displayed a non-associative prototypicality shift in the expected direction and both faces were perceived as being more prototypical of the outgroup post-extinction. The win conditioning treatment detected no change in exemplar prototypicality among participants aware of the relationship between the face and stimulus. No effects were detected on the sorting task. When the analyses were expanded to include face variations, contingent unaware participants' results suggest that the paired face and its variations were more prototypical post-extinction. Other generalisation effects found showed no consistent pattern in results to interpret.

Discussion

Summary and Interpretation of Key Findings

This research contributes to existing research in two ways. Firstly, I investigated prototypicality shifts in outgroup exemplars when exemplars were paired with losing money. Losing money is a negatively valenced experience that I did not expect to be anxiety provoking in my experimental context. The absence of an anxiety provoking unconditioned stimulus in the negatively valenced stimulus was intended to allow disentangling the relative contribution of fear and negative valence to changes in perceived prototypicality of outgroup faces. Based on an evaluative-fit explanation, I predicted a negatively valenced but non-anxiety provoking stimulus paired with an outgroup exemplar would shift exemplar prototypicality in a similar direction observed in previous chapters and that the paired outgroup exemplar would be perceived as more outgroup-like.

I found evidence for a non-associative prototypicality shift on the self-reported measure. Both the paired and unpaired face were perceived as being more prototypical of the outgroup post-extinction, than at pre-test. The direction of the prototypicality shift extended from target exemplars and generalised to face variations and new exemplars. This set of results suggests that in both the win and lose conditioning treatments, all Black exemplars presented in this study were perceived as being outgroup-like post-extinction. In contrast, the non-associative prototypicality shift was found only in the lose conditioning treatment for contingent aware participants and no shifts found in the win conditioning treatment. Furthermore, prototypicality shifts generalised to paired faces at a larger rate compared to unpaired faces. Together, these results provide evidence that non-associative prototypicality shifts occurred and were greater in the lose conditioning treatment when participants were aware of the face-outcome pairing.

Secondly, I predicted a positively valenced stimulus paired with an outgroup exemplar would shift exemplar prototypicality in the opposite direction to that found in previous chapters and the paired outgroup exemplar would be perceived as less outgroup-like. Unexpectedly I did not find a shift in exemplar prototypicality in the hypothesised direction. Therefore, there is no evidence to support my second hypothesis.

A face type x time x conditioning treatment interaction was found in the SCR data. In the losing conditioning treatment participants experienced higher SCR responses towards the paired face post-acquisition relative to pre-acquisition. SCR changes towards the unpaired face were not found in the losing conditioning treatment. This result aligns with Delgado et al. (2006) and provides evidence that conditioning occurred when participants lost money against the outgroup exemplar. However, no effects were found in the SCR data for participants in the win conditioning treatment. Likewise, no significant effect was found on self-reported likability or EMG measures. Therefore, there is partial support for effective conditioning.

To explain non-associative prototypicality shifts and conditioning in the expected direction for the lose conditioning treatment but not the win conditioning treatment I focus my attention to magnitude effects. In this research I attempted to match the magnitude between conditioning treatment groups by matching the amount of money won and lost between groups. Despite my attempt, the magnitude of losing money outweighed the magnitude of winning money. A plethora of research indicates responses to positive and negative outcomes are not equal (for a review see Baumeister, Bratslavsky, Finkenauer & Vohs 2001; Rozin & Roysman, 2001). These reviews highlight the unequal strength between positive and negative outcomes, demonstrating that negative outcomes are far more influential than positive outcomes. Evidence suggests the psychological significance of losing something (i.e., in this study losing money) is stronger than winning something (i.e., in this study winning money; Atthowe, 1960; Costantini & Hoving, 1973; Kahneman & Tversky, 1984; Myers, Reilly & Taub, 1961). Hence, in my study aversive conditioning was effective at developing associations between faces and the negatively valenced stimulus, whereas a positively valenced stimulus was not (Baumeister et al., 2001; Baeyens et al., 1990; Costantini & Hoving, 1973). The unequal magnitude between positive and negative experiences could explain why I found non-associative prototypicality shifts and conditioning when participants lost money, but not when money was won.

It is possible that the context of the gambling game in the lose conditioning treatment caused both the paired and unpaired faces to be perceived as being more prototypical of outgroup. The non-associative prototypicality shift could be due to both faces acting as a signal for no money. In my previous studies an unsafe face was paired with a mild electrical stimulation and the safe face never paired with an electrical stimulation. The distinction between receiving an electrical stimulation or not is significant, whereas in the current study the distinction was not. In the context of this study participants believed they could either win, draw or lose against an outgroup opponent. In the lose conditioning treatment, even though participants drew against an opponent and monetary balance remained stable the outcome could still be perceived as negative because no money was won. At the beginning of the study participants were told all money they won would be kept and even though drawing did not involve losing money, drawing did not involve winning money either. Thus, I believe any outcome that did not result in winning money might have been perceived as a negative experience. Therefore, this explanation would suggest my results are in line with an evaluative-fit mechanism because both the paired and unpaired face were associated with negativity and both perceived as being more prototypical of the outgroup (although I would expect a more marked response for the paired face).

Limitations and Future Research Ideas

The complex nature of the methodology could explain why the study did not work as intended. The research method was adapted from Delgado et al.'s (2006) procedure, in which conditioning between geometric shapes and money occurred. Geometric shapes do not share the same psychological characteristics as outgroup exemplars. Social groups are influenced by motives, emotions and membership within multiple groups that make group dynamics more complex than geometric shapes (Mackie & Smith, 1998). Modifications were made to the gambling game employed by Delgado et al. (2006) to make the experience feel more real to participants and applicable to social situations. For example, participants played numerous gambling games which all had a different set of instructions and "playing modes". In different games players were gambling by themselves in the presence of an outgroup exemplar, gambling against the outgroup exemplar with and without feedback and watching the outgroup exemplar gamble. This was a complex set of games for participants to follow and instructions may not have been fully understood.

Due to the complexity of the gambling game, I believe a high cognitive load and fatigue effects might have influenced participants' performance and prevented this study from working as intended (Schulz, Fischbacher, Thoni & Utikal, 2014). For example, research demonstrated that in a complex task, a participant's reaction time becomes slower because they become fatigued (Singleton, 1953; Welford, 1980). Prototypicality shifts on the sorting task may not be detected because participant's attention faded throughout the experiment, which impacted on reaction times on the sorting task.

To reduce the complexity of the gambling and reduce fatigue effects, a more simplified version of the game could be implemented. Money could still be accrued at the beginning of the experiment, similar to Task 1 used in this research. However, to simplify conditioning a gambling game would not be played. Money could simply be added or subtracted from the participant's balance when an outgroup exemplar appeared on the screen, replicating the CS+/unsafe face pairings used in Chapters 3, 4 and 5. Another outgroup exemplar would appear on the screen and the participant's balance would not change, replicating the CS-/safe face pairings used in previous chapters. A procedure similar to that described would allow winning or losing money to form a positive or negative association with the outgroup exemplar because the task is simplified and not distracting. Hence, outgroup exemplars would be associated with a positive or negative valence and perceived prototypicality could be assessed more accurately with effective conditioning.

Future research could address the psychological difference between appetitive and aversive conditioning. It is difficult to equate between a positive and negative experience and in this study I objectively equated the two with comparable increases and decreases in money. However, subjectively equating between the positive and negative experience is another approach that could be implemented by including a calibration tool prior to conditioning. The calibration tool would allow participants to allocate different weights to the positive and negative experiences they will experience during conditioning. By subjectively equating (vs objectively equating) the psychological magnitude between positive and negative experiences the amount of money needed to win in order to have the same psychological significance as losing money could be more accurately determined.

Future research could also distinguish between evaluative-fit and emotion-fit mechanisms using the methodology implemented in my previous chapters. I paired an electrical stimulation with an anxiety provoking group exemplar, which conflated the effects of negativity and anxiety. In order to distinguish between the two mechanisms I could pair the electrical stimulation with a non-anxiety provoking group exemplar. For example, pairing a homosexual exemplar with an anxiety provoking stimulus is not expected to shift exemplar prototypicality from an emotion-fit perspective because anxiety does not fit this group (Cottrell & Neuberg, 2005; Dasgupta et al., 2009). If a shift in prototypicality was detected in the homosexual exemplar, it could be concluded that negativity rather than anxiety was behind the shift.

Conclusion

To conclude, to the best of my knowledge this is the first research that has attempted to investigate the consequences that aversive conditioning with a negative but non-anxiety provoking stimulus had on an outgroup exemplar's perceived prototypicality. Furthermore, I attempted to investigate how an association with a positively valenced stimulus shifted an outgroup exemplar's perceived prototypicality. Results provided some support that an evaluative-fit mechanism may contribute to perceived exemplar prototypicality shifts in the absence of emotion through non-associative prototypicality shifts in the lose conditioning treatment. Further research is needed to determine whether evaluative-fit mechanisms underlie changes in outgroup exemplar prototypicality following an association with a positively valenced stimulus. Future research in this area should better equate the psychological significance between positive and negative stimuli used to form associations. If my hypotheses were accurate, this research could lead to new and better developed intervention strategies that reduce stereotypical categorisation and improve the quality of stereotypical responses towards different groups.

Endnotes

1. Research was conducted across two years using two participant samples. The first participant sample was recruited from SONA, an online platform used by the School of Psychology to recruit psychology and other interested students to research studies. The second participant sample recruited male participants from the university, as gender was expected to influence results and an uneven sample of males were collected in the first participant sample. The second sample of participants were matched with another participant on their level of competitiveness between the win and lose treatment groups. Identical measures were used between the two participant samples, except for the implementation of a stereotyping measure in the second participant sample. These participants completed a stereotyping measure pre- and post-test after all other pre-existing measures data were collected (see below for more information).
2. Participants' level of competitiveness was then assessed through Houston et al.'s revised competitive index (Houston, Harris, McIntire & Francis, 2002). This index results from nine items, such as "I like competition" and "I find competitive situations unpleasant" (1 =*strongly disagree,* 7 = *strongly agree*), and formed a reliable index (Cronbach's alpha = .94). I explored the effect the competitive index had on prototypicality data, however, no effect was found.
3. The stereotyping measure was incorporated during the second batch of testing and data was available for 35 participants (17 win conditioning treatment, 18 lose conditioning treatment). The stereotyping measure consisted of two questions that asked participants to estimate the probability that a series of faces (3 Black and 3 White faces presented one at a time) matched a description at pre-test and post-extinction. The description was designed to be stereotypical consistent with the

White group, whilst the other description was stereotypically consistent with the Black group. I compared changes from pre-test to post-extinction towards each individual face and a combined average of all 3 faces for each ethnic group. However, no effects were found, all *p*'s > .126. Despite finding no significant effect, sample sizes are small and further, well powered analysis should be investigated further.

4. An exploratory analysis was carried out to determine if gender moderated the relationship between conditioning treatment and pre-test to post-extinction changes in face prototypicality on self-reported data. No statistically reliable effects when entering gender into the model emerged or by exploring results separately for male and females.
5. An exploratory analysis was carried out to determine if gender moderated the relationship between conditioning treatment and pre-test to post-extinction changes in face prototypicality on sorting task data. No statistically reliable effects when entering gender into the model emerged or by exploring results separately for male and females.
6. A chi-square test of independence was performed to examine the relationship between contingency awareness and the face stimuli that were counterbalanced. The relationship between these two variables was non-significant, X^2 (1, N = 92) = .13, p = .714. This non-significant result suggests my post-hoc grouping of contingency awareness did not undo face stimuli counterbalancing.

References

Aberson. C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A Meta-analysis. *Personality and Social Psychology Review, 4* (2), 157-173.

Atthowe, J. M. (1960). Types of conflict and their resolution: A reinterpretation. *Journal of Experimental Psychology, 59*, 1-9.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5* (4), 323-370.

Baeyens, F., Eelen, P., Van den Bergh, O., Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation, 21*, 434-455.

Both, S., Laan, E., Spiering, M., Nilsson, T., Oomens, S., & Everaerd, W. (2008). Appetitive and aversive classical conditioning of female sexual response. *The Journal of Sexual Medicine, 5*(6) 1386-1401.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Breiter, H. C., & Rosen, B. R. (1999). Functional magnetic resonance imaging of brain reward circuitry in the human. *Annals of the New York Academy of Sciences, 877*, 523-547.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. A. (1986). Electromyography activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50* (2), 260-268.

Costantini, A. F., Hoving, K. L. (1973). The effectiveness of reward and punishment contingencies on response inhibition. *Journal of Experimental Child Psychology, 16*, 484-494.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice, *Emotion, 9*(4), 585-591.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning and Behavior, 37* (1), 1-20.

Delgado, M. R., Labouliere, C. D., & Phelps, E. A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Social Cognitive and Affective Neuroscience*, *1*(3), 250–259.

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes, *Psychological Science*, *15*(5), 319-324.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience, 20,* 6159-6165.

Gottfried, J. A., O’Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*(24), 10829-10837.

Franken, I. H. A, Huijding, J., Nijs, I. M. T., & van Strien, J. W. (2011). Electrophysiology of appetitive taste and appetitive taste conditioning in humans. *Biological Psychology*, *86*(3), 273–278.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18*(3), 232-239.

Hermann, C., Ziegler, S., Birbaumer, N., & Flor, H. (2000). Pavlovian aversive and appetitive odor conditioning in humans: Subjective, peripheral, and electrocortical changes. *Experimental Brain Research, 132*(2), 203–215.

Houston, J., Harris, P., McIntire, S., & Francis, D. (2002). Revising the competitiveness index using factor analysis. *Psychological Reports, 90*(1), 31-34.

Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39*, 341-350.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researchers handbook.* United States of America: Pearson Prentice Hall.

Locke, V., Macrae, C. N., & Eaton, J. L. (2005). Is person categorization modulated by exemplar typicality? *Social Cognition*, *23*(5), 417–428.

Mackie, D. M., & Smith, E. R. (1998). Integrative relations: Insights from a theoretically integrative approach. *Psychological Review, 105*, 499-529.

Martin-Soelch, C., Linthicum, J., & Ernst, M. (2007). Appetitive conditioning: Neural bases and implications for psychopathology. *Neuroscience and Biobehavioral Reviews*, *31*(3), 426–440.

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1589), 670–679.

Myers, J. L., Reilly, R. E., & Taub, H. A. (1961). Differential cost, gain, and relative frequency of reward in a sequential choice situation. *Jounral of Experimental Psychology, 62*, 357-360.

Navarrete, C. D., McDonald, M. M., Asher, B.D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an out-group face in a minimal group context. *Evolution and Human Behavior, 33* (5), 590-593.

Reynolds, K. J., Turner, J. C., & Haslam, S. (2000). When are we better than them and they worse than us? A closer look at social discrimination in positive and negative domains. *Journal of Personality and Social Psychology, 78*, 64-80.

Richeson, J. A., & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology*, *89*(4), 517–530.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5* (4), 296-320.

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review*, *2*, 40-62.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis)categorization of unattractive brides and attractive prostitutes. *Experimental Psychology*, *55*(3), 182–188.

Schulz, J. F., Fischbacher, U., Thoni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, *41*, 77-87.

Singleton, W. T. (1953). Deterioration of performance on a short-term perceptual-motor task. In W. F. Floyd and A. T. Welford (Eds.), *Symposium on Fatigue* (pp. 163-172). London: H. K. Lewis and Co..

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-177.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of social conflict. In W. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp.33-47). California: Brooks/Cole.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis. *Psychological Science*, *18*(1), 19–23.

Van Vugt, M., & Park, J. H. (2010). The tribal instinct hypothesis: Evolution and the social psychology of intergroup relations. In S. Sturmer, & M. Snyder (Eds.), *The psychology of prosocial behaviour: Group processes, intergroup relations, and helping* (pp.13-32). Malaysia: Wiley-Blackwell.

Walther, E., Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavoir Processes, 32* (4), 454-459.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19*, 175-196.

Walther, E., Weil, R., & Dusing, J. (2011). The role of evaluative conditioning in attitude formation. *Current Directions in Psychological Science*, *20*(3), 192-196.

Welford, A. T. (1980). Choice reaction time: Basic concepts. In A. T. Welford (Ed.), *Reaction Times* (pp. 73-128). New York: Academic Press.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, *73*(1), 73–90.

Chapter 7: General Discussion

Past research has found that exemplars paired with a negative/anxiety provoking stimulus evoke worsened evaluative and affective responses (Mallan, Sax, & Lipp, 2009; Olson & Fazio, 2006; Olsson, Ebert, Banaji, & Phelps, 2005; Weisbuch, Pauker, & Ambady, 2009). The research presented in this thesis extended past research and investigated changes of perceived exemplar prototypicality following an exemplar's association with a valenced/anxiety producing stimulus. I consider changes in exemplar prototypicality as changes in cognitive representations because prototypicality is a fundamental dimension in the cognitive representation of social groups (Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000). In a similar vein, changes in exemplar prototypicality are likely to impact stereotype formation and change (Coats, Latu & Haydel, 2007; Rothbart & John, 1985).

Broadly, I took an innovative and unique approach and used conditioning procedures to investigate changes in exemplar prototypicality. By combining two distinct branches of psychology I was in a position to determine the effects that contingency, time and valence had on a social psychological process. The seven studies reported in the four empirical chapters of this thesis provide initial evidence for the existence of a prototypicality shift effect and explored possible mechanisms for these changes in cognitive representations of ingroup and outgroup exemplars.

In this final chapter, target exemplar and manipulation check results from each experimental chapter are first summarised. Next, I discuss three key topics that the research presented within this thesis addresses: (1) Prototypicality shifts and their robustness; (2) Mechanisms underlying prototypicality shifts; (3) Generalisability of prototypicality shifts. For each topic, I begin by discussing complexities in understanding and interpreting the data, limitations and ideas for future research. Notwithstanding the additional complexities, I continue the discussion within each topic and explain to what extent results align with previous theory and research, which includes the broader implications my results have for each topic.

Summary of Results

The research presented within this thesis has a primary focus on measuring changes in perceived exemplar prototypicality. I measured participants' perceived exemplar prototypicality with self-reported prototypicality measures and a speeded sorting task. I also measured participant's perceived anxiety/likeability of exemplars as a manipulation check to determine if evaluations/emotions towards each exemplar changed following an association with a valenced/anxiety producing stimulus. Manipulation checks were measured through a self-reported questionnaire and skin conductance responses (SCRs).

To assist in understanding the complex set of results discussed in this thesis, a summary table is provided at the end of this section. Table 3 provides an overview of each study and the basic design. In particular, Table 3 summarises results for the basic prototypicality shifts and whether the shift was contingent-specific (associative) or non-associative, thereby assisting in appreciating recurrent dissociations between parallel measures of the same construct. I also included in Table 3 a summary of generalisation results for 25% and 50% face variations and new exemplars. The main focus of this research was to investigate prototypicality shifts of target exemplars involved in conditioning and, as a result, I excluded ancillary data from the summary table, such as exploratory analyses that involved contingency awareness. In addition to summarising target exemplar, face variation and new exemplar prototypicality shifts, Table 3 also summarises manipulation checks and whether valence/anxiety became associated with an exemplar after acquisition and extinguished following the extinction procedure. Table 3 is visible at the end of summary of results section.

Chapter 3 presents very first evidence that exemplars' perceived prototypicality shifted after (relative to before) aversive learning and explored possible underlying processes. In three studies a change in how Black outgroup exemplars were perceived by White Anglo-Saxon Australians was found. Study 3.1 randomly allocated participants to a direct or vicarious aversive conditioning procedure and measured perceived exemplar prototypicality following an extinction procedure. Participants were exposed to or witnessed another ingroup member receive a mild electro-tactile stimulation paired with one outgroup exemplar but not another outgroup exemplar. Negativity/anxiety towards the unsafe exemplar (but not the safe exemplar) developed through aversive conditioning and the association subsequently made ambiguous through repeated, non-reinforced presentations of the exemplar (extinction; see Table 3, Study 3.1 manipulation checks). Self-reported prototypicality measures of individual exemplars were collected after the extinction procedure. Results revealed that cognitive representations of the exemplar shifted after being paired with the aversive stimulus and the outgroup exemplar was perceived to be more prototypical of the outgroup category (but no changes were observed for the safe exemplar; i.e., a contingent specific prototypicality shift; See Table 3, Study 3.1 perceived prototypicality). Prototypicality shifts were of similar magnitude between direct and vicarious aversive learning conditions, providing initial evidence that exemplar's perceived prototypicality not only changed when one directly experiences an aversive pairing, but also when one merely witnesses a negative association occurring. Altogether, these shifts of perceived exemplar prototypicality are in line with evaluative/emotion-fit mechanisms because exemplars that were associated with negativity/anxiety were perceived as more outgroup-like.

In this first study, prototypicality shifts were assessed and detected post-extinction and on a self-reported measure. In Study 3.2, perceived prototypicality was measured at two different time points to provide a better understanding of the effect extinction had. In Study 3.2 perceived exemplar prototypicality was measured immediately after acquisition (but before extinction) or, after the extinction procedure. In addition to varying the time point prototypicality was measured, I also investigated whether prototypicality shifts were limited to more deliberate/self-reported processes or occurred at a more implicit level. Perceived exemplar prototypicality was again measured with self-reported prototypicality, but also with a speeded sorting task, which provided a more implicit measure of prototypicality (Richeson & Trawalter, 2005).

Manipulation checks in Study 3.2 detected contingent specific conditioning on the self-reported measure, but non-discriminative effects on SCRs. Results from Study 3.2 prototypicality measures suggested that, following aversive learning, the exemplar associated with negativity/anxiety (but not the safe exemplar) was perceived to be more prototypical of the outgroup after extinction, but not beforehand (after acquisition – before extinction), in both direct and vicarious learning conditions on the implicit measure (see Table 3, Study 3.2 prototypicality shifts). The implicit prototypicality shift was not mapped by the self-reported prototypicality measure and a non-associative prototypicality shift was detected instead on the explicit indicator. The implicit prototypicality shift suggests that extinction plays a role in shifting exemplar prototypicality because the effect was detected at post-extinction only, and not beforehand. Pre-post changes in the response latencies in the speeded sorting task suggested contingent specific shifts in exemplar prototypicality following aversive conditioning took place at an implicit level, in addition to an explicit level detected in Study 3.1. The results found on both explicit (Study 3.1) and implicit (Study 3.2) measures suggest contingent specific changes in exemplar's perceived prototypicality can be detected at different levels of cognitive processing. However, there are inconsistencies between the self-reported results between Study 3.1 and 3.2, and I discuss this and other dissociations in measurement later in this chapter.

Exploratory analyses in Study 3.2 suggested contingency awareness might be involved in perceived exemplar prototypicality shifts and Study 3.3 was developed to further investigate the effect contingency awareness had. Contingency awareness was manipulated in Study 3.3 via a masking procedure. A similar direct and vicarious aversive conditioning procedure with outgroup exemplars to Study 3.2 was used in Study 3.3. However, in this study, the negative/anxiety provoking association was prevented from entering conscious awareness because the outgroup exemplars were masked during acquisition. Exemplar perception was masked during acquisition by replacing the exemplar image with a mask after a short time period (17 to 27 milliseconds). This method limited participants' awareness of which exemplar was associated with the aversive stimulus. A non-associative prototypicality shift was observed in this study on both the explicit and implicit measures and both the unsafe and safe face were perceived as being more prototypical of the outgroup post-extinction (but not post-acquisition; See Table 3, Study 3.3 prototypicality measures). I suggest contingent specific exemplar prototypicality shifts did not occur because negativity/anxiety became associated with the Black skin from the mask, rather than a specific exemplar (non-discriminative conditioning; See Table 3, Study 3.3 manipulation checks). The mask contained similar characteristics to Black exemplars in that it followed the outline of a face and contained dark physiognomic features typical of Black exemplars. Rather than negativity/anxiety becoming associated with an individual exemplar, I suggest an association was created between the mask and the aversive stimulus. Thus use of the masking procedure, in addition to non-associative prototypicality shifts suggest that contingency awareness might be an underlying process involved in changing exemplar perception.

Together, the three studies from Chapter 3 provide initial evidence of shifts in the perceived prototypicality of outgroup exemplars as a result of aversive conditioning. First, there is some evidence that outgroup exemplars paired with an aversive stimulus were perceived as being more outgroup-like after aversive conditioning and extinction compared to an exemplar never paired with an aversive stimulus (although some prototypicality shifts were non-associative). Second, perceived prototypicality shifts were of a similar magnitude between direct and vicarious aversive conditioning. Third, the time-point that exemplar prototypicality was measured at (i.e., after acquisition or after extinction) and ancillary analysis suggest that extinction and contingency awareness are at least partially involved in perceived shifts of exemplar prototypicality. Given that the results in Chapter 3 suggest the implication of one or more processes active during the extinction procedure in shifts of exemplar prototypicality, the extinction process is what I chose to focus on next.

Chapter 4 investigated the role that extinction had in shifting exemplar prototypicality. Contingent specific prototypicality shifts in Study 3.1 (explicit measure) and 3.2 (implicit measure) and non-associative prototypicality shifts in Study 3.3 were found post-extinction and not immediately after acquisition. This pattern suggests that the mechanism(s) underpinning changes in perceived exemplar prototypicality might operate during extinction. In Study 4.1 participants underwent direct aversive conditioning following the same procedure used in the studies of Chapter 3 with the exception of extinction, which was manipulated in this study. One group underwent extinction as described in earlier studies and viewed repeated presentations of the two outgroup exemplars presented throughout acquisition in the absence of the aversive stimulus (individuality group). The other group (category membership group) repeatedly viewed two images of the outgroup exemplars, but the exemplar could not be identified. Rather, the face was scrambled to remove any facial features cues, leaving only the category membership cue (Black skin). Overall, a non-associative prototypicality shift occurred in both conditions and both the unsafe and safe face were perceived as being more prototypical of the outgroup (see Table 3, Study 4.1 prototypicality measures). However, when factoring in contingency awareness in the model, in both groups exemplar prototypicality shifted in the direction of greater outgroup likeness only when the exemplar was paired with the aversive stimulus, relative to when the exemplar was never paired with the aversive stimulus. This pattern of results suggests that repeated presentations of ‘intact’ outgroup exemplars during extinction was *not* necessary for shifts of exemplar prototypicality; repeated presentations of category membership cues (i.e., Black skin) were sufficient to observe non-associative shifts of exemplar prototypicality. However, contingent specific learning or some degree of contingency awareness is required for contingent specific prototypicality shifts. I came to this conclusion because contingent specific shifts were found only during the post-hoc allocation of participants into contingent aware groups based on their accurate (vs inaccurate) recognition of the face-stimulus pairing and not in the larger model disregarding contingency awareness.

Chapter 3 and 4 focused exclusively on outgroup exemplars. A natural logical extension was to investigate whether *ingroup* exemplar prototypicality would shift under aversive conditioning. In Chapter 5, two studies extended the scope of Chapters 3 and 4 by investigating perceived exemplar prototypicality shifts of ingroup, as well as outgroup exemplars. Chapter 5's methodology used a similar direct aversive learning condition to that reported in Chapters 3 and 4. However, target exemplars presented throughout conditioning included ingroup exemplars, as well as outgroup exemplars.

In Study 5.1 ingroup and outgroup target membership was established via minimal groups and I investigated the effects aversive learning had on responses to ingroup and outgroup exemplars. Manipulation checks found evidence of conditioning of the unsafe faces across both SCR and self-reported anxiety measures and SCR data confirmed successful extinction of physiological arousal. The results on my key measure of perceived prototypicality displayed a pattern consistent with evaluative/emotion-fit: The safe exemplar (never paired with the aversive stimulus) was perceived as being more prototypical of the ingroup. There was no change in prototypicality for the unsafe exemplar, suggesting a contingent specific prototypicality shift on the self-reported measure (see Table 3, Study 5.1 perceived prototypicality). Thus, ingroup exemplar prototypicality shifts were detected in this study and the safe exemplar was perceived to be more prototypical of the ingroup.

In Study 5.2, ingroup and outgroup target membership was established along ethnicity, based on White participants' exposure to White or Black exemplars. SCR manipulation checks suggested contingent specific conditioning in the ingroup condition, but non-discriminative conditioning in the outgroup condition. Similar to Study 5.1 successful extinction was evident on SCR data (see Table 3, Study 5.2 manipulation checks). Overall, non-significant results were detected on the perceived prototypicality measures; there was also a disparity in the trends across self-reported and sorting measures across the ingroup and outgroup conditions. A trend for a non-associative prototypicality shift was found in Study 5.2 in the outgroup condition but not in the ingroup condition on the self-reported measure. In contrast, a trend for a non-associative prototypicality shift was found in the ingroup condition but not the outgroup condition on the implicit measure of prototypicality (see Table 3, Study 5.2 perceived prototypicality). Results were non-significant in Study 5.2, but the different trends observed between the ingroup and outgroup condition do align with evaluative/emotion-fit mechanisms: The trend data suggest that there was a (weak and measure specific) indication for ingroup exemplars to be perceived as less ingroup-like and outgroup exemplars to be perceived as more outgroup-like following an association with a negative/anxiety provoking stimulus. Together, Studies 5.1 and 5.2 provided preliminary evidence that negativity/anxiety might be inherently linked to the outgroup and safety/positivity to the ingroup and appraisals of evaluative/emotion fit (vs. misfit) might be implicated. Hence, there was a weak indication that ingroup exemplars associated with anxiety/negativity are psychologically excluded from the cognitive perception/representation of the ingroup; whilst safe/positive exemplars are psychologically included in the ingroup representation.

The next step in my research aimed to establish whether exemplar perceived prototypicality shifts were due to a negative stimulus being paired with exemplars (evaluative-fit), or whether it was due to the anxiety triggered by the electro-tactile stimulation (emotion-fit). Furthermore, I sought to determine whether exemplar prototypicality shifts found in the earlier studies could be reversed under appetitive conditioning. To this end, Study 6.1 manipulated the valence associated with the stimulus paired with an exemplar. In this study money loss (vs money gain) during a gambling game was used to associate outgroup exemplars and negative valence (or positive valence). In one condition participants lost money against an outgroup opponent (i.e., a negative but non-anxiety provoking event) and tied with another outgroup opponent. In the other condition participants won against an outgroup opponent (i.e., a positive event) and tied with another outgroup opponent. The manipulation of the type of reinforcer paired with exemplars removed as much as possible the anxiety component that was associated with the electro-tactile stimulation in previous studies. Additionally, it provided an opportunity to investigate the effects that a positively valenced stimulus had on outgroup exemplars' perceived prototypicality. SCR results suggest contingent specific conditioning occurred in the group who lost money, but no conditioning was detected in the group who won money. No prototypicality shift was detected on the implicit measure. A non-associative prototypicality shift was detected on the self-reported measure of prototypicality in Study 6.1 and both faces were perceived as being more prototypical of the outgroup (see Table 3, Study 6.1 perceived prototypicality) – importantly this effect was not qualified by the type of conditioning participants underwent. Hence, unexpectedly, conditioning treatment was not involved in the interaction and this suggests the direction of exemplar prototypicality shifts did not differ based on an association with a positive or negatively valenced stimulus (in the absence of anxiety-safety cues). Hence, null findings could not provide conclusive evidence to distinguish between evaluative and emotion-fit mechanisms. Furthermore, I did not establish whether exemplar prototypicality could be shifted in the opposite direction with a positive stimulus, but manipulation check results suggest conditioning never occurred and this could explain the null result.

Table 3: Summary of Studies' Designs and key Results on Manipulation Checks and Perceived Exemplar Prototypicality

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
3.1	Aversive conditioning WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Learning type (direct vs vicarious).	SCR: Unsafe face had higher SCR levels compared to safe face at post-acquisition (vs pre-acquisition). SR: Not available.	SCR and SR: Successful extinction.	Unsafe face perceived as being more outgroup-like than safe face post-extinction (vs pre-test).	*25% and 50% face variations*: 25% unsafe face perceived as being more outgroup-like than safe face post-extinction (vs pre-test). No effect detected for 50% variation faces. *New Exemplar*: New exemplars similar to safe face perceived as less outgroup-like in direct learning condition. No other effects detected.	Not available.	*25% and 50% face variations*: Not available. *New Exemplar:* Not available.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
3.2	Aversive conditioning WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Learning type (direct vs vicarious) and post-test position (after acquisition vs after extinction).	SCR: Both faces had higher SCR at post acquisition (vs pre acquisition). SR: Unsafe face rated higher in anxiety compared to the safe face at post-acquisition (vs pre-acquisition).	SCR and SR: Successful extinction.	Both unsafe and safe face perceived as being more outgroup-like post-test (post-acquisition and post-extinction vs pre-test).	*25% and 50% face variations*: No effect detected. *New exemplar:* No effect detected.	Unsafe face perceived as being more outgroup-like (vs safe face) post-extinction only (not after acquisition) – In both direct and vicarious learning.	*25% and 50% face variations:* No effect detected. *New Exemplar:* New exemplar 1 perceived as more outgroup like post-extinction (vs pre-test). Not qualified by similarity to target faces. No effect detected for new exemplar 2.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
3.3	Masked aversive conditioning WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Learning type (direct vs vicarious) and post-test position (after acquisition vs after extinction).	SCR: Both faces had higher SCR at post acquisition (vs pre-acquisition). SR: No effect found on self-reported measure.	SCR and SR: Successful extinction.	Both unsafe and safe face perceived as being more outgroup-like post-extinction only (not after acquisition) – In both direct and vicarious learning.	*25% and 50% face variations:* No effect detected. *New exemplar:* No effect detected.	Both unsafe and safe face perceived as being more outgroup-like post-extinction only (not after acquisition) – In both direct and vicarious learning.	*25% and 50% face variations:* Both unsafe (25% and 50%) and safe face (25% and 50%) perceived as being more outgroup-like post-extinction only (vs pre-test) – In both direct and vicarious learning. *New Exemplar:* New exemplar 1 perceived as more outgroup like post-extinction (vs pre-test). Not qualified by similarity to target faces. No effect detected for new exemplar 2.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
4.1	Aversive conditioning WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Type of extinction (individuality group, category membership group.	SCR: Unsafe face had higher SCR levels compared to safe face at post-acquisition (vs pre-acquisition). SR: Not available.	SCR: Successful extinction. SR: Higher anxiety ratings for both unsafe face and safe face post-extinction (vs pre-test).	Both unsafe and safe face perceived as being more outgroup-like post-extinction (vs pre-test).	*25% and 50% face variations:* No effect detected. *New exemplar:* New exemplar 1 and 2 perceived as more outgroup like post-extinction (vs pre-test). Not qualified by similarity to target faces.	Both unsafe and safe face perceived as being more outgroup-like post-extinction (vs pre-test). Effect found in both types of extinction.	*25% and 50% face variations:* Both unsafe (25% and 50%) and safe face (25% and 50%) perceived as being more outgroup-like post-extinction only (vs pre-test). *New Exemplar:* No effect detected.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
5.1	Aversive conditioning with minimal groups WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Target group (ingroup vs outgroup) and post-test position (after acquisition vs after extinction).	SCR: Unsafe face had higher SCR levels compared to safe face at post-acquisition (vs pre-acquisition). Further qualified by target group and higher SCR towards outgroup vs ingroup exemplars. SR: Unsafe face rated higher in anxiety compared to the safe face at post-acquisition (vs pre-acquisition). Not qualified by target group.	SCR: Successful extinction. SR: Higher anxiety ratings for both unsafe face and safe face post-extinction (vs pre-test).	Safe face perceived as being more ingroup-like (vs unsafe face) post-acquisition and extinction (vs pre-test). Safe face perceived as being less outgroup-like (vs unsafe face) post-acquisition and extinction (vs pre-test).	*Physiognomy and Background faces:* Safe face perceived as being more ingroup-like (vs unsafe face) post-test (vs pre-test). Not qualified by BS factors. *New exemplar:* Not available.	Not available.	*Physiognomy and Background faces:* Not available. *New Exemplar:* Not available.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
5.2	Aversive conditioning with minimal groups WS factors: Face type (unsafe, safe) and time (pre-, post-test). BS factors: Target ethnicity (White/ ingroup vs Black/ outgroup).	SCR: White/Ingroup: Unsafe ingroup face had higher SCR levels compared to safe ingroup face at post-acquisition (vs pre-acquisition). Black/outgroup: Both outgroup faces higher SCR at post acquisition (vs pre-acquisition).	SCR: Successful extinction. SR: Higher anxiety ratings for both unsafe face and safe face post-extinction (vs pre-test).	White/Ingroup: No effect. Black/outgroup: Trend for both unsafe and safe face to be perceived as being more outgroup-like post-extinction (vs pre-test).	*25% and 50% face variations:* All faces perceived as less prototypical of their group. Not qualified by target ethnicity. *New exemplar:* New exemplar 1 and 2 perceived as being more prototypical of their group post-extinction (vs pre-test). Not qualified by similarity to target faces.	White/ Ingroup: Trend for both unsafe and safe faces to be perceived as being less ingroup-like post-extinction (vs pre-test). Black/ outgroup: No effect.	*25% and 50% face variations:* No effect detected. *New Exemplar:* No effect detected.

Study	Design	Manipulation check		Exemplar prototypicality			
		Acquisition	Extinction	Self-reported		Sorting task	
				Target Exemplar	Generalisation	Target Exemplar	Generalisation
6.1	Aversive and appetitive conditioning WS factors: Face type (paired, unpaired) and time (pre-, post-test). BS factors: Conditioning treatment (win vs lose).	SCR: Paired face had higher SCR levels compared to the unpaired face at post-acquisition (vs pre-acquisition). This effect was only in the lose condition. No effect in win condition. SR: No effect on self-reported likeability.	SCR and SR: Successful extinction.	Both paired and unpaired outgroup faces perceived as being more prototypical of the outgroup post-extinction (vs pre-test). This effect was across both the win and lose condition.	*25% and 50% face variations:* No effect detected. *New exemplar:* New exemplar 1 perceived as being more outgroup like post-extinction (vs pre-test). Not qualified by similarity to target faces or win and lose condition. No effect detected for new exemplar 2.	No effect in either win or lose condition.	*25% and 50% face variations:* Both unsafe (25% and 50%) and safe face (25% and 50%) perceived as being more outgroup-like post-extinction only (vs pre-test) in lose condition. No effect detected in win condition. *New Exemplar:* No effect detected.

Note. BS-Between subjects; SCR- Skin conductance responses: differences in skin conductance responses were measured; SR-Self reported measures

In summary, the research presented in this thesis aimed to extend earlier research through investigations into intergroup categorisations as a function of aversive (Studies 3.1 – 6.1) and appetitive conditioning (Study 6.1). In the midst of evident complexities and (more than anticipated noise; see below), my research extends previous investigations by providing some preliminary evidence for a perceived shift in the cognitive representation of group exemplars following conditioning, as indexed by pre-post measures of exemplar prototypicality.

Across seven experiments, I showed effective acquisition on SCR responses in three studies (Studies 3.1, 4.1 and 5.1) and on self-reported anxiety in three other studies (Studies 3.2, 3.3 and 5.2). In this context, I detected evidence for contingent specific prototypicality shifts in the expected direction under aversive conditioning in three studies (Studies 3.1, 3.2 and 5.1) and non-associative prototypicality shifts, again in the expected direction under aversive conditioning, in four studies (Studies 3.3, 4.1, 5.2 and 6.1).

From the results' summary, it is evident that these novel effects on prototypicality indices manifested in the context of a number of significant disparities and inconsistencies. Therefore, prior to drawing some broader conclusions about the implications of these findings for theory and interventions, I first consider and discuss dissociations and inconsistencies of findings.

Prototypicality Shifts: Their Existence and Robustness

The main objective of this thesis was to investigate exemplar prototypicality shifts following an association with a negative/anxiety provoking stimulus. I begin this section by discussing the central issues associated with changes in perceived exemplar prototypicality and the robustness of the effect. Complexities and limitations surrounding the prototypicality shift effect are discussed, and ideas for future research proposed. Notwithstanding the complexities and limitations raised, I conclude this section by discussing implications that changes in perceived prototypicality have for theory and research.

One inconsistency in the results that is clearly visible from Table 3 is whether prototypicality shifts were associative (changes in one face but not another) or non-associative (both faces changed in their perception). Throughout this thesis I expected associative processes to facilitate contingent-specific prototypicality shifts and exemplars paired (vs. not paired) with negativity/anxiety to become more outgroup-like/less ingroup-like. Results from Studies 3.1, 3.2 and 5.1 provide support for associative processes. However, non-associative prototypicality shifts (or trends for a shift of this kind) were detected in Studies 3.3, 4.1, 5.2 and 6.1, which suggest non-associative processes might be implicated in prototypicality shifts too.

There are two explanations to account for this inconsistency between associative and non-associative processes. Firstly, it is possible that non-associative/contingent non-specific prototypicality shifts were observed because both exemplars became associated with negativity/anxiety, which would influence prototypicality shifts of both exemplars. For example, in Study 3.3 non-discriminative conditioning was detected on the SCR manipulation check, suggesting both exemplars were perceived as being more negative/anxiety provoking after acquisition. The non-associative prototypicality shift found in this study could therefore be explained by the evaluative/emotion-fit mechanism for both the unsafe and safe exemplar. Secondly, prototypicality shifts could result from sensitisation and habituation effects, two non-associative processes. During conditioning sensitisation may occur and participants respond to both faces similarly due the presence of an aversive stimulus in the conditioning block (Cevik, 2014).

Habituation is the process of diminishing a response by repeatedly presenting a stimulus, and this could mimic the extinction procedure in my research because negativity/anxiety diminishes following repeated exemplar presentations (Best et al., 2008). Despite inconsistencies between contingent specific and non-associative prototypicality shifts across studies, it is important to note that the shift (or trend for a shift) always occurred in the direction predicted by an evaluative-fit and emotion-fit mechanism.

Throughout my research there were a number of disparities of results between explicit and implicit measures. In studies 3.2, 5.2 and 6.1 perceived prototypicality shifts differed between measures, whilst in studies 3.2, 4.1, 5.1, 5.2 and 6.1 manipulation checks differed between measures (see Table 3). In some instances, one measure showed an effect whilst another measure would not; in other instances both measures detected an effect that was different in nature. Because the disparity between measures varied across individual studies and their unique designs without an obvious pattern, it is difficult to identify the key reason for these disparities. Differences in results between studies could be explained by the different ‘contexts’ each study was conducted in: Some studies involved manipulations during conditioning and others during extinction, other studies involved different designs with measures being collected at different time points using different stimuli (Blair, 2002; Milad, Orr, Pitman & Rauch, 2005). Furthermore, internal transitions in the procedure could also represent a change in context for participants. Vansteenwegen et al. (2005) provided evidence that the lighting of a room influenced participant’s responses because of the context perceived. Similarly, the removal of equipment for psychophysiological measures could represent a change in context for participants. As equipment was removed at different time points in each study (i.e. measures collection at different time points – after acquisition or after extinction), participants may have viewed data collection as a different phase of the study not associated with the previous context. Whilst these differences can explain why results differed between studies, they cannot explain the differences within studies. Apart from studies being conducted in different contexts, the underlying reasons for the disparity between measures within studies could be similar.

One possible explanation is that fatigue associated factors influenced results. Explicit measures are easily controlled for and extremely susceptible to the context in which they are presented (Gawronski & LeBel, 2008; Maass, Castelli & Arcuri, 2000; Milad et al., 2005). Whilst explicit measures are more susceptible to context, implicit measures are also susceptible to context (Wittenbrink, Judd & Park, 2001). It is possible that context biased responses on the explicit measure, but also on the implicit measure. Furthermore, testing sessions across all studies were long and participants could have become fatigued, limiting the amount of control they had over explicit measures. In most instances (apart from Studies 3.1-3.3) explicit measures were collected after the implicit measure, making these measures more susceptible to fatigue effects. Similarly, my implicit prototypicality measure required participants to make speeded decisions, and fatigue could influence the performance on this measure due to the length of the second testing session compared to the first testing session. The fact that participants were generally faster at post-test could be due to task repetition and does not limit the fatigue argument (Pashler & Baylis, 1991).

Another possible explanation that could impact results is motivational forces (Fazio & Olson, 2003; Hofmann, Gawronski, Gschwendner, Le & Schmitt, 2005; Maass et al., 2000). Participants could have been motivated to hide their responses due to social desirability or answered questions in a certain way causing response biases (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Whilst I attempted to minimise fatigue and response biases by having breaks throughout the study and not disclosing the aim of research to participants, it is possible that biases affected results and caused differences between measures. Overall, the disparity of results between measures clouds the reliability of the prototypicality shift and contextual changes, fatigue and motivational factors could possibly explain the variance between measures.

Furthermore, in Chapter 3 a secondary line of investigation was to determine the interplay between affect and cognitions in prototypicality shifts. To explore the interplay I manipulated the time point prototypicality was measured and collected responses before or after extinction. Extinction does not eliminate the association developed during acquisition, but rather renders the association ambiguous and effects can re-emerge (Bouton, 2002). For example, anxiety/negativity may re-emerge during the collection of prototypicality measures via AAB renewal (Bouton, 2002; Bouton & Ricker, 1994). Acquisition and extinction occurred in the same context in my research and participants might have perceived a context change when prototypicality measures were collected because the equipment used during acquisition and extinction was removed. Hence, it is possible that participants viewed the collection of my prototypicality measures as occurring in a new context, allowing the conditioned response to be renewed in the new context. In addition, spontaneous recovery of the conditioned response could have occurred because there was a period of time between extinction and the collection of prototypicality measures, which could represent a change in the temporal context (Bouton, 2002; Rescorla, 2004). Therefore, I cannot determine whether changes in prototypicality are independent or dependent on changes in affective responses.

Another unexpected result detected in this thesis was the persistence of heightened self-reported (vs. physiological) anxiety after extinction in Studies 4.1, 5.1

and 5.2. In these studies self-reported anxiety persisted beyond extinction, raising the possibility that participants left the experiment with conditioned perceptions. Evidence suggests that self-reported valence requires more trials to extinguish compared to electrodermal responses, which could explain the disparity between the two measure (Lipp, Oughton & LeLievre, 2003). Renewal/ and spontaneous recovery (previously discussed) offer another an account as to why the self-reported anxiety measure result at post-extinction was different to SCR data, which provided evidence for successful extinction. The conditioned response could have renewed or spontaneously recovered after physiological extinction, which could explain why relatively higher self-reported anxiety ratings were detected.

There is also a range of evidence to suggest participants left the experiment with no heightened anxiety. Firstly, although post-extinction self-reported responses were slightly elevated in these studies in comparison to pre-test responses, the absolute level of these responses was still below the midpoint of the scale. Thus, self-reported anxiety was low. Moreover, SCR results detected no increases in responses during extinction, which suggests anxiety was extinguished (i.e., there are inconsistencies between measures). As mentioned above, SCR require less extinction trials compared to self-reported anxiety (Lipp et al., 2003). Furthermore, conditioned anxiety was established in a controlled setting and a similar context is important to predict the presentation of a conditioned response (Hall & Honey, 1990; Milad et al., 2005; Vansteenwegen et al., 2005). As a result it is unlikely that effects detected in these studies will persist beyond extinction outside the laboratory. Finally, participants underwent a positive visualisation task for the ethnic group of interest in the study (e.g., a speech from Barack Obama welcoming a new era of hope and change) at the conclusion of the study to extinguish any lingering anxiety. The positive visualisation task and subsequent debriefing is likely to act similar to instructed extinction, furthering bringing down anxiety (Mallan et al., 2009). Therefore, it is highly unlikely that participants left the experiment with conditioned perceptions that could impact their daily lives.

Moreover, another limitation is that my research cannot distinguish whether evaluative-fit or emotion-fit mechanisms are responsible for exemplar prototypicality shifts. I attempted to address this limitation in Study 6.1 by pairing outgroup exemplars with a stimulus that was not associated with any anxiety but was negatively valenced (i.e., results could have ruled out emotion-fit mechanisms of prototypicality shifts). However, when I adjusted the procedure used in Studies 3.1 to 5.1 and implemented a different type of reinforcer in Study 6.1, methodological issues may have occurred and prevented the key research question from being adequately addressed. In particular, cognitive demand was high throughout conditioning compared to previous studies and conditioning may not have occurred as it did in previous studies. A simplified research design that reduces the number of tasks required of the participant should allow the association between exemplar and positivity to develop more readily, which could be addressed in future research.

It is desirable to isolate the exact mechanisms responsible for prototypicality shifts so that effective solutions to address problematic intergroup relations can be implemented. Different theoretical frameworks suggest different intervention strategies to improve intergroup relations in society. For example, evaluative-fit explanations focus on promoting a positive experience in intergroup relations. In contrast, emotion-fit explanations focus on eliminating any fear/anxiety that could be associated with another group. Understanding whether evaluative-fit or emotion-fit facilitates shifts of exemplar prototypicality would allow appropriate policies and procedures to be implemented that address problematic intergroup relations. My data does not provide a sufficiently strong basis to recommend one approach over the other.

The changes in perceived prototypicality following an association with a negative/anxiety provoking stimulus reported in this thesis are consistent with evaluative-fit and emotion-fit psychological explanations of ingroup-outgroup bias and stereotyping (Cottrell & Neuberg, 2005; Dasgupta, DeSteno, Williams & Hunsinger, 2009; DeSteno, Dasgupta, Bartlett & Cajdric, 2004; Miller, Maner & Becker, 2010; Oakes, Haslam & Turner, 1994; Oakes, Turner & Haslam, 1991; Rubin & Hewstone, 1998; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). Central to both theoretical traditions (Ellemers, Spears & Doosje, 2002; Scheepers, Spears, Doosje & Manstead, 2006) is a bias to favour the ingroup over the outgroup because of either a motivation to achieve or maintain differential group status (valence) or because of differential perceived safety/threat (emotions). My results suggest that evaluative-fit/emotion-fit mechanisms not only change affective responses to exemplars, but might be implicated in how an exemplar is categorised, which has consequences for the maintenance and reinforcement of negative group schemas and stereotypes in society.

The studies in Chapter 3 and Study 4.1 (among contingent aware participants) showed that an outgroup exemplar who became associated with negativity/anxiety shifted in perceived prototypicality and was perceived as more outgroup-like. When the contingent association between negativity/anxiety and the exemplar did not occur and the association rather formed with the Black group, contingent non-specific changes in exemplar prototypicality occurred and both exemplars were perceived as being more prototypical of the outgroup. Despite non contingent-specific changes occurring in some studies, the direction of prototypicality shifts align with evaluative/emotion-fit mechanisms. Thus, in line with evaluative/emotion-fit mechanisms, negatively valenced and anxiety provoking associations resulted in an increase in the perceived prototypicality of outgroup exemplars (Hogg, 2000; Oakes et al., 1994; Rubin & Hewstone, 1998).

Correspondingly, Study 5.1 found an ingroup exemplar associated with non-negativity/safety was perceived as being more prototypical of the ingroup. This shift of exemplar prototypicality aligns with evaluative/emotion-fit mechanisms also because the ingroup is typically associated with positivity and safety. Furthermore, Studies 5.1 and 5.2 showed a tendency for an ingroup exemplar associated with negativity/anxiety to be less prototypical of the ingroup. Whilst shifts of exemplar prototypicality in this direction were not statistically reliable, these trends suggest that there was a tendency for the unsafe ingroup exemplar to be excluded cognitively from the ingroup through a decrease in how representative they were perceived to be of the ingroup. Together these results suggest that an exemplar associated with positivity/safety increased in their inclusiveness within the ingroup representation and decreased in their inclusiveness within the outgroup representation. In contrast, an exemplar associated with negativity/anxiety increased in their inclusiveness within the outgroup representation and decreases their inclusiveness within the ingroup representation.

An extension of this reasoning is that positive and safety associations should lead to improved exemplar perceptions through decreased inclusiveness within the outgroup representation and increased inclusiveness within the ingroup representation. This is the prediction I attempted to test in Study 6.1 by developing an appetitive conditioning paradigm where participants won money against outgroup exemplars. Unfortunately, this novel gambling game paradigm appeared to have quite different psychological effects on participants than the aversive classical conditioning paradigm I had used in the earlier studies. Hence, no firm conclusions could be made from this study but the theoretical assumptions I based my predictions on are still valid; they are consistent with previous research and are worth further investigation.

Notwithstanding the complexities and limitations presented previously, this series of studies provides preliminary evidence that forming negative associations with individual group exemplars changes exemplar cognitive representation through shifts in exemplar perceived prototypicality. This pattern of results is consistent with, and extends past categorisation research that has focused on evaluatively/emotionally loaded ingroup and outgroup exemplars. Extant research indicates that positively valenced exemplars are more likely categorised as ingroup members, whilst negatively valenced exemplars are more likely categorised as outgroup members (Miller et al., 2010; Richeson & Trawalter, 2005; Ruys, Dijksterhuis & Corneille, 2008). Furthermore, positive facial expressions (i.e., happy faces) are more likely categorised as ingroup members and negative facial expressions (i.e., sad faces) are more likely categorised as outgroup exemplars (Hugenberg 2005; Hugenberg & Bodenhausen, 2003). Thus, evaluations of exemplars and applicable emotions guide exemplar categorisations into ingroup vs outgroup.

Previous research is extended in this thesis through incidental pairings of exemplars with valenced/emotionally loaded stimuli through conditioning procedures. Past social categorisation research manipulated the valence of exemplars with integral qualities (e.g., personality). Integral valence is difficult to change because valence is intrinsically associated with the exemplar (Bodenhausen, 1993). By investigating incidental pairings of valence, I found that extrinsic associations with environmental stimuli that have no inherent association with an exemplar are capable of shifting how an exemplar is perceived and categorised. My research therefore suggests that negative/anxiety associations with group exemplars serve not only as important markers of group membership (as already demonstrated in past research, e.g. Olsson et al., 2005; Weisbuch et al., 2009), but actively *shape* and *change* the inclusion-exclusion status of group exemplars. Hence, by associating exemplars with incidental valence I investigated changes in perceptions of several different exemplars (i.e., CS+/unsafe exemplar vs CS-/safe exemplar) over time and ascertained how perceptions shifted following a controlled extrinsic association with a negative/anxiety provoking stimulus.

My results and the implication they have align with the ideas proposed in various models (MacInnis & Page-Gould, 2015; Paolini, Harris & Griffin, 2016). These researchers argue that divergent conclusions made by the intergroup interaction literature and the intergroup contact literature are aspects of the same process. *Intergroup interaction* research typically involves short interactions between ingroup and outgroup exemplars, which are unstructured and produce negative outcomes and anxiety (MacInnis & Page-Gould, 2015). Intergroup interaction possibly maps onto the acquisition process in my research because participants have brief contact with outgroup exemplars and negative evaluations and anxiety are learnt (i.e., evaluative-fit and emotion-fit mechanisms). *Intergroup contact* research typically involves longer-term contact with outgroup exemplars, which are structured and produce positive outcomes and a reduction in anxiety (MacInnis & Page-Gould, 2015). Intergroup contact is similar to the extinction procedure in my research because participants had prolonged exposure to outgroup exemplars and negative evaluations/anxiety are at this point reduced. MacInnis and Page-Gould (2015) argue that intergroup interaction and intergroup contact are part of the same larger process and intergroup interaction processes just precede intergroup contact processes. The model proposed by Paolini et al. (2016) incorporates a similar temporal sequence to MacInnis and Page-Gould (2015) but the authors inject contingency and valence into their model. Contingency and valence, along with changes across time, are important components for my research because they allowed me to investigate changes in exemplar prototypicality for different exemplars. Thus my research demonstrates that following acquisition (aka intergroup interaction) and extinction (aka intergroup contact), exemplar prototypicality is increased for specific exemplars (cf. non-associative processes prototypicality shifts). Together, both models suggest an avenue that outlines how exemplar prototypicality changes could lead to broader impacts in society, which are discussed later.

In summary, the evidence presented in this thesis suggests the robustness of the prototypicality shift effect unstable because there were different types of shifts between studies (associative vs non-associative shifts) and disparities between explicit and implicit measures of prototypicality. Whilst I provide evidence that exemplar prototypicality shifts are in line with evaluative-fit and emotion-fit mechanisms, I am unable to ascertain which mechanism is responsible. Despite these complexities and limitations, this is the first evidence to suggest that exemplar prototypicality can be shifted following an association with a negative/anxiety provoking stimulus. As such, these results suggest the inclusion/exclusion status within a group is actively shaped and changed by negative/anxiety provoking associations, which could have broader consequences for stereotypes and intergroup relations.

Mechanisms Underlying Prototypicality Shifts

The main focus of this research was to test whether prototypicality shifts as a function of conditioning. A secondary aim was to initiate an exploratory investigation into the underlying factors responsible for changes in exemplar prototypicality. However, given the novelty of the effect, I treated efforts in this direction as theoretically-informed but exploratory in nature. This led to a number of complexities and limitations of my results in regards to conclusions that can be made about underlying mechanisms. I begin this section by discussing these complexities and limitation in greater detail, whilst providing direction for future research. Despite these complexities and limitations, I conclude this section with a discussion of three underlying mechanisms that results from this research suggest could be implicated in prototypicality shifts: (1) Repeated presentations and passage of time in the extinction process; (2) Familiarity coupled with attention during the extinction process; (3) Contingency awareness.

Interpretation of the results of this research is limited by stretched sample size and insufficient power throughout, which might have led to false-negative results (Button et al., 2013; Cohen, 1992) and unreliable positive effects (Christley, 2010; Simmons, Nelson & Simonsohn, 2011; Verhoeven, Simonsen & McIntyre, 2005). The designs used in my thesis were appropriately powered for a focus on key between-subjects effects (i.e., Group membership manipulation, time point post-test prototypicality measures were collected etc.), rather than the exploration of additional moderators (i.e., contingency awareness). Yet, throughout my studies I still attempted to explore possible mechanisms involved in prototypicality shifts by assessing mediating variables and checking (a-posteriori) for the role of additional moderating factors. I treated and regarded these ancillary analyses as merely exploratory because of these power issues. In hindsight, I should have designed my studies taking into consideration these additional factors at the designing stage. Testing sessions were long due to the complex research paradigm, high in resource allocation (i.e., psychophysiological measures restricted testing to one participant at a time). As a result, I had appropriate power to test for basic prototypicality shifts, but I did not have enough participants in my studies to test the underpinning psychological mechanisms potentially involved in switching exemplar prototypicality shifts on and off. Therefore, whilst I provide a first look at the effects these underpinning factors may have, results should be interpreted with caution and future studies that investigate underlying mechanisms should use appropriately powered designs.

In my research I interpreted changes in exemplar prototypicality as reflective of the exemplar shifting in prototypicality towards or away from the group prototype. However, it is equally possible that my findings reflect the group prototype shifting towards or away from the exemplar. I leaned towards the former rather than the latter interpretation because of processes that underlie categorisation models (prototype, exemplar, mixed; Hilton & Von Hippel, 1996; Kruschke, 2011; Minda & Smith, 2002; Minda & Smith, 2011; Rosch, 1978; Smith & Zarate, 1990). These models suggest category exemplars become more representative of their group based on their similarity to a group prototype or group of exemplars. In these models central tendency (or the group prototype) is a slow changing parameter. Therefore the group prototype is not expected to shift from limited reinforced associations. Research shows central tendency required prolonged exposure to many exemplars (Hamberger, 1994; Hewstone, 1994; McIntyre, Paolini & Hewstone, 2016; Paolini, Hewstone, Cairns & Voci, 2004). However, it is still logically possible that my research reflects the group prototype (also or exclusively) shifting towards the exemplar.

Future research ideas to distinguish whether the exemplar and/or central tendency shifts include the incorporation of additional measures. Current measures assessed an exemplar's perceived prototypicality but did not ask about the general ingroup/outgroup stereotype. To distinguish if the exemplar and/or central tendency of the group shifted, questions similar to those used by Hewstone and Hamberger (2000) could be incorporated into the research design. Hewstone and Hamberger asked participants to read descriptions of exemplars that contained both stereotypical consistent and inconsistent information. Stereotype descriptions provided details about a group and group exemplars such as how they looked and behaved. Participants rated how typical each exemplar was to the stereotype description. Additionally, participants rated how similar the general perception of the group was to the stereotype description. Thus, a stereotypical rating was collected for individual exemplars and the group as a whole. Including a measure similar to Hewstone and Hamberger's into my experiment would allow an assessment of group and exemplar perceptions to be compared pre- to post-test. If individual exemplar features shifted and the group prototype representation remained stable, I could have unequivocally determined that exemplar prototypicality shifted towards the outgroup. However, if the exemplar representation remained stable and the group prototype perception changed, then the prototype would have shifted towards the exemplar.

Notwithstanding these limitations, I discuss possible mechanisms that could facilitate prototypicality shifts. One mechanism that I explored in this research was extinction processes and the influence it had on prototypicality shifts. During extinction two exemplars observed during acquisition were presented repeatedly in the absence of an aversive stimulus. Repeated presentations in the absence of an aversive stimulus create a new association with the exemplar, thereby making the association developed during acquisition ambiguous (Bouton, 1994; 2002; 2004; 2014). I measured exemplar prototypicality shifts before and after extinction and found prototypicality shifts only after extinction in most cases (cf. Study 5.1, see Table 3 perceived prototypicality shifts). I concluded that extinction processes must be involved in shifting perceived exemplar prototypicality. Another mechanism that could facilitate prototypicality shifts is contingency awareness. Evidence suggests contingency awareness is a necessary process in conditioning (Hofmann, De Houwer, Perugini, Baeyens & Crombez, 2010; Pleyers, Corneille, Luminet & Yzerbyt, 2007), but the effect it has for prototypicality shifts has not been explored. Hence, various mechanisms involved in extinction and contingency awareness are discussed.

Extinction Process: Repeated Presentations and Passage of Time

One mechanism involved during extinction that I suggest might assist shifts of exemplar prototypicality are repeated presentations of ethnic cues. Study 3.1-3.3 repeatedly presented Black exemplars, whose ethnic cue was accessible (i.e., dark skin). In all 3 studies in Chapter 3, Black skin was repeatedly presented to participants during extinction. Changes in perceived exemplar prototypicality was found only after extinction (and not beforehand), suggesting repeated presentations of exemplars without an aversive stimulus were at least partially responsible for prototypicality shifts. Furthermore, the number of exemplar presentations required to extinguish conditioned anxiety varied in Study 3.2 and this was investigated as a possible mediator. Variability in the number of exemplars that were presented during extinction statistically mediated shifts in perceived outgroup prototypicality, suggesting repeated exemplar presentations were at least partially responsible for prototypicality shifts.

Study 4.1 provided further evidence that repeated presentations of ethnic cues in the absence of an aversive stimulus could be responsible for prototypicality shifts. During extinction one group of participants viewed repeated exemplar presentations, whilst the other group viewed distorted images of exemplars that retained ethnic cues but could not be individually identified. Thus, with this manipulation I systematically contrasted repeated exemplar presentations against repeated presentations of ethnic cues, which are otherwise embedded in the basic extinction process. Study 4.1 found prototypicality shifts in both groups, which suggests repeated presentations of ethnic cues, as opposed to repeated exemplar presentations, were sufficient to cause exemplar prototypicality shifts.

Chapter 5 results provide mixed evidence for repeated visual/ethnic cue presentations as driving prototypicality shifts. Study 5.1 found prototypicality shifts in safe exemplars away from the outgroup prototype and towards the ingroup prototype. Prototypicality shifts were found following extinction as in previous studies, but this time shifts were also found before extinction. The inclusion of ingroup exemplars may have resulted in a smaller number of visual cue presentations being required compared to other studies (i.e., due to meta-contrast principle; Corneille & Judd, 1999; Oakes et al., 1991), but further research is needed to determine this. Furthermore, Study 5.1 group membership was defined by minimal groups and the amount of information about groups available to participants was limited. Navarrete et al. (2012) argue that a dissociation between learning effects and resistance to extinction in their results for social vs. minimal groups was due to differences in pre-existing cognitions about the groups across studies; something similar might have occurred in my research. Hence, it is possible that the investigation using minimal groups altered the need for repeated presentations in prototypicality shifts.

One might speculate that repeated ethnic cue presentations change exemplar prototypicality independent from whether conditioning preceded extinction. For example, Macrae and Bodenhausen (2000) suggest repeated exposure to stimulus events gradually changes categorisation. It is possible that in my research simply exposing participants to category cues changed prototypicality as categorical thinking becomes more accessible. Although this possibility was not directly tested, this type of effects seems improbable because changes in prototypicality most often occurred towards certain exemplars (e.g., unsafe faces) and the direction of the shifts differed in instances.

Study 3.1 – 5.1 found shifts towards the exemplar paired with an aversive stimulus. If repeated ethnic cue presentations alone were responsible for this effect prototypicality shifts should have been found for both exemplars (cf. non-associative prototypicality shifts in Study 3.3 and 4.1; Study 4.1 contingent aware results do align with this argument). Furthermore, prototypicality shift direction varied in Study 5.1 – A trend for unsafe faces to shift towards the outgroup and safe faces shifted towards the ingroup. I would expect exemplar prototypicality shifts to be in a consistent direction if repeated ethnic cues alone exacerbated categorical thinking. Altogether, results suggest repeated ethnic cues following acquisition in the absence of an aversive stimulus are at least partially responsible in shifting exemplar prototypicality, but do not account for the whole story.

Other factors could contribute to explain my results; however the data collected in this investigation cannot establish this with confidence. The inability to pinpoint the exact mechanism is made more difficult because in my basic paradigm repeated presentation of membership cues in my studies was confounded with a simple passage of time. For example, a passage of time might help to consolidate memory traces that could be responsible for prototypicality shifts and changes in cognitions (Feld & Diekelmann, 2015; Landman et al., 2016; Rauchs et al., 2011). Whilst I am not in a position to isolate the exact mechanism underlying my results, from the data collected repeated presentation of ethnic cues remains one of the most plausible explanations.

Extinction process: Familiarity Coupled with Attention

Another possible mechanism underlying prototypicality shift effects following extinction is familiarity coupled with attention effects. Previously I suggested repeated presentation of ethnic cues during extinction might help consolidate participants' perceptions of exemplars' group representativeness. Here I take this explanation a step further and suggest repeated presentations of ethnic cues might increase the (stimulus-specific) link between the aversive stimulus, the unsafe exemplar and group membership.

Increases in exemplar familiarity have been demonstrated to increase stereotypical processing (Garcia-Marques & Mackie, 2007; Smith, Miller, Maitner, Crump, Garcia-Marques & Mackie, 2006). Familiarity alone however cannot explain when prototypicality shifts occur in one exemplar but not another (Studies 3.1, 3.2, 5.1), as in my paradigms safe and unsafe faces were presented the same number of times. To provide a complete explanation, attentional effects in conjunction with familiarity might need to be considered. Visual attention towards stimuli is demonstrated to increase categorical representation (Macrae, Bodenhausen, Milne & Calvini, 1999). Following acquisition, attention is directed towards the unsafe face and the association with negativity/threat, causing contingent specific prototypicality shifts (Armony & Dolan, 2002). Non-associative prototypicality shifts could be a result of a lack of familiarity and attention being directed to the unsafe face. Hence, familiarity and attention may be possible underlying mechanisms facilitating prototypicality shifts, but future research is needed to explore this possibility further.

Contingency Awareness

Contingency awareness is another possible mechanism involved in exemplar prototypicality shifts. Contingency awareness refers to when an individual is aware of the association between a stimulus and reinforcer (Kattner, Ellermeier & Tavakoli, 2012). Within the learning literature the need for contingency awareness in conditioning has been debated (Baeyens, Eelen, Van den Bergh & Crombez, 1990; Lovibond & Shanks, 2002; Walther & Nagengast, 2006). The debate has been largely resolved and evidence suggests contingency awareness is required for fear and evaluative conditioning (see Hofmann et al., 2010; Pleyers et al., 2007). However, the role contingency awareness has for categorisation was yet to be tested.

Throughout this research fear/evaluative conditioning was important in facilitating prototypicality shifts. Without negative/anxiety provoking associations with stimuli, evaluative-/emotion-fit could not occur. My research attempted to go a step further and sought to explore the mechanisms involved in conditioning that facilitated prototypicality shifts. Contingency awareness was one factor of conditioning that I explored and could play an important role in exemplar prototypicality shifts. I found initial evidence for contingent specific prototypicality shifts when participants were aware of the contingency relationship between the exemplar and unconditioned stimulus but not when participants were contingent unaware. Exploratory analysis in Study 3.2, 4.1 and 5.1 found stronger shifts of prototypicality when participants were contingent aware compared to contingent unaware participants. Furthermore, Study 3.3 masking procedure removed the ability for participants to be contingent aware and contingent specific prototypicality shifts were not found. Therefore, my results provide preliminary evidence that contingency awareness is another mechanism involved in shifting exemplar prototypicality.

If prototypicality shifts depend on contingency awareness, other factors associated with contingency awareness could influence prototypicality shifts. Research suggests contingency awareness shapes conditioning and conditioning effects (i.e., anxiety) are stronger amongst contingent aware participants (Labrenz, Icenhour, Benson & Elsenbruch, 2015; Van Dessel, De Houwer & Gast, 2016). As a result of stronger conditioned effects, a greater number of extinction trials should be required to extinguish the effect. In addition to higher levels of anxiety and greater number of extinction trials, research also suggests memory and awareness are key factors in facilitating contingency awareness (Gawronski & Walther, 2012; Lovibond & Shanks, 2002). Therefore, it is possible that factors such as greater anxiety, number of extinction trials, memory and awareness were involved in prototypicality shifts, but were concealed by the overarching contingency awareness/conditioning theme.

In summary, limitations of my exploratory analysis into underlying mechanisms involved in prototypicality shifts include power concerns for moderational analysis and not being able to conclude whether the exemplar or prototype (central tendency) shifts. Despite these limitations, I provide initial evidence that extinction processes and contingency awareness might be underlying mechanisms that facilitate prototypicality shifts. Future research should follow up the effect these mechanisms have with well powered designs in order to provide more conclusive evidence for the effect they have in shifting exemplar prototypicality.

Generalisability of Prototypicality Shifts

In this section I expand on my previous discussion around prototypicality shifts and discuss how this effect could generalise to similar and new exemplars, which has broader consequences for intergroup relations and society in general. Firstly, I provide an overview of generalisation results. Next I discuss complexities and limitations of my research around generalisability of effects to other exemplars and more broadly into society. I conclude this section by discussing the implications results have for intergroup relations and stereotypes, notwithstanding the limitations already discussed.

In my research two types of generalisation stimuli were used. A 25% and 50% variation of each target face that moved progressively away from group prototype along a generalisation gradient (towards a computational average of all FaceGen White, Black, Middle-eastern and Asian faces) were used to measure generalisation to face variations. Two new exemplars that were unrelated to the target faces were also used to determine if prototypicality shifts extended to new exemplars. Contingent specific changes to face variations in the expected direction were detected in Study 3.1 and 5.1, whilst non-associative changes in the expected direction were detected in Study 3.3, 4.1 and 6.1. Evidence for changes in prototypicality were also found in new exemplar data and prototypicality shifts were detected for at least one new exemplar in Study 3.1, 3.2, 3.3 and 4.1. Prototypicality shifts for new exemplars were also detected in Study 5.2 and 6.1, but results cannot be interpreted because the necessary between subjects interaction was not significant. For example the same new exemplar prototypicality shift was detected in the ingroup vs outgroup condition and win vs lose conditioning treatment, which prevented interpretable conclusions being made. A summary of all the generalisation results is found in Table 3 under the generalisation heading.

My generalisation results provide initial evidence that prototypicality shifts extend from one exemplar who becomes associated with negativity/anxiety to another exemplar. Generalisation effects map onto traditional distinct operationalisations of generalisation in learning and social psychology (Pettigrew & Tropp, 2006; Ranganath & Nosek, 2008; Verosky & Todorov, 2013), and have larger consequences for intergroup relations and stereotyping. Rothbart and John (1985) argue that typical group exemplars are more influential at changing group perceptions compared to atypical group exemplars (Rothbart & Lewis, 1988; Wilder, Simon, & Faith, 1996). Thus, if negative/anxiety associations generalise to other exemplars a likely consequence is negatively perceived stereotypes and intergroup relations will be maintained and further worsened. However, prototypicality shifts detected varied across measures (implicit vs explicit) and were not consistent across the generalisation and new exemplar stimuli.

Therefore, there are limitations associated with results and caution should be taken when interpreting implications from these results.

Another limitation regarding generalisation of my effects is that this research investigated only one source of perceived prototypicality across visually accessible stimuli: I investigated perceived prototypicality with reference to ethnicity and skin colour because previous research demonstrated ethnicity is one of the quickest activated categories that is socially very consequential across the globe (Ito & Urland, 2003; Zhao & Bentin, 2008). Apart from only focusing on one type of ethnicity, my research only focused on visual prototypicality. Future research could address this limitation and investigate other forms of prototypicality to determine if changes in perceived exemplar prototypicality are limited to visual features or extend to other features. Voices for example are an important tool in communication and could be used to categorise individuals (Babel & McGuire, 2015; Miller et al., 2010). Miller et al. (2010) for example showed that voices that are more male/threatening cause the exemplar to be categorised as more typical of the outgroup than ingroup. Determining whether perceived voice prototypicality shifts after aversive or appetitive associations develop would provide evidence that shifts in exemplar prototypicality can be extended outside of visual perception. Identifying other sources of prototypicality shifts will have broader real-world consequences as globalisation results in everyday contact between different groups via the phone, which could help form the basis of intergroup perceptions and stereotypes.

In addition to the source of prototypicality investigated, future research could investigate how different types of contact influence prototypicality. Results from Chapter 3 provide initial evidence that changes in exemplar prototypicality did not require associations to be learnt first-hand and occurred equally when an association was learnt through watching another ingroup member (i.e., a White model) experience the aversive associations. Results are consistent with research that suggests different forms of exposure (that is not limited to direct exposure) change intergroup perceptions (Bandura, 1977; Mallan et al., 2009; Mazziotta, Mummendey, & Wright, 2011; Miller & Dollard, 1941; Olsson & Phelps, 2004). Hence, my results suggest that other forms of socially mediated experiences, in particular vicarious exposure, can change how an exemplar from a stereotyped group is cognitively perceived (Hewstone, 2010; Paolini et al., 2004; Wright, Aron, McLaughlin-Volpe & Ropp, 1997).

Changes in perceived prototypicality through different forms of exposure is relevant in today's society because increasing technological advancements have the power to spread messages about different exemplars and groups instantaneously to a broad audience. Other forms of social exposure (vicarious, media, imagined) could explain how segregated societies or people not willing to engage in intergroup contact form and change perceptions towards a group and group exemplars (Greenberg, Mastero & Brand, 2002; Tan, Zhang, Zhang & Dalisay, 2009). Hence, by showing exemplar perceived prototypicality changes through vicarious exposure, I demonstrated a unique way in which exemplar prototypicality may change, and as a result, how stereotypes develop and are maintained. Future research may focus on the impact that different forms of exposure have in shaping categorisation and how these effects generalise to the broader society when direct contact does not occur (for example see Harwood, Paolini, Joyce, Rubin & Arroyo, 2011; Paolini et al., 2014).

Another limitation relates to the faces used in this research. Facial stimuli presented to participants throughout this research were computer generated, young, male faces (with the exception of Study 5.1). The decision to present male exemplars only was based on Navarrete, Olsson, Ho, Mendes, Thomsen and Sidanius' (2009) research, which found greater levels of anxiety developed towards male faces (vs females) during fear conditioning. Thus, male faces were chosen because they were expected to produce more affective learning, potentially resulting in greater prototypicality shifts. The decision to use male targets was consistent with evolutionary-based theories because more polarised evaluations associated with male exemplars are expected to generate more fear in participants (Cottrell & Neuberg, 2005; Navarrete et al., 2009). Future research could investigate whether prototypicality shifts develop towards female exemplars to provide evidence that shifts in perceived prototypicality following conditioning are not limited to males only.

Future research could also present real exemplars (vs. computer-generated faces), which would improve the ecological/external validity of this research. Computer generated faces (except for Study 5.1) were used in this research so that facial qualities could be controlled. This allowed for the selection of typicality, anxiety and attractiveness based on the Facegen database. Additionally, this approach allowed for generalisation faces to be created that varied in their typicality from target faces in a controlled manner. Thus, in order to improve the ecological/external validity of prototypicality shifts, real faces should be used.

Future research could use real world stimuli from other groups . White and Black exemplars were only used in my research (except for minimal groups in Study 5.1). Stereotypes and prejudice towards Asian and Muslim exemplars is well documented in the literature (e.g., French, Franz, Phelan & Blaine, 2013; Schug, Alt & Klauer, 2015) and a possible avenue for this methodology to extend into. Therefore results could be extended to investigate perceived exemplar prototypicality of Asian or Middle Eastern ethnic groups, different age groups, sporting groups etc. Extending the generalisability of the prototypicality shift effect to other groups would indicate that cognitive representations of all group exemplars can shift after associative learning. Finding these generalised effects would allow new methods to be developed to address a larger range of problematic intergroup relations.

Extending research into different groups is beneficial because certain groups are associated with different emotions. For example, Dasgupta et al. (2009), DeSteno, et al. (2004) and Cottrell and Neuberg (2005) provide evidence that different groups are associated with specific emotions. Black people are associated more with fear and aggression, whilst homosexuals are associated with disgust. If exemplars need to be paired with group specific emotions in order to form associations and change exemplar perceived prototypicality, then the aversive stimulus used in my research might only change perceptions towards Black exemplars and other groups associated with fear (i.e., Muslims). Consistent with this group-specificity, Golkar, Bjornstjerna and Olsson (2015) found fear persisted for Black outgroup exemplars but not Muslim outgroup exemplars after aversive conditioning, suggesting effects could be different for each outgroup. Thus, investigating prototypicality shifts for different groups would allow a more stringent test of emotion-fit hypothesis, rather than sticking to one type of outgroup and one type of association. The theoretical and practical implication of this would be targeted strategies are needed to address problematic intergroup relations. Together the issues raised and proposed ideas for future research will allow this research to inform implementable strategies for real world settings.

Taking these limitations and ideas for future research into consideration, my results provide preliminary evidence that prototypicality shifts might extend to exemplars not involved in conditioning. These generalisation effects are consistent with previous research that investigated evaluations generalising from one exemplar to another (Lissek et al., 2010; Pettigrew & Tropp, 2006; Ranganath & Nosek, 2008;

Verosky & Todorov, 2013). My research provides preliminary evidence that extends this research because changes in perceived prototypicality (or categorisation) generalised to similar exemplars, instead of evaluations as demonstrated by previous research.

My generalisation results could help to explain negative stereotyping and intergroup relations in society because changes in exemplar-level prototypicality under aversive conditioning changed perceived prototypicality of other exemplars. As a result, the prototypical representation of the group could ultimately worsen and negative stereotypes become more common via a process called member to group generalisation (Paolini et al., 2004). Member to group generalisation is the social cognitive process whereby individual exemplar experiences generalise and change the perception of the group as a whole (Brown & Hewstone, 2005; Rothbart & John, 1985; Rothbart, Sriram, & Davis-Stitt, 1996; Sherman, 1996). My results suggest changes in exemplar-level prototypicality following an aversive association cause heightened outgroup prototypicality in similar exemplars never associated with an aversive experience. Heightened outgroup prototypicality that extends from one exemplar as a result of aversive conditioning to another exemplar not involved in conditioning reinforces and maintains the perceived aversive fit with the outgroup. Hence, problematic changes in exemplar cognitive representations, which could lead to problematic intergroup categorisation and stereotyping, might extend from an interaction with one exemplar to other group exemplars. By extending to exemplars not involved in associative learning my results provide initial evidence for a unique way in which negative intergroup stereotypes could develop and are maintained.

Future research should include a measure to determine if changes in exemplar prototypicality extend to and influence changes in stereotyping (Blair, Judd, Sadler &

Jenkins, 2002; Hewstone & Hamberger, 2000). Blair et al. (2002) and Hewstone and Hamberger (2000) presented participants with short descriptions of information that contained stereotypically consistent and inconsistent information. Each exemplar was presented and rated on their fit with the short descriptions. A similar measure could be incorporated into research that further investigates the effects presented in this thesis. If this type of measure was included in my research, I would expect exemplars who were perceived as being more prototypical of their group to be rated as a better fit with descriptions that contained more stereotypical consistent information after (vs before) associative learning. In contrast, exemplars that were perceived as being less prototypical of the group after associative learning were expected to fit less with stereotypical descriptions. By including a stereotyping measure in future research the link between changes in exemplar perceived prototypicality and stereotyping can be empirically demonstrated within a single research design and paradigm. This would strengthen my research by showing changes in exemplar prototypicality influence stereotypes more broadly, which has broader implications for intergroup relations in society.

The assumption that changes in exemplar prototypicality are linked to stereotyping is supported by previous research (Blair, Judd & Chapleau, 2004; Blair et al., 2002; Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000; Rothbart & John, 1985; Rothbart et al., 1996). In particular, research in the area of stereotype change suggests intergroup attitudes and perceptions are changed more readily when one is exposed to group exemplars that are perceived to be representative of that group, or in other words, more prototypical of their group (for a review of evidence, Brown, Eller, Leeds & Stace, 2007; Brown & Hewstone, 2005; McIntyre et al., 2016; Rothbart & John, 1985; Rothbart et al., 1996). Thus, if an aversive experience persists after prototypicality has been heightened an already negative stereotype will be worsened. However, positive effects can be exhibited from heightened prototypicality. If prototypicality has been heightened, knowledge that a typical exemplar is a smart, industrious and a well-liked person could impact the group perception in a positive manner and improve negative stereotypes (Birtel & Crisp, 2012; Kunda, Davies, Adams & Spencer, 2002). Thus, an experience that increases exemplar prototypicality is the first step to changing stereotypes, but the experience following an increase in prototypicality should be key to the direction of the stereotype change.

Previously I described how exemplar-level changes in prototypicality could influence and worsen group-level stereotypes. At the onset of an exemplar-level experience, categorisation occurs and stereotypes are activated to facilitate intergroup interactions (Hamilton & Sherman, 1994). Thus, group-level stereotypes shape responses at the exemplar-level via group-to-exemplar generalisation (Chen & Ratliff, 2015; Gawronski & Quinn, 2013). A circular relationship is likely to develop, because changes at the exemplar-level influence group level stereotypes and vice versa (Gawronski & Quinn, 2013). Hence problematic intergroup stereotypes at the group-level (developed from exemplar-level experiences) are activated when interacting with new group exemplars, thus potentially fuelling long-term and far-reach effects of stereotypes and prototypicality shifts.

In summary, my research provides preliminary evidence that exemplar prototypicality can be shifted following acquisition and subsequent extinction of negativity/anxiety. Limitations around the generalisability of my results to other exemplars and stereotypes more broadly include disparities in the data and only one type of prototypicality being investigated with a limited number of faces. Increases in prototypicality are linked to greater changes in stereotypes because typical exemplars are more representative of a group and channel larger and more stable exemplar-to-group generalisations. Thus, any experience following heightened exemplar prototypicality should be more likely to be incorporated into the group stereotype.

Coda

Altogether, the research presented in this thesis investigated changes in exemplar prototypicality using a unique approach that has not been undertaken before. In particular, changes in a social psychological construct (prototypicality) are investigated with conditioning techniques. Previous research used a similar approach to investigate changes in exemplar evaluations (Olsson et al., 2005; Olsson & Phelps, 2004), but I am not aware of any research that uses this approach to investigate changes in perceived prototypicality. Changes in perceived prototypicality are an important determinant of category activation and highly influential in social categorisation and stereotyping (Blair et al., 2004; Bodenhausen & Peery, 2009; Macrae & Bodenhausen, 2000). By using conditioning techniques, I was in a novel position to determine changes in exemplar categorisation across time for specific exemplars (contingent effects) associated with different types of valence. Investigations that incorporate contingency, time and valence are important because the psychological consequences an individual experience with a limited number of group exemplars has for new exemplars and group level responses will ultimately shed light on stereotypes and intergroup perceptions (Paolini et al., 2016). Therefore, if the robustness of the effect detected in this research is confirmed, it will have consequences for stereotyping, intergroup relations and attitudes that could lead to strategies to improve these in society more broadly.

In conclusion, this thesis reported seven experiments that investigated changes in exemplar prototypicality before and after an association developed between group exemplars and a valenced/anxiety provoking stimulus. Preliminary evidence suggested that exemplar prototypicality shifts were consistent with evaluative-fit (Hogg, 2000; Tajfel & Turner, 1979) and emotion-fit mechanisms (Cottrell & Neuberg, 2005; Dasgupta et al., 2009). Exemplars associated with an aversive experience were perceived as being more prototypical of the outgroup and less prototypical of the ingroup. In addition, exploratory analyses into possible mechanisms that facilitate prototypicality shifts suggested repeated presentations of ethnic cues and contingency awareness could be responsible. It is possible that changes in exemplar prototypicality could extend and generalise to other group exemplars through different types of contact (directly, vicariously etc.), which would have important consequences for intergroup relations in society. Therefore, the research presented in this thesis contributes in a number of ways to an ever-growing body of literature that aims to reduce negative stereotypes and improve intergroup relations in society.

References

Armony, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. *Neuropsychologia, 40*(7), 817-826.

Babel, M., & McGuire, G. (2015). Perceptual fluency and judgements of vocal aesthetics and stereotypicality. *Cognitive Science, 39*, 766-787.

Bandura, A. (1977). *Social learning theory*. New York: General Learning Press.

Baeyens, F., Eelen, P., Van Den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation, 21*, 434-455.

Best, J. D., Berghmans, S., Hunt, J. J. F. G, Clarke, S. C., Fleming, A., Goldsmith, P., & Roach, A. G. (2008). Non-associative learning in larval zebrafish. *Neuropsychopharmacology, 33*, 1206-1215.

Birtel, M. D., & Crisp, R. J. (2012). “Treating” prejudice: An exposure-therapy approach to reducing negative reactions towards stigmatized groups. *Psychological Science, 23*(11), 1379-1386.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242-261.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*(1), 5-25.

Bodenhausen, G. V. (1993). Emotions, arousal, and stereotypic judgments: A heuristic model of affect and stereotyping. In D. M. Mackie, & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp.

13-37). San Diego: Academic Press.

Bodenhausen, G. V, & Peery, D. (2009). Social categorization and stereotyping in vivo : The VUCA challenge. *Social and Personality Psychology Compass, 2*(3), 133–151.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(3), 219-231.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory, 11*, 485-494.

Bouton, M. E. (2014). Why behaviour change is difficult to sustain. *Preventive Medicine, 68*, 29-36.

Bouton, M. E., & Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning and Behavior, 22*(3), 317-324.

Brown, R., Eller, A., Leeds, S., & Stace, K. (2007). Intergroup contact and intergroup attitudes: A longitudinal study. *European Journal of Social Psychology, 37*, 692-703.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, Vol. 37 (pp. 255-342). San Diego, CA: Elsevier.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews, 14*, 365-376.

Cevik, A. O. (2014). Habituation, sensitization, and Pavlovian conditioning. *Frontiers in Integrative Neuroscience, 8*, 13.

Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black-White biracial group members. *Social Psychological and Personality Science, 6* (5), 544-550.

Christley, R. M. (2010). Power and error: Increased risk of false positive results in underpowered studies. *The Open Epidemiology Journal*, *3*, 16-19.

Coats, S., Latu, I. M., & Haydel, L. A. (2007). The facilitative effect of evaluative fit on social categorization. *Current Research in Social Psychology, 12,* 54 – 67.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.

Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorisation of multifaceted stimuli. *Journal of Personality and Social Psychology*, *77*(5), 927–941.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology, 88* (5), 770-789.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice, *Emotion, 9*(4), 585-591.

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes, *Psychological Science, 15*(5), 319-324.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53,* 161-186.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Feld, G. B., & Diekelmann, S. (2015). Sleep smart – optimizing sleep for declarative learning and memory. *Frontiers in Psychology, 6.* A622.

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing Muslim/Arab stereotypes through evaluative conditioning. *The Journal of Social Psychology, 153* (1), 6-9.

Garcia-Marques, T., & Mackie, D. M. (2007). Familiarity impacts person perception. *European Journal of Social Psychology, 37,* 839-855.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355-1361.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology, 49* (1), 120-125.

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology, 48,* 617-623.

Golkar, A., Bjornstjerna, M., & Olsson, A. (2015). Learned fear to social out-group members are determined by ethnicity and prior exposure. *Frontiers in Psychology, 6*, 123.

Greenberg, B. S., Mastero, D., & Brand, J. E. (2002). Minorities and the mass media: Television into the 21st century. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 333–352). Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, G., & Honey, R. C. (1990). Context-specific conditioning in the conditioned-emotional-response procedure. *Journal of Experimental Psychology: Animal Behavior Processes, 16*(3), 271-278.

Hamburger, Y. (1994). The contact hypothesis reconsidered: Effects of the atypical outgroup member on the outgroup stereotype. *Basic and Applied Social Psychology, 15*(3), 339-358.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–68). Taylor & Francis Group.

Harwood, J., Paolini, S., Joyce, N., Rubin, M., & Arroyo, A. (2011). Secondary transfer effects from imagined contact: Group similarity affects the generalization gradient. *British Journal of Social Psychology, 50,* 180-189.

Hewstone, M. (1994). Revision and change of stereotypic beliefs: In search of the elusive subtyping model. *European Review of Social Psychology, 5*(1), 69-109.

Hewstone, M. (2010). The contact space: A novel framework for intergroup contact research. *Journal of Language and Social Psychology, 29* (2), 147-177.

Hewstone, M., & Hamberger, J. (2000). Perceived variability and stereotype change. *Journal of Experimental Social Psychology, 36*, 103-124.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237-271.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390–421.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369-1385.

Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology, 11*, 223-255.

Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion, 5*(3), 267–276.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*(6), 640–643.

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.

Kattner, F., Ellermeier, W., & Tavakoli, P. (2012). Both trace and delay conditioning of evaluative responses depend on contingency awareness. *Learning and Motivation, 43*(1-2), 35–47.

Kruschke, J. K. (2011). Models of attention learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 120-152). New York: Cambridge University Press.

Kunda, Z., Davies, P. G., Adams, B. D., & Spencer, S. J. (2002). The dynamic time course of stereotype activation: Activation, dissipation, and resurrection. *Journal of Personality and Social Psychology, 82*(3), 283-299.

Labrenz, F., Icenhour, A., Benson, S., & Elsenbruch, S. (2015). Contingency awareness shapes acquisition and extinction of emotional responses in a conditioning model of pain-related fear. *Frontiers in Behavioural Neuroscience, 9*, 318.

Landman, N., Kuhn, M., Maier, J. G., Feige, B., Spiegelhalder, K., Riemann, D., & Nissen, C. (2016). Sleep strengthen but does not reorganize memory traces in a verbal creativity task. *SLEEP, 39*(3), 705-713.

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there. *Learning and Motivation*, *34*(3), 219-239.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry, 167*, 47-55.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3-26.

Maass, A., Castelli, L., Arcuri, L. (2000). Measuring prejudice: Implicit versus explicit techniques. In D. Copozza & R. Brown (Eds.), *Social identity processes* (pp. 96-116). London: Sage.

MacInnis, C. C., & Page-Gould, E. (2015). How can intergroup interaction be bad if intergroup contact is good? Exploring and reconciling an apparent paradox in the science of intergroup relations. *Perspectives on Psychological Science, 10*(3), 307-327.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93–120.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Calvini, G. (1999). Seeing more than we can know: Visual attention and category activation. *Journal of Experimental Social Psychology, 35*, 590-602.

Mallan, K. M., Sax, J., & Lipp, O. V. (2009). Verbal instruction abolishes fear conditioned to racial out-group faces. *Journal of Experimental Social Psychology, 45*(6), 1303–1307.

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects Applying social-cognitive theory to intergroup contact research. *Group Processes & Intergroup Relations, 14*(2), 255-274.

McIntyre, K., Paolini, S., & Hewstone, M. (2016). Changing people's views of outgroups through individual-to-group generalisation: Meta-analytic reviews and theoretical considerations. *European Review of Social Psychology, 27*(1), 63-115.

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology, 42*(4), 456-464.

Miller, N., & Dollard, J. (1941). *Social learning and imitation.* New Haven, NJ: Yale University Press.

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them". *Journal of Personality and Social Psychology, 99*(1), 62–77.

Minda, J. P., & Smith, D. J. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(2), 275-292.

Minda, J. P., & Smith, D. J. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 40-64). New York: Cambridge University Press.

Navarrete, C. D., McDonald, M. M., Asher, B.D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an out-group face in a minimal group context. *Evolution and Human Behavior, 33* (5), 590-593.

Navarrete, C. D., Olsson, A., Ho, A. K., Mendes, W. B., Thomsen, L., & Sidanius, J. (2009). Fear extinction to an out-group face: The role of target gender. *Psychological Science, 20* (2), 155-158.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotypes and social reality.* Oxford: Blackwell.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30,* 125-144.

Olson, M. A, & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433.

Olsson, A., & Phelps, E. A. (2004). Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science, 309*(5735), 785–787.

Paolini, S., Harris, N. C., & Griffin, A. S. (2016). Learning anxiety in interactions with the outgroup: Towards a learning model of anxiety and stress in intergroup contact. *Group Processes and Intergroup Relations*, *19* (3), 275-313.

Paolini, S., Hewstone, M., Cairns, E., & Voci, A. (2004). Effects of direct and indirect cross-group friendships on judgments of Catholics and Protestants in Northern Ireland: The mediating role of an anxiety-reduction mechanism. *Personality and Social Psychology Bulletin*, *30*, 770-786.

Pashler, H., Baylis, G. C. (1991). Procedural learning: II. Intertrial repetition effects in speeded-choice task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(1), 33-48.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytical test of the intergroup contact theory. *Journal of Personality and Social Psychology*, *90,* 751-783*.*

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis) liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*(1), 130-144.

Podsakoff, P. M., Mackenzie, S. B., Lee, J-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Ranganath, K. A, & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*(3), 249–254.

Rauchs, G., Feyers, D., Landeau, B., Bastin, C., Luxen, A., Maquet, P., & Collette, F. (2011). Sleep contributes to the strengthening of some memories over others, depending on hippocampal activity at learning. *The Journal of Neuroscience, 31*(7), 2563-2568.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning and Memory*, *11*, 501-509.

Richeson, J. A, & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology*, *89*(4), 517–530.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Oxford, England: Lawrence Erlbaum.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effect of intergroup contact. *Journal of Social Issues, 41*, 81-104.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology*, *32*(4), 309-336.

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review, 2*, 40-62.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis) categorization of unattractive brides and attractive prostitutes. *Experimental Psychology*, *55*(3), 182–188.

Scheepers, D., Spears, R., Doosje, B., & Manstead, A. S. R. (2006). The social functions of ingroup bias: Creating, confirming, or changing social reality. European Review of Social Psychology, 17, 359-396.

Schug, J., Alt, N. P., & Klauer, K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology*, 56, 121-125.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70*(6), 1126-1141.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.

Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., & Mackie, D. M. (2006). Familiarity can increase stereotyping. *Journal of Experimental Social Psychology, 42*, 471-478.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition, 8*(3), 243-262.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks/Cole.

Tan, A., Zhang, Y., Zhang, L., & Dalisay, F. (2009). Stereotypes of African Americans and media use among Chinese high school students. *The Howard Journal of Communication, 20*, 260-275.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., & Wetherell, M.S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Blackwell.

Van Dessel, P., De Houwer, J., & Gast, A. (2016). Approach-avoidance training effects are moderated by awareness of stimulus-action contingencies. *Personality and Social Psychology Bulletin, 42*(1), 81-93.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behaviour Research and Therapy, 43*(3), 323-336.

Verhoeven, K. J. F., Simonsen, K. L., & McIntyre, L. M. (2005). Implementing false discovery rate control: increasing your power. *Oikos, 108*, 643-647.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology, 49*, 661-669.

Walther, E., Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes, 32* (4), 454-459.

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science, 326*(5960), 1711–1714.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*(5), 815-827.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73-90.

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review, 15* (6), 1093-1099.

Appendix A: Ethics Approval for Main Study

HUMAN RESEARCH ETHICS COMMITTEE

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Notification of Expedited Approval

To Chief Investigator or Project Supervisor:	**Dr Andrea Griffin**
Cc Co-investigators / Research Students:	**Dr Stefania Paolini** **Mr Nicholas Harris** **Miss Jenna Pickard** **Mr Paul Williams**
Re Protocol:	**What's in a face? Intergroup Learning Study (Phase 3)**
Date:	**13-May-2009**
Reference No:	**H-2009-0104**
Date of Initial Approval:	**13-May-2009**
Approved To:	**12-May-2012**

Thank you for your **Response to Deferred** submission to the Human Research Ethics Committee (HREC) seeking approval in relation to the above protocol.
Your submission was considered under **Expedited** review by the HREC Panel.
I am pleased to advise that the decision on your submission is **Approved** effective **13-May-2009**.

Approval is granted to the date indicated above or until the project is completed, whichever occurs first. If the approval of an External HREC has been "noted" the approval period is as determined by that HREC.

The full Committee will be asked to ratify this decision at its next scheduled meeting. A formal *Certificate of Approval* will be available upon request. Your approval number is **H-2009-0104**.

If the research requires the use of an Information Statement, ensure this number is inserted at the relevant point in the Complaints paragraph prior to distribution to potential participants You may then proceed with the research.

Conditions of Approval

This approval has been granted subject to you complying with the requirements for *Monitoring of Progress*, *Reporting of Adverse Events*, and *Variations to the Approved Protocol* as detailed below.

PLEASE NOTE:
In the case where the HREC has "noted" the approval of an External HREC, progress reports and reports of adverse events are to be submitted to the External HREC only. In the case of Variations to the approved protocol, or a Renewal of approval, you will apply to the External HREC for approval in the first instance and then Register that approval with the University's HREC.

- ***Monitoring of Progress***

Other than above, the University is obliged to monitor the progress of research projects involving human participants to ensure that they are conducted according to the protocol as approved by the HREC. A progress report is required on an annual basis. You will be advised when a report is due.

- ***Reporting of Adverse Events***

1. It is the responsibility of the person **first named on this Approval Advice** to report adverse events.
2. Adverse events, however minor, must be recorded by the investigator as observed by the investigator or as volunteered by a participant in the research. Full details are to be documented, whether or not the investigator, or his/her deputies, consider the event to be related to the research substance or procedure.
3. 3. Serious or unforeseen adverse events that occur during the research or within six (6) months of completion of the research, must be reported by the person first named on the Approval Advice to the (HREC) by way of the Adverse Event Report form within 72 hours of the occurrence of the event or the investigator receiving advice of the event.
4. Serious adverse events are defined as:
   - o Causing death, life threatening or serious disability.
   - o Causing or prolonging hospitalisation.
   - o Overdoses, cancers, congenital abnormalities, tissue damage, whether or not they are judged to be caused by the investigational agent or procedure.
   - o Causing psycho-social and/or financial harm. This covers everything from perceived invasion of privacy, breach of confidentiality, or the diminution of social reputation, to the creation of psychological fears and trauma.
   - o Any other event which might affect the continued ethical acceptability of the project.

5. Reports of adverse events must include:
   - o Participant's study identification number; o date of birth;
   - o date of entry into the study; o treatment arm (if applicable); o date of event; o details of event;

- the investigator's opinion as to whether the event is related to the research procedures; and
- action taken in response to the event.

6. Adverse events which do not fall within the definition of serious, including those reported from other sites involved in the research, are to be reported in detail at the time of the annual progress report to the HREC.

 Variations to approved protocol

If you wish to change, or deviate from, the approved protocol, you will need to submit an *Application for Variation to Approved Human Research*. Variations may include, but are not limited to, changes or additions to investigators, study design, study population, number of participants, methods of recruitment, or participant information/consent documentation. **Variations must be approved by the (HREC) before they are implemented** except when Registering an approval of a variation from an external HREC which has been designated the lead HREC, in which case you may proceed as soon as you receive an acknowledgement of your Registration.

Linkage of ethics approval to a new Grant

HREC approvals cannot be assigned to a new grant or award (ie those that were not identified on the application for ethics approval) without confirmation of the approval from the Human Research Ethics Officer on behalf of the HREC.

Best wishes for a successful project.

Associate Professor Alison Ferguson
Chair, Human Research Ethics Committee

For communications and enquiries:

Human Research Ethics Administration

Research Services

Research Office
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 18999
F +61 2 492 17164
Human-Ethics@newcastle.edu.au

Appendix B: Ethics Approval for Pilot Studies

HUMAN RESEARCH ETHICS COMMITTEE
APPROVAL TO CONDUCT HUMAN RESEARCH

To Chief Investigator or Project Supervisor:	**Doctor Stefania Paolini**
Cc Co-investigators / Research Students:	**Miss Jenna Pickard** **Mr Nicholas Harris** **Mr Paul Williams** **Doctor Andrea Griffin**
Re Protocol:	**What's in a face? Intergroup Learning Study (Pilot Testing Phases 1 and 2)**
Date:	**09-Nov-2009**
Reference No:	**H-2009-0044**

Thank you for your recent application to the University of Newcastle Human Research Ethics Committee (HREC) for approval of the protocol identified above.

Details of previous approvals for Initial, Renewal and Variation applications are available upon request.

A *Certificate of Approval* is enclosed.

The *Certificate* and this advice are to be retained

They are important documents

- Note any comments related to the approval.
- **Where the HREC is the lead or primary HREC, if the research requires the use of an Information Statement, ensure the Reference No. is inserted into the complaints paragraph in the approved document(s) prior to distribution to potential participants.**
- Where the research is the project of a higher degree candidate, it is the responsibility of the project supervisor to ensure that the candidate receives this approval advice.

Conditions of Approval

This approval has been granted subject to you complying with the requirements for *Monitoring of Progress*, *Reporting of Adverse Events*, and *Variations to the Approved Protocol* as detailed below.

PLEASE NOTE:

In the case where the HREC has "noted" the approval of an External HREC, progress reports and reports of adverse events are to be submitted to the External HREC only. In the case of Variations to the approved protocol, you will apply to the External HREC for approval in the first instance and then Register that approval with the University's HREC.

- ***Monitoring of Progress***

Other than above, the University is obliged to monitor the progress of research projects involving human participants to ensure that they are conducted according to the protocol as approved by the HREC. The *Certificate of Approval* identifies the period for which approval is granted and your progress report schedule. A progress report is required on an annual basis, you will be advised when a report is due.

- ***Reporting of Adverse Events***

1. It is the responsibility of the person **first named on the Certificate** to report adverse events.

2. Adverse events, however minor, must be recorded by the investigator as observed by the investigator or as volunteered by a participant in the research. Full details are to be documented, whether or not the investigator, or his/her deputies, consider the event to be related to the research substance or procedure.

3. Serious or unforeseen adverse events that occur during the research or within six (6) months of completion of the research, must be reported by the person first named on the Certificate to the (HREC) by way of the Adverse Event Report form within 72 hours of the occurrence of the event or the investigator receiving advice of the event.

4. Serious adverse events are defined as:
    - Causing death, life threatening or serious disability.
    - Causing or prolonging hospitalisation.
    - Overdoses, cancers, congenital abnormalities, tissue damage, whether or not they are judged to be caused by the investigational agent or procedure.
    - Causing psycho-social and/or financial harm. This covers everything from perceived invasion of privacy, breach of confidentiality, or the diminution of social reputation, to the creation of psychological fears and trauma.
    - Any other event which might affect the continued ethical acceptability of the project.

5. Reports of adverse events must include:
    - Participant's study identification number;
    - date of birth;
    - date of entry into the study;
    - treatment arm (if applicable);
    - date of event;
    - details of event;
    - the investigator's opinion as to whether the event is related to the research procedures; and
    - action taken in response to the event.

6. Adverse events which do not fall within the definition of serious, including those reported from other sites involved in the research, are to be reported in detail at the time of the annual progress report to the HREC.

- ***Variations to approved protocol***

If you wish to change, or deviate from, the approved protocol, you will need to submit an *Application for Variation to Approved Human Research*. Variations may include, but are not limited to, changes or additions to investigators, study design, study population, number of participants, methods of recruitment, or participant information/consent documentation. **Variations must be approved by the (HREC) before they are implemented** except when Registering an approval of a variation from an external HREC which has been designated the lead HREC, in which case you may proceed as soon as you receive an acknowledgement of your Registration.

Linkage of ethics approval to a new Grant

HREC approvals cannot be assigned to a new grant or award (ie those that were not identified on the application for ethics approval) without confirmation of the approval from the Human Research Ethics Officer on behalf of the HREC.

With best wishes for a successful project.

Associate Professor Alison Ferguson

Chair, Human Research Ethics Committee

For communications and enquiries:

Human Research Ethics Administration

Research Services
Research Office
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 18999
F +61 2 492 17164
Human-Ethics@newcastle.edu.au

Linked University of Newcastle administered funding:

Funding body	Funding project title	First named investigator	Grant Ref
		,	

HUMAN RESEARCH ETHICS

COMMITTEE

Certificate of Approval

Applicant: (first named in application)	Doctor Stefania Paolini
Co-Investigators / Research Students:	Miss Jenna Pickard Mr Nicholas Harris Mr Paul Williams Doctor Andrea Griffin
Protocol:	What's in a face? Intergroup Learning Study (Pilot Testing Phases 1 and 2)

In approving this protocol, the Human Research Ethics Committee (HREC) is of the opinion that the project complies with the provisions contained in the *National Statement on Ethical Conduct in Human Research, 2007*, and the requirements within this University relating to human research.

Note: Approval is granted subject to the requirements set out in the accompanying document ***Approval to Conduct Human Research***, and any additional comments or conditions noted below.

Details of Approval

HREC Approval No: H-2009-0044 **Date of Initial Approval:** 17-Mar-2009

Approved to: 16-Mar-2010

Approval is granted to this date or until the project is completed, whichever occurs first. If the approval of an External HREC has been "noted" the approval period is as determined by that HREC.

Progress reports due: Annually.

If the approval of an External HREC has been "noted", the reporting period is as determined by that HREC.

Approval Details

Initial Application

15-Apr-2009

Approved

The Committee ratified the approval granted by the Chair on 17 March 2009 under the provisions for expedited review.

Appendix C: Sample Information Sheet Presented in Studies Chapter 3, 4 and 5

FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY

Dr Andrea Griffin,

School of Psychology
University of Newcastle
Callaghan NSW2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

INFORMATION SHEET

A Study on Face Processing and Emotions

V#7: 6/08/11 (all, lab)

Please read this Information Sheet carefully and be sure you understand its contents before you consent to participate. If there is anything you do not understand, or if you have any questions about the study, please ask the researcher.

Welcome and thank you for your interest in participating in this work. The purpose of our research is to investigate the role of learning in the acquisition of anxiety responses. Research in this area is important because by understanding whether healthy people learn to behave anxiously in certain situations, we can better help them overcome their fears.

Students from the University of Newcastle, as well as members of the general population, can participate in this research. Participation is entirely voluntary and you may withdraw at any time without giving a reason. However, we do not expect any serious risks or discomfort to arise from your participation.

This study will be conducted over two laboratory sessions, with the second laboratory session taking place at least five days and no more than four weeks after the first laboratory session. The first lab session should take approximately 30-45 minutes to complete whilst the second laboratory session will take approximately 1 hour and 45 minutes to 2 hours to complete.

During the first lab session you will be asked to complete a series of computerised tasks asking you to look at and judge a set of pictures (e.g., human faces), as well as complete some social demographic questions and report on past experiences with people similar and different from you. During the second lab session, while viewing pictures, you will receive a small number (between 5 and 10) of mild electric stimulations to your dominant hand's index finger. The level of the mild electrical stimulation that you will receive will be set yourself at the start of the laboratory session. The second laboratory session will also be recorded via a camera. This recording will be used by the research team to ensure that the tasks you complete are being attended to in the correct manner. This involves checking to make sure you are looking at the picture presentations and observing any hand movements that could affect the physiological recordings. The recording will only be used for this purpose. The recording will be kept confidential and will be stored for a maximum of three years. After this period of time the recording will deleted.

As you may feel somewhat apprehensive about the idea of being electrically stimulated, the following information has been designed to provide you with sufficient information about the techniques used to enable you to make an informed decision as to whether you wish to participate in this study. If you would like the technical details, please read these paragraphs; otherwise start again where text is in large font.

Electrical stimulation is widely used in research around the world to induce a slightly heightened level of anxiety or arousal. To give you an idea of the level of arousal we expect you will experience, the sensation will be similar to that you might experience when you watch a thriller, or when someone unexpectedly touches you on the shoulder while you work on your computer. Mild electrical stimulation is delivered by a purpose built device called a stimulus isolator that has passed internationally recognised, and very strict safety standards imposed by International Electrotechnical Commission (IEC; IEC60601), and which is widely used in human psycho-physiological research, including undergraduate teaching labs for students in Psychology and Human Physiology.

If you agree to participate, the first (index) finger of your dominant hand will be cleaned using an alcohol wipe before covering a small area on the last segment of the finger with a thin layer of standard (NaCl electrolyte) cream. This cream is to maximise the skin-electrode conductance. The cream will cause no discomfort. A small electrode will then be attached to the surface of your finger using tape. You will then be taken through a standard ‘work-up’ procedure, which allows you to determine your own individual level of stimulation intensity.

As part of the ‘work-up’ procedure, the researcher will show you the computer to which the electrode is connected and its operation. Then you will be asked to sample the stimulation, starting at the lowest intensity, which you will not be able to feel, and gradually increase the intensity in small increments yourself until a level is reached that you judge to be definitely uncomfortable, but not painful. The sensation you will experience is very similar to the one you have when you come into contact with static electricity.

You can be reassured that this procedure will not be painful, will not burn you and will not put you in any danger. Even if you were to set the level of stimulation at the maximum level this device can deliver, the current you would receive would be so small, and so short that it would not cause you any injury. In fact, several participants in our laboratory have chosen the maximum level of stimulation and still reported a positive experience with this research.

Once you have completed this procedure, you will be ready to take part in the lab portion of the study. Once again, you are free to withdraw from the project at any point during the ‘work-up’ procedure or there after.

Exclusion Criteria

Participants who have previously completed the “What’s In a Face Study” are ineligible to participate in this study as the required tasks are of a similar nature. To be eligible for this study you must come from a White Australian (Anglo/European) background.

Although risks to the healthy participant are negligible, please note that following the safety recommendations of the PowerLab manufacturer, people with specific health conditions should not undergo electrical stimulation. You will not be allowed to take part in this study if you respond affirmatively, to one or more of the questions below:

1. I have clinical anxiety
2. I have compromised peripheral sensation
3. I have a cardiac condition
4. I am equipped with an implanted or external pacemaker
5. I have a history of epileptic episodes
6. I have suffered a stroke
7. I am currently pregnant

We aim to collect anonymous data during this study. However, should the data collected during the laboratory session suggest that the study has caused your arousal to increase; the researcher will encourage you to provide your contact details and give permission for the researcher to contact you a few days after the laboratory session to ensure that you are feeling ok. Moreover, if you feel that anything in this study has brought up feelings of distress, please contact the University of Newcastle Counselling Service located in the Hunter building, either by telephone (49215801) or by e-mail (counselling@newcastle.edu.au). Their service is free of charge.

You will be free to withdraw from the study without giving a reason and without incurring an academic penalty. Your anonymity will be protected at all times during this study by use of a dummy code. The data from this study will be analysed and reported in an aggregated and de-identified manner and stored securely in the School of Psychology for a minimum of 5 years. You can request a summary of the research results by e-mailing the researchers at Andrea.Griffin@newcastle.edu.au any time after 1st December 2012. Alternatively, you will be able to complete a *Results Feedback Options* form at the end of the laboratory session to ask for the summary to be emailed to you after 1st December 2012.

First year psychology students are eligible to receive 5 points in course credit once the second laboratory session has been completed. Research volunteers are eligible to receive $25 travel reimbursement once the second laboratory session has been completed.

To participate, you will need to read and sign the Electric Stimulation Check Form and the Consent Form. Please keep this information sheet for your own reference. Any additional enquiries about the study may be directed to Dr A. Griffin, whose contact details appear at the top of this information sheet.

Thank you for considering this invitation. Your contribution to research in this area is important. By understanding mild anxious responses in healthy people, we can inform better interventions to help anxious people overcome their fears.

Research Team:

Chief Investigators:
Dr A. Griffin, Dr S. Paolini

Research Students:
N.Harris, K. Baldwin, S.Turnbull

Chief Investigators: Dr Griffin, Dr Paolini, Research Team: Nicholas Harris, Scott Turnbull, Alison Gosling, Elissa Ellis School of Psychology, The University of Newcastle, Telephone: (02) 49217161. This project has been approved by the University's Human Research Ethics Committee, Approval No. H-2009.0104. Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone 02 49216333, e-mail Human-Ethics@newcastle.edu.au

Appendix D: Sample Information Sheet Presented in Study 6

FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Dr Andrea Griffin,
School of Psychology
University of Newcastle
Callaghan NSW2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

INFORMATION SHEET
A Study on Bodily Responses During Gambling Under Varied Conditions.

V#12: 10/01/14 (all, lab)

Please read this Information Sheet carefully and be sure you understand its contents before you consent to participate. If there is anything you do not understand, or if you have any questions about the study, please ask the researcher.

Welcome and thank you for your interest in participating in this work. The purpose of our research is to investigate the role that various gambling situations has on you. Research in this area is important because by understanding whether healthy people learn to behave differently in certain situations, we can better help them overcome their feelings.

Students from the University of Newcastle, as well as members of the general population, can participate in this research. Participation is entirely voluntary and you may withdraw at any time without giving a reason. However, we do not expect any serious risks or discomfort to arise from your participation.

This study will be conducted over two laboratory sessions, with the second laboratory session taking place at least five days and no more than four weeks after the first laboratory session. The first lab session should take approximately 30 minutes to complete whilst the second laboratory session will take approximately 2 hours to complete.

During the first lab session you will be asked to complete a series of computerised tasks asking you to look at and judge a set of pictures (e.g., human faces), as well as complete some social demographic questions and report on past experiences with people similar and different from you. During the second lab session, you will first play a card guessing where you have the chance to win money. You will then play the card guessing game against other computerised players under various conditions. During the second laboratory session tasks you will be connected to various psychophysiological instruments to record you biofeedback throughout the tasks. These instruments are safe to use and we do not expect you to feel any discomfort. The second laboratory session will also be recorded via a camera. This recording will be used by the research team to ensure that the tasks you complete are being attended to in the correct manner. This involves checking to make sure you are looking at the screen and observing any hand movements that could affect the physiological recordings. The recording will only be used for this purpose. The recording will be kept confidential and will be stored for a maximum of three years. After this period of time the recording will deleted.

Exclusion Criteria
Participants who have previously completed the "What's In a Face Study" or "Study of Face Processing and Emotions " are ineligible to participate in this study as the required tasks are of a similar nature. To be eligible for this study you must come from a White Australian (Anglo/European) background.

To ensure the psychophysiological recordings work and valid data is collected, the following exclusion criteria have been identified.
You will not be allowed to take part in this study if you respond affirmatively, to one or more of the questions below:

1. I have clinical anxiety
2. I have compromised peripheral sensation
3. I have had surgery conducted to my face (ie., plastic surgery, botox)
4. I participate in gambling activities on a regular basis and consider myself to have a gambling problem.

If you feel that anything in this study has brought up feelings of distress, please contact the University of Newcastle Counselling Service located in the Hunter building, either by telephone (49215801) or by e-mail (counselling@newcastle.edu.au). Their service is free of charge.

You will be free to withdraw from the study without giving a reason and without incurring an academic penalty. Your anonymity will be protected at all times during this study by use of a dummy code. The data from this study will be analysed and reported in an aggregated and de-identified manner and stored securely in the School of Psychology for a minimum of 5 years. You can request a summary of the research results by e-mailing the researchers at Andrea.Griffin@newcastle.edu.au any time after 1st December 2014. Alternatively, you will be able to complete a *Results Feedback Options* form at the end of the laboratory session to ask for the summary to be emailed to you after 1st December 2014.

Psychology students are eligible to receive 5 points in course credit once the second laboratory session has been completed. Research volunteers are eligible to receive $20 travel reimbursement once the second laboratory session has been completed.

To participate, you will need to read and sign the Consent Form. Please keep this information sheet for your own reference. Any additional enquiries about the study may be directed to Dr A. Griffin, whose contact details appear at the top of this information sheet.

Thank you for considering this invitation. Your contribution to research in this area is important.

Research Team:

Chief Investigators:
Dr A. Griffin, Dr S. Paolini

Research Students:
N.Harris, S.Turnbull & A. Allan

Chief Investigators: Dr Griffin, Dr Paolini, Research Team: Nicholas Harris, Scott Turnbull, Alison Gosling, Elissa Dance, Olivia Gritten, Kaitlin Fitzgerald, Caroline Kuhne , Emma Sherwood, Alexandra Allan & Michael Cook School of Psychology, The University of Newcastle, Telephone: (02) 49217161. This project has been approved by the University's Human Research Ethics Committee, Approval No. H-2009.0104. Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone 02 49216333, e-mail Human-Ethics@newcastle.edu.au

Appendix E: Sample Consent Form

FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Dr Andrea Griffin,
School of Psychology
University of Newcastle
Callaghan NSW 2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

CONSENT FORM
A Study on Face Processing and Emotions

I have been invited to participate in the research project entitled "A Study on Face Processing and Emotions", supervised by Dr Griffin and Dr Paolini. I agree to participate in the above research project and give my consent freely by signing this form.

I understand that:
- ❖ The study will be carried out as described in the Information Sheet, a copy of which I have retained.
- ❖ The information that I provide during the investigation will be strictly confidential to the research team.
- ❖ I can withdraw from the project or any procedure **at any time** without penalty and do not have to give any reason for withdrawing.
- ❖ I have had all my questions answered and I understand that I will be fully debriefed about the rationale of the study at the end of my participation.

I consent to:
- o Have my data used for this experiment.
- o Attend two laboratory sessions with the second laboratory session taking place at least five days and no more than four weeks after the first laboratory session.
- o Having my second laboratory session recorded
- o The research team watching the recorded laboratory session when analysing the data
- o Look at a series of pictures presented on a computer.
- o Complete some rating scales to express my emotions to pictures.
- o Provide non-invasive psycho-physiological responses to the visual stimuli (heart rate, breathing and sweating). I understand that this procedure is neither painful nor unpleasant.
- o Receiving electric stimulation to my fingers or forearm while acknowledging that I do not have clinical anxiety, compromised peripheral sensation, a cardiac condition, an implantable or external cardiac pacemaker, or a history of epileptic episodes; nor have I suffered from stroke.
- o Receive 5-10 electrical stimulations to my fingers or forearm at a level that I will be free to set prior to the beginning of the testing session.
- o Participate in a study that may make me feel anxious.

Please print your name, add your signature and the date in the spaces provided below:

Name: …………………………………………………………………..

Signature: ………………………………………………………………

Date: ……………………………………………………………………..

V#6 04/03/13

Chief Investigators: Dr Griffin, Dr Paolini Research Team Nicholas Harris, Scott Turnbull, Alison Gosling, Elissa Dance, Olivia Gritten, Kaitlin Fitzgerald, Caroline Kuhne & Emma Sherwood School of Psychology, The University of Newcastle, Telephone: (02) 49217161. This project has been approved by the University's Human Research Ethics Committee, Approval No. H-2009-0104 Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone 02 49216333, e-mail Human-Ethics@newcastle.edu.au

Appendix F: Sample Electrical Stimulation Check Form

FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Dr Andrea Griffin,
School of Psychology
University of Newcastle
Callaghan NSW 2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

ELECTRIC STIMULATION CHECK FORM
A Study on Face Processing and Emotions

I have read the information sheet and understand that this study involves electric stimulation that will be delivered to me during the laboratory session of this study. I accept that I will receive electric stimulation and agree to participate. I state that I do NOT have

- Clinical anxiety
- Compromised peripheral sensation
- A cardiac condition
- An implantable or external cardiac pacemaker
- Any history of epileptic episodes
- A history of stroke
- Are pregnant

I have read this form and acknowledge that I do not have any of the 6 above conditions. I also confirm that the researcher has verbally checked that I do not have any of the above six conditions before asking me to read and sign the study's Consent Form.

Please print your name, add your signature and the date in the spaces provided below:

Name: ………………………………………………………………..

Signature: ……………………………………………………………

Date: …………………………………………………………………

V#2: 28/02/13

Chief Investigators: Dr Griffin, Dr Paolini Research Team: Nicholas Harris, Scott Turnbull, Alison Gosling, Elissa Dance, Olivia Gritten, Kaitlin Fitzgerald, Caroline Kuhne & Emma Sherwood School of Psychology, The University of Newcastle, Telephone: (02) 49217161. This project has been approved by the University's Human Research Ethics Committee, Approval No. H-2009-0104 Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone 02 49216333, e-mail Human-Ethics@newcastle.edu.au

Appendix G: Sample Debriefing Sheet Presented in Studies in Chapter 3, 4 and 5

FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Dr Andrea Griffin,
School of Psychology
University of Newcastle
Callaghan NSW 2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

A Study on Face Processing and Emotions
Debriefing Sheet

V#6: 28/02/13

Thank you very much for the time you have taken to participate in this study. We appreciate your contribution to our research project. Below you will find a description of the research rationale.

The Need for Confidentiality (All)
Since there is ample evidence that knowing the research hypotheses of a psychological study before taking part can invalidate its data, we would be grateful if you did not talk about this study and its nature to your friends and colleagues, who may also decide to participate. This way, future data collection will not be affected by their pre-existing knowledge about the nature of the study. We appreciate your confidentiality.

The Procedure and Research Hypotheses

(Study 3.1, 3.2 and 3.3)
At the start of this study, you were informed that the research was looking at the role of learning in the acquisition of anxiety responses. In particular, the research focused on how the anxiety affected your judgements about people from different groups and how these judgments affected other members from this group. This is important because by understanding whether healthy people learn to behave anxiously in certain situations, we can better help them overcome their fears.

We expect that the first lab session would be directly related to your learning results in the second lab session. Psychological research suggests that the extent to which one is familiar with a group will affect the extent to which we learn about this group in the future.

During the learning task, half of the participants were randomly allocated to a direct learning condition, where they directly experienced pairing of human faces with stimulation. The other half of participants were allocated to an indirect learning condition, where they did not have a personal experience of the pairing of human faces with stimulation. Instead they watched a video with another person experiencing the pairing.

Half of the participants from the direct and indirect condition were allocated to a slow face presentation condition, while the other half were allocated to a quick face presentation condition. In the slow face presentation condition, participants perceived faces with no interference. Participants in the quick face presentation condition were briefly presented with the face followed by a scrambled picture limiting the recognition of the first face in order to sample less controlled learning of the target faces.

(Study 4)
At the start of this study, you were informed that the research was looking at the role of learning in the acquisition of anxiety responses. In particular, the research focused on how the anxiety affected your judgements about people from different groups and how these judgments affected other members from this group. This is important because by understanding whether healthy people learn to behave anxiously in certain situations, we can better help them overcome their fears.

During the learning task, all participants were randomly allocated to a direct learning condition, where they directly experienced pairing of human faces with stimulation. After the learning task, half of the participants went on to complete an extinction procedure, where the human faces observed during the learning task were presented in the absence of any stimulation, so to extinguish any heightened anxiety developed as a result of the learning task. The other half of participants were allocated to a 'time condition' and just asked to complete a filler task, during which repeated presentations of unidentifiable faces were presented.

We expect the anxiety developed during the learning task and the repeated presentations of human faces during the extinction process to influence how participants categorise certain faces.

(Study 5.1)
At the start of this study, you were informed that the research was looking at the role of learning in the acquisition of anxiety responses. In particular, the research focused on how the anxiety affected your judgements about people from different groups and how these judgments affected other members from this group. This is important because by understanding whether healthy people learn to behave anxiously in certain situations, we can better help them overcome their fears.

We expect that the first lab session would be directly related to your learning results in the second lab session. Psychological research suggests that the extent to which one is familiar with a group will affect the extent to which we learn about this group in the future.

During the learning task, all participants were randomly allocated to a direct learning condition, where they directly experienced pairing of human faces with stimulation. However, half of the participants completed the task to a high familiarity group (White faces), while the other half of participants completed the task to a low familiarity group (Black faces). This manipulation was implemented to determine how familiarity affects learning.

(Study 5.2)

During the learning task, half of the participants were randomly allocated to be an under-estimator and the others were assigned to be an over-estimator. Following this, all participants viewed faces from each group, with one over-estimator and one under-estimator face being paired with an electric stimulation, whereas an additional over-estimator and under-estimator face was never paired with an electric stimulation.

This phase is known as acquisition. During acquisition, one face (CS+) is paired with an event, whereas another face (CS-) is never paired with an event. We are interested in generalisation, or spread, of this learned response. That is, it is expected that people change their physiological responses (sweating, breathing) and self-reported responses (anxiety, attractiveness) to a face paired with an event compared to a face not paired with such an event.

Merits and Significance of This Research (All)
This research is important because it aims to guide the development of effective strategies to reduce intergroup anxiety and intergroup tension in society at the individual-level and societal-level.

By combining learning theory, intergroup contact theory, and physiological measurements of autonomic arousal, the proposed work will form the basis of a powerful and novel approach to examining the role of learning in intergroup relationships. Understanding the nature of the learning mechanisms responsible for heightened autonomic responses to ethnic stimuli will guide the direction of future research and the development of effective strategies to reduce intergroup tension.

Final Note (All)
This data will clarify the far-reaching impact on intergroup relations of anxiety learning. Now that you have been fully informed about the nature of this research, you may wish to reconsider your decision to consent to participate. If this is the case, you do not need to give any reason for withdrawing from the research at this stage.

If you feel that you no longer wish to participate, inform the researcher and your data will be destroyed as soon as practical. If you wish to allow your data to be included in the research, then please indicate this decision to the researcher now. For practical reasons, we will be unable to withdraw your (de-identified) data at a later time unless you ask for this now.

Contact Details (All)
Again, thank you for taking part in this study. Your help is greatly appreciated. If you have any comments, queries, or complaints, or if you would like to know more about the research, then please contact Dr Andrea Griffin.

University Counselling and Health Services (All)
Should you feel that anything in this study has brought up feelings of distress, please contact the University of Newcastle Counselling Service located in the Hunter building on the Callaghan campus or the student support unit near the library at Ourimbah campus by telephone (49215801 at Callaghan or 4348 4060 at Ourimbah) or e-mail counselling@newcastle.edu.au). Their service is free of charge.

Should you feel that anything in this study has given rise to any physical health concern, please contact the University of Newcastle Health Service and make an appointment to see a General Practitioner. The service can be contacted by phone on 4921 6000 at Callaghan or 4348 4060 at Ourimbah. This service is free of charge.

Appendix H: Sample Debriefing Sheet Presented in Studies 6

FACULTY OF SCIENCE AND
INFORMATION TECHNOLOGY

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

Dr Andrea Griffin,
School of Psychology
University of Newcastle
Callaghan NSW 2308
Ph: +61 2 4921 7161
Fax: +61 2 4921 6980
E-mail: Andrea.Griffin@Newcastle.edu.au

A Study On Bodily Responses During Gambling Under Varied Conditions
Debriefing Sheet

V#9: 12/03/14

Thank you very much for the time you have taken to participate in this study. We appreciate your contribution to our research project. Below you will find a description of the research rationale.

The Need for Confidentiality
Since there is ample evidence that knowing the research hypotheses of a psychological study before taking part can invalidate its data, we would be grateful if you did not talk about this study and its nature to your friends and colleagues, who may also decide to participate. This way, future data collection will not be affected by their pre-existing knowledge about the nature of the study. We appreciate your confidentiality.

The Procedure and Research Hypotheses
At the start of this study, you were informed that the research was examining how you were affected by various gambling situations. In particular, the research focused on how gambling outcomes affected your judgements about people from different groups and how these judgments affected other members from this group. This is important because understanding how people learn to behave in positive and negative situations, we can better help them overcome their stereotypes and emotions.

All participants played gambling games against computerised players, where they experienced pairing of human faces with winning or losing money. Half of the participants were randomly allocated to an appetitive condition, where they always won money to a paired. In contrast, participants randomly allocated to the aversive condition always lost money to one paired face. Immediately prior to debriefing stage all participants completed the same card guessing game again in which the amount of money was made equal among participants. Hence everyone received the same cash reward no matter what condition they were allocated to. We expect the emotions developed during these two conditions to differ and influence how participants categorise certain faces.

Merits and Significance of This Research

This research is important because it aims to guide the development of effective strategies to reduce intergroup anxiety and intergroup tension in society at the individual-level and societal-level.

By combining learning theory, intergroup contact theory, and physiological measurements of autonomic arousal, the proposed work will form the basis of a powerful and novel approach to examining the role of learning in intergroup relationships. Understanding the nature of the learning mechanisms responsible for heightened autonomic responses to ethnic stimuli will guide the direction of future research and the development of effective strategies to reduce intergroup tension.

Final Note

This data will clarify the far-reaching impact on intergroup relations of learning. Now that you have been fully informed about the nature of this research, you may wish to reconsider your decision to consent to participate. If this is the case, you do not need to give any reason for withdrawing from the research at this stage.

If you feel that you no longer wish to participate, inform the researcher and your data will be destroyed as soon as practical. If you wish to allow your data to be included in the research, then please indicate this decision to the researcher now. For practical reasons, we will be unable to withdraw your (de-identified) data at a later time unless you ask for this now.

Contact Details

Again, thank you for taking part in this study. Your help is greatly appreciated. If you have any comments, queries, or complaints, or if you would like to know more about the research, then please contact Dr Andrea Griffin.

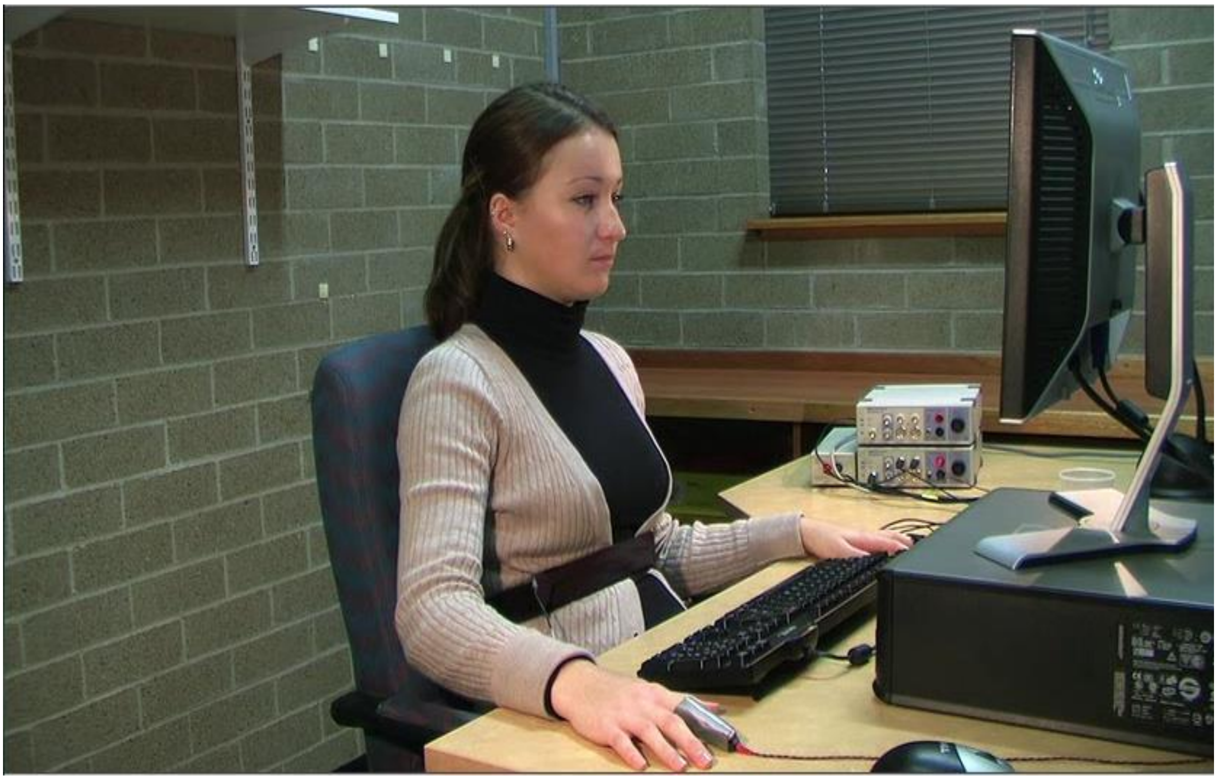
University Counselling and Health Services

Should you feel that anything in this study has brought up feelings of distress, please contact the University of Newcastle Counselling Service located in the Hunter building on the Callaghan campus or the student support unit near the library at Ourimbah campus by telephone (49215801 at Callaghan or 4348 4060 at Ourimbah) or e-mail counselling@newcastle.edu.au). Their service is free of charge.

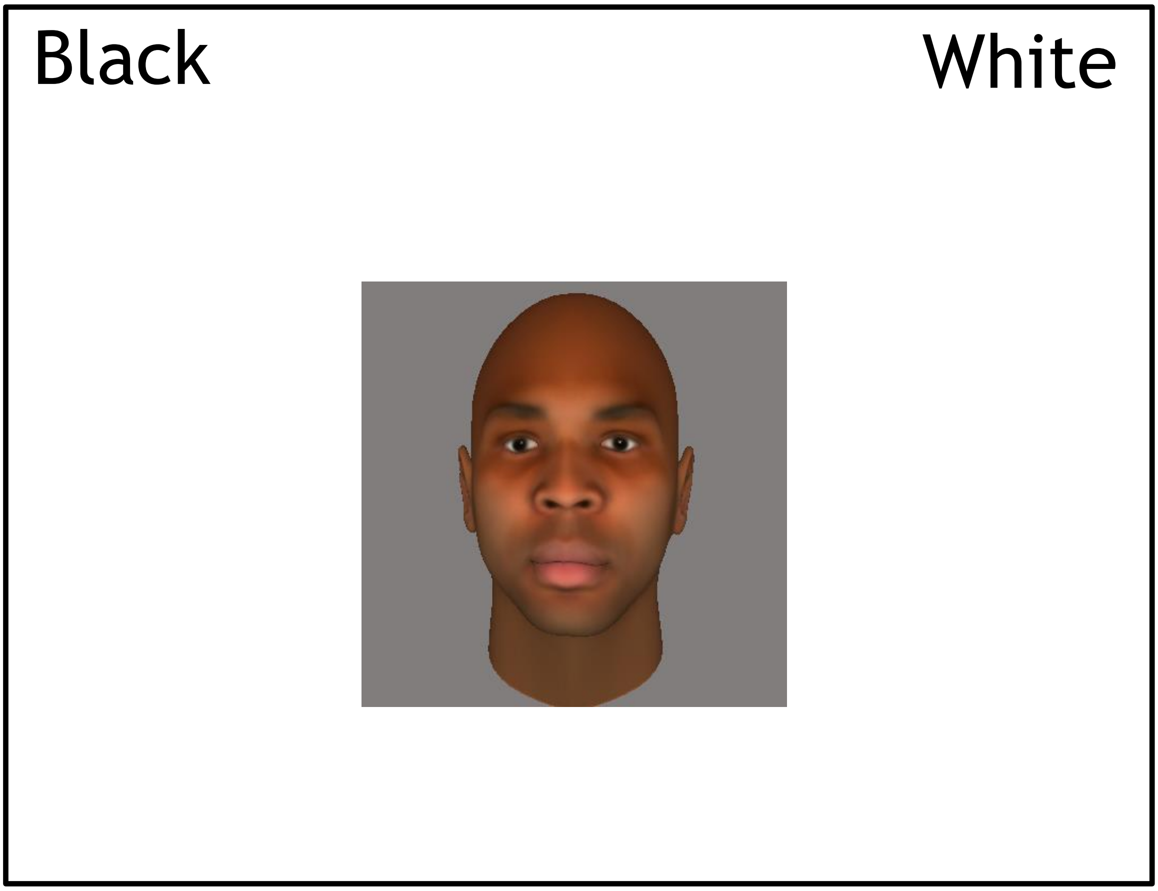
Should you feel that anything in this study has given rise to any physical health concern, please contact the University of Newcastle Health Service and make an appointment to see a General Practitioner. The service can be contacted by phone on 4921 6000 at Callaghan or 4348 4060 at Ourimbah. This service is free of charge.

We expect that the first lab session would be directly related to your learning results in the second lab session. Psychological research suggests that the extent to which one is familiar with a group will affect the extent to which we learn about this group in the future.

Appendix I: Still from the Vicarious Learning Video



Appendix J: Speeded Sorting Task Screenshot



Appendix K: Self-Reported Typicality Item Instructions

Prototypicality Rating Task

Soon you will be presented with a small set of faces.

We are interested in you indicating how PROTOTYPICAL you see each face is of Black people in general. In other words, we want to know how well each face fits your general idea about how Black individuals look. Remember: For the sake of this research, with the label 'Black' we refer to African-looking individuals and not Australian aboriginal individuals.

You will be asked to express your ratings on rating scales like the one below

PROTOTYPICALLY BLACK

Not at	○	○	○	○	○	○	○	*Very*
all	1	2	3	4	5	6	7	*much*

During the task, you can choose any number between 1 and 7 that best describes your most immediate reaction to the face, with 1 indicating that the face is not at all prototypical of black people, and 7 that the face is very prototypical of black people.

Please remember to respond based upon your first impression and be as frank as you can.

When you are ready to complete the prototypicality ratings, click the "Continue" button.

(new Screen)

Appendix L: Self-Reported Typicality Example Item

PROTOTYPICALLY BLACK

Not at all	1	2	3	4	5	6	7	*Very much*

Appendix M: Self-Reported Anxiety Instructions

Anxiety Rating Task

Soon you will be presented with a small set of faces.

For this block of ratings, we are interested in you indicating how ANXIOUS you anticipate you would be if you encounter each of these individuals while you walk on your own along a street at daytime.

You will be asked to express your ratings on rating scales like the one below

ANXIOUS

Not at all	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	*Very much*

During the task, you can choose any number between 1 and 7 that best describes your most immediate reaction to the face, with 1 indicating that the face does not make you at all anxious, and 7 indicating that the face makes you very anxious.

Please remember to respond based upon your first impression and be as frank as you can.

If you are unclear about these instructions, please ask the researcher now.

When you are ready to complete the anxiety ratings, click the "Continue" button.

Appendix N: Self-Reported Anxiety Example Item

ANXIOUS

Not at all	1	2	3	4	5	6	7	*Very much*
	○	○	○	○	○	○	○	

Appendix O: Contingency Awareness Questions

Phase 2 Thoughts

The questions below are interested in your thoughts of Phase 2. Remember that Phase 2 was when you viewed faces on the monitor whilst connected to the electrodes. Please answer the following questions about your thoughts of phase 2 only.

- How many different black faces did you see during Phase 2? (Provide a box with a drop down menu ranging from 1 to 30 and allow participants to select one)
- Do you think you reacted the same way to all the faces presented on the monitor in phase 2? (Yes/No)
- If “no” to the above question = Do you think you had a particular reaction pattern in the way you behaved towards the faces? Please describe this pattern in the open-box below. (Provide a text box for them to write response in with a minimum of 30 characters and max of 250)
- Which face was shown when you received the electrical stimulations? (Show the CS+ and CS- faces (face 2a and 4a) and have an option below each face allowing participants to select that face as the one that paired with the electrical stimulation).
- How confident are you that the face you selected above was shown when you received the electrical stimulations? (1 = not at all confident, 7 = very confident).

Appendix P: Additional Data Analysis

The following analyses report data collected from studies in this thesis that are not presented in-text. A large amount of data was collected in this thesis and presented in an already complex result section, which has the potential to convolute my key message. Results were reported in this appendix rather than in the main text when analyses supplemented results or when analyses focused on aspects that could potentially be of interest to researchers but did not inform the main focus of this research – exemplar prototypicality shifts.

Appendix P1: Study 3.1

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 learning type x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis conducted in the main text with a 2 learning type (direct, vicarious) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR data and last extinction block) mixed model ANOVA with face type and SCR test block as repeated measures. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. As expected a main effect of extinction block was detected, *F* (1, 64) = 40.60, *p* < .001, η_p^2 = .39. A paired samples *t*-test confirmed SCR decreased from post-test SCR data to the last block of extinction for the unsafe face (Post-test SCR: *M* = 1.38, *SD* = .44; Last extinction block *M* = 1.00, *SD* = .00), *t* (65) = 7.03, *p* <.001 and safe face (Post-test SCR: *M* = 1.10, *SD* = .18; Last extinction block *M* = 1.00, *SD* = .00), *t* (65) = 4.30, *p* <.001. This supplemental analysis suggests SCR responses decreased following extinction.

A one samples *t*-test (with a test statistic of 1 as this was added to data prior to square root transformation) was carried out to ensure SCR levels were brought to baseline. This analyses was carried out but the test could not be conducted because there was no variability in responses towards the unsafe and safe face (*M* = 1.00, SD = .00). Thus, all responses were 0 and represents no increases in SCR at the conclusion of extinction, demonstrating successful extinction.

Appendix P2: Study 3.2

Skin Conductance Responses: Measuring Negativity/Anxiety Generalisation to Face Variations and New Exemplars

I expanded the model from the unsafe and safe target faces used to check manipulations that are presented in the main text and included *variations* of the safe and unsafe target faces uninvolved in conditioning to test whether negativity/anxiety generalised to similar exemplars. A 2 learning type x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation gradient and time as repeated measures was conducted. A time main effect, F (1, 59) = 37.71, p < .001, ηp2= .39 and a time x generalisation gradient interaction were detected, F (2, 118) = 24.87, p < .001, ηp2= .30. More importantly a learning type x time x generalisation gradient was detected, F (2,118) = 11.42, p < .001,

$\eta p2 = .16$. This interaction was subsequently followed up by looking at direct and vicarious learning individually. A time x generalisation gradient interaction was found in both the direct, $F (2, 60) = 4.10$, $p = .021$, $\eta p2 = .12$ and vicarious learning conditioning, $F (2, 58) = 20.64$, $p < .001$, $\eta p2 = .42$.

In the direct learning condition, SCR's were higher at post-test than at pre-test for the target faces and 25% generalisation faces. No difference was detected for the 50% variation of faces. Results have been presented in Table 4 for simplicity. These results suggest anxiety generalised to similar faces (25% variations) but not to faces that varied greater in appearance (50% variations).

Table 4. Statistics for SCR effects detected in the Direct Learning condition

Exemplar	Mean		SD		Statistics	
	Pre-test	Post-test	Pre-test	Post-test	t-value (df=30)	p-value
target	1.02	1.14	.04	.19	-3.44	.002
25% variations	1.01	1.08	.01	.15	-2.65	.013
50% variations	1.02	1.06	.03	.13	Non-significant	Non-significant

In the vicarious learning condition, SCR were higher at post-test than at pre-test for the target faces and variations of the target faces. This results in presented in Table 5 and suggests non-associative anxiety shifts for all faces.

Table 5. Statistics for SCR effects detected in the Vicarious learning condition.

Exemplar	Mean		SD		Statistics	
	Pre-test	Post-test	Pre-test	Post-test	t-value (df=30)	p-value
target	1.01	1.44	.02	.41	-5.66	<.001
25% variation	1.03	1.16	.05	.24	-3.23	.003
50% variation	1.01	1.11	.03	.18	-2.93	.007

To test whether anxiety generalised to *new exemplars*, participant's similarity ratings were used as described previously for Study 3.1. The new variable distinguished between those who perceived each new exemplar as being more similar to the unsafe face than the safe face, as being more similar to the safe face than the unsafe face, or equally similar to both. Individual analyses were carried out for each new Black face, using a 2 learning type x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on the SCR data. For the first new Black face, a marginal main effect of time indicated higher SCRs responses were found at post-test (M = 1.09, SD = .26) than at pre-test (M = 1.02, SD = .05), F (1, 53) = 3.81, p = .056, η_p^2 = .07. A main effect of time was found for the second new Black face and higher SCRs responses were found at post-test (M = 1.12, SD = .26) than at pre-test (M = 1.01, SD = .03), F (1, 53) = 4.97, p = .030, η_p^2 = .0. This suggests anxiety generalised to the new Black faces, irrespective of whether they were similar to the unsafe face or not. No other effects were significant and face similarity did not qualify the effects.

Self-reported Anxiety: Measuring Negativity/Anxiety Generalisation to Face Variations and New Exemplars

I expanded the model from the unsafe and safe target faces used to check manipulations that are presented in the main text and included *variations* of the safe and unsafe target faces uninvolved in conditioning to test whether negativity/anxiety generalised to similar exemplars. A 2 learning type by 2 post-test position by 2 face type by 3 generalisation gradient (target, 25%, 50%) by 2 time mixed model ANOVA with face type, generalisation and time as repeated measures was conducted on self-reported anxiety data. A generalisation main effect was found, F (2,112) = 23.74, p <.001, η_p^2= .30; the expected generalisation gradient was found with target faces rated highest in anxiety (M = 3.46, SD = 1.61), followed by the 25% variation (M = 3.06, SD

= 1.33) and finally the 50% variation (*M* = 2.85, *SD* = 1.31). Unexpectedly, all other effects were non-significant, all *p'*s > .086. These results suggest anxiety did not generalise across time to face variations on the self-reported anxiety measure. No generalisation effects to face variations were detected; hence, no further analyses were carried out on the variation data.

To test whether anxiety generalised to *new exemplars*, participant's similarity ratings were used as described previously for Study 3.1. The new variable distinguished between those who perceived each new exemplar as being more similar to the unsafe face than the safe face, as being more similar to the safe face than the unsafe face, or equally similar to both. Individual analyses were carried out for each new Black face, using a 2 learning type x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure on the self-reported anxiety data. No effects were detected for either exemplar, all *p*'s > .162.

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 learning type x 2 post-test position (after acquisition, after extinction) x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in the main text with a 2 learning type (direct, vicarious) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR data and last extinction block) mixed model ANOVA with face type and SCR test block as repeated measures. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. As expected a main effect of extinction block was detected, F (1, 57) = 5.50, p = .023, η_p^2 = .09. A paired samples t-test confirmed SCR decreased from post-test SCR to the last block of extinction for the unsafe face (post-test SCR: M = 1.26, SD = .30; last extinction block M = 1.00, SD = .01), t (60) = 6.68, p <.001 and safe face (post-test SCR: M = 1.22, SD = .29; last extinction block M = 1.00, SD = .00), t (60) = 5.95, p <.001. All other effects were non-significant and this analysis suggests anxiety was extinguished.

A one samples t-test with a test statistic of 1 (a constant of 1 was added prior to square root transformation) confirmed heightened anxiety was brought back to baseline for the unsafe face, t (60) = 1.622, p = .110. This analyses was carried on the safe face data but the test could not be conducted because there was no variability in responses towards the safe face (M = 1.00, SD = .00). Thus, all responses were brought back to baseline, demonstrating successful extinction.

Appendix P3: Study 3.3

Skin Conductance Responses: Measuring Negativity/Anxiety Generalisation to Face Variations and New Exemplars

I expanded the model from the unsafe and safe target faces used to check manipulations that are presented in the main text and included *variations* of the safe and unsafe target faces uninvolved in conditioning to test whether negativity/anxiety generalised to similar exemplars. A 2 learning type x 2 face type x 3 generalisation gradient (target, 25%, 50%) x 2 time mixed model ANOVA with face type, generalisation gradient and time as repeated measures. The time main effect previously found held, and more importantly, I found a generalisation gradient x time interaction, F (2,112) = 17.56, $p < .01$, $\eta_p^2 = .24$, which I followed up by looking at each face variation separately. A non-associative specific change in negativity/anxiety in the same direction as target faces was found for 25% variation faces as evidenced by a time main effect, F (1, 56) = 8.68, $p = .005$, $\eta_p^2 = .13$ (pre-test M = 1.02, SD = .04; post-test M = 1.07, SD = .18). The main effect was further qualified by learning type, F (1, 56) = 8.96, $p = .004$, $\eta_p^2 = .14$. The interaction with learning type demonstrated the vicarious learning condition was larger than the direct learning condition. The main effect of time was replicated for 50% variation faces, F (1, 55) = 9.82, $p = .003$, $\eta_p^2 = .15$ (pre-test M = 1.01, SD = .03; post-test M = 1.09, SD = .22). Hence, a non-associative shift in negativity/anxiety was found for all generalisation faces on the SCR measure, which was similar for the effect detected for target faces.

I investigated whether this effect extended to new exemplars with a 2 learning type x 3 new face similarity x 2 time mixed model ANOVA with time as the repeated measure carried out individually for each exemplar. The time main effect observed for generalised faces held for the first new exemplar, F (1, 54) = 5.94, $p = .018$, $\eta_p^2 = .01$. The time main was replicated for the second new exemplar, F (1, 54) = 9.45, $p = .003$, $\eta_p^2 = .14$, but was further qualified by learning type, F (1, 54) = 4.03, $p = .050$, $\eta_p^2 = .07$. I investigated this effect further by looking at each level of learning type separately and found the changes in negativity/anxiety were larger in the vicarious learning condition than the direct learning condition. All other effects were non-significant. Together, these results suggest non-associative negativity/anxiety changes generalised to similar and new Black exemplars.

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 learning type x 2 post-test position (after acquisition, after extinction) x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in the main text with a 2 learning type (direct, vicarious) x 2 post-test position (after acquisition, after extinction) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR data and last extinction block) mixed model ANOVA with face type and SCR test block as repeated measures on SCR. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. A non-significant main effect of extinction block was detected, F (1, 54) = 2.50, p = .120, η_p^2 = .04. Despite the non-significant result a paired samples t-test confirmed the SCR decreased from post-test SCR data to the last block of extinction for the unsafe face (post-test SCR: M = 1.22, SD = .26 ; last extinction block M = 1.00, SD = .00), t (57) = 6.25, p <.001 and safe face (post-test SCR: M = 1.21, SD = .31 ; last extinction block M = 1.00, SD = .00), t (60) = 5.07, p <.001. All other effects were non-significant and this analysis suggests anxiety was extinguished.

A one samples t-test with a test statistic of 1 (a constant of 1 was added prior to square root transformation) confirmed heightened anxiety was brought back to baseline for the unsafe face, *t* (57) = 1.00, *p* = .322 and safe face, *t* (57) = 1.00, *p* = .322. Together, these analyses suggest all responses were brought back to baseline, demonstrating successful extinction.

Appendix P4: Study 4.1

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 type of extinction x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in main text with a 2 type of extinction (individuality group, category membership group) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR and last extinction block) mixed model ANOVA with face type and SCR test block as repeated measures. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. As expected, a significant main effect of extinction block was detected, F (1, 57) = 4.72, p = .034, η_p^2 = .08. A paired samples *t*-test confirmed SCR decreased from post-test SCR to the last block of extinction for the unsafe face (post-test SCR: *M* = 1.20, *SD* = .27; last extinction block *M* = 1.01, *SD* = .06), *t* (58) = 4.98, *p* <.001, and safe face (post-test SCR: *M* = 1.09, *SD*

= .18; last extinction block *M* = 1.01, *SD* = .02), *t* (58) = 4.06, *p* <.001. All other effects were non-significant and this analysis suggests anxiety was extinguished.

A one samples *t*-test with a test statistic of 1 (a constant of 1 was added prior to square root transformation) confirmed heightened anxiety was brought back to baseline for the unsafe face, *t* (58) = 1.24, *p* = .218. Together, these analyses suggest responses were brought back to baseline, demonstrating successful extinction.

Appendix P5: Study 5.1

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition vs after extinction) x 2 target group (ingroup and outgroup) x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in main text with a 2 participant's group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition vs after extinction) x 2 target group (ingroup and outgroup) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR data and last extinction block) mixed model ANOVA with target group, face type and extinction block as repeated measures. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. As expected, there was a significant main effect of extinction block was detected, F (1, 49) = 9.86, p = .003, η_p^2 = .17. A paired samples t-test confirmed SCR decreased from post-test SCR to the last block of extinction for the unsafe face (post-test SCR: M = 1.10, SD = .19; last extinction block M = 1.00, SD = .00), t (52) = 4.00, p <.001 and safe face (post-test SCR: M = 1.02, SD = .07; last extinction block M = 1.00, SD = .00), t (52) = 2.21, p <.032. The 3 way interaction involving target group, face type and extinction block was also significant. The 3 way interaction was driven by the outgroup unsafe face decreasing by a larger amount compared to the other faces, but all last block extinction means were the same. All other effects were non-significant and this analysis suggests anxiety was extinguished.

A one samples t-test with a test statistic of 1 (a constant of 1 was added to data prior to square root transformation) was carried out to ensure SCR levels were brought back to baseline. This analyses was carried out but the test could not be conducted because there was no variability in responses towards the unsafe and safe face (M = 1.00, SD = .00). Thus, all responses were 0 and represents no increases in SCR at the conclusion of extinction, demonstrating successful extinction.

Self-reported Anxiety Data: Acquisition and Extinction Analyses

I also checked for effective acquisition and extinction with self-reported anxiety data. I performed a 2 participant group membership (under-estimator or over-estimator) x 2 post-test position (after acquisition vs after extinction) x 2 target group (ingroup and outgroup) x 2 face type (unsafe and safe face) x 2 time (pre- and post-test) mixed model ANOVA with target group, face type and time as repeated measures. A face type x time interaction, F (1, 53) = 4.46, p = .039, ηp^2 = .08 and a face type x time x target group interaction were detected, F (1, 53) = 4.97, p = .030, ηp^2 = .09. I investigated the face type x time x target group interaction further and conducted the same analyses separately for each target group (ingroup and outgroup).

In the Ingroup condition the face type x time interaction held, F (1, 53) = 8.98, p = .004, ηp^2 = .15. Paired samples *t*-test suggested the unsafe ingroup face was perceived as being more anxiety provoking post-test (*M* = 3.93, *SD* = 1.98) than at pre-test (*M* = 2.56, *SD* = 1.28), *t* (56) = 5.48, $p < .001$. A trend for the ingroup safe face to be perceived as being more anxiety provoking at post-test (*M* = 3.30, *SD* = 1.84) than at pre-test (*M* = 2.84, *SD* = 1.54) was detected but the effect only marginally significant, *t* (56) = 1.99, *p* = .052. This set of results suggest greater anxiety developed for the unsafe exemplar. The non-significant interaction involving post-test position provides some evidence that self-reported anxiety was slightly elevated after extinction, despite the rating being relatively low (below the midpoint of a 7 point scale). All participants underwent additional steps after extinction to diffuse any residual change in self-reported anxiety.

When following up the 3 way interaction for the Outgroup condition, the face type x time interaction was non-significant, F (1, 53) = .46, p = .501, ηp^2 = .01. Instead a time main effect was detected, F (1, 53) = 23.34, $p < .001$, ηp^2 = .36. Paired samples-*t*-tests suggested the outgroup unsafe face was perceived as being more anxiety provoking at post-test (*M* = 4.30, *SD* = 2.10) than at pre-test (*M* = 2.86, *SD* = 1.41), *t* (56) = 5.10, $p < .001$. Similarly, the outgroup safe face was perceived as being more anxiety provoking at post-test (M = 3.77, *SD* = 1.92) than at pre-test (*M* = 2.61, *SD* = 1.46), *t* (56) = 4.19, $p < .001$. Together these results suggest both outgroup faces were perceived as being more anxiety provoking after acquisition (vs before). The non-significant interaction involving post-test position provides some evidence that self-reported anxiety was slightly elevated after extinction, despite the rating being relatively low. Similar to the ingroup condition, all participants underwent additional steps after extinction to diffuse any residual change in self-reported anxiety.

Appendix P6: Study 5.2

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a target ethnicity x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in main text with a 2 target ethnicity (ingroup vs outgroup) x 2 face type (unsafe and safe face) x 2 SCR test block (post-test SCR data and last extinction block) mixed model ANOVA with face type and SCR test block as repeated measures. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will compare SCR post-test data (collected immediately after acquisition) to the final presentations during extinction to ensure anxiety was extinguished. As expected, a significant main effect of extinction block was detected, F (1, 60) = 7.45, p = .008, η_p^2 = .11. A paired samples t-test confirmed SCR decreased from post-test SCR to the last block of extinction for the unsafe face (post-test SCR: M = 1.17, SD = .26; last extinction block M = 1.00, SD = .02), t (61) = 5.34, p <.001 and safe face (post-test SCR: M = 1.11, SD = .18; last extinction block M = 1.01, SD = .06), *t* (61) = 3.87, *p* <.001. All other effects were non-significant and this analysis suggests anxiety was extinguished.

A one samples *t*-test with a test statistic of 1 (a constant of 1 was added prior to square root transformation) confirmed heightened anxiety was brought back to baseline for the unsafe face, *t* (61) = 1.74, *p* = .087 and safe face, *t* (61) = 1.88, *p* = .065. Together, these analyses suggest responses were brought back to baseline, demonstrating successful extinction.

Appendix P7: Study 6.1

Electromyography (EMG) Method and Results

Facial expressions were measured via muscle activity as a marker for effect of the faces through electromyography (EMG) using an ADInstruments octal bio amp (ML 138) with MLA4105 biopotenial electrodes. MLA4105 shielded biopotenial are domed Ag/AgCl 4mm diameter electrodes, which were connected to an octal bio amp and filled with standard electrolyte (Surgicon Electro Gel). A ground electrode was attached to the midline or centre of the forehand, two electrodes attached to the corrugator supercilii and two electrodes attached to the zygomatic major following the guidelines proposed by Fridlund & Cacioppo (1986). The corrugator supercilii measures negative affect towards stimuli and the zygomatic major measures positive affect towards stimuli, potentially providing a measure of valenced responses in the win/lose scenarios (see below) (Cacioppo, Petty, Losch & Kim, 1986; Larsen, Norris & Cacioppo, 2003; Witvliet & Vrana, 1995). Thus, higher corrugator supercilii activity towards exemplars paired with money loss post conditioning was expected in the lose conditioning treatment. Higher zygomatic major activity towards the exemplar paired with money gain was expected in the win conditioning treatment post-conditioning.

EMG data was collected and analysed following standard guidelines (Fridlund & Cacioppo, 1986; Tassinary & Cacioppo, 2000). A ground electrode was attached to the midline of the participant's forehead, approximately 3-4cm superior to the upper borders of the inner brow. One pair of electrodes was attached to the skin over the corrugator supercilii above the left eye and another of pair electrodes attached over the zygomatic major on the left hand side of the face. Recording sites were prepared with alcohol swabs, followed by a mild skin abrasive. A Butterworth filter with a frequency band of 30-500Hz was applied to data by software as part of the data. Recordings that corresponded to artefacts (a response exceeding 100 units in the 500 ms prior to face onset were excluded from the analysis. Mean muscle activity across each face presentation at pre-test (2 presentations) and at post-test (2 presentations) were calculated as the difference from baseline activity, measured 500 ms before stimulus presentation.

EMG data was analysed with a 2 conditioning treatment (win vs lose) x 2 face type (paired and unpaired face) x 2 time (pre-conditioning/post-extinction) mixed model ANOVA was conducted with face type and time as the repeated measures on the corrugator muscle activity. Due to technical error with the equipment, analyses on EMG data were restricted to 55 participants (out of 92 participants). As corrugator activity increases with frowning, this analysis investigates whether negative valenced evaluations became associated with an exemplar following conditioning. Unexpectedly, no interactions involving face type, time and conditioning treatment were detected, *p'*s > .358.

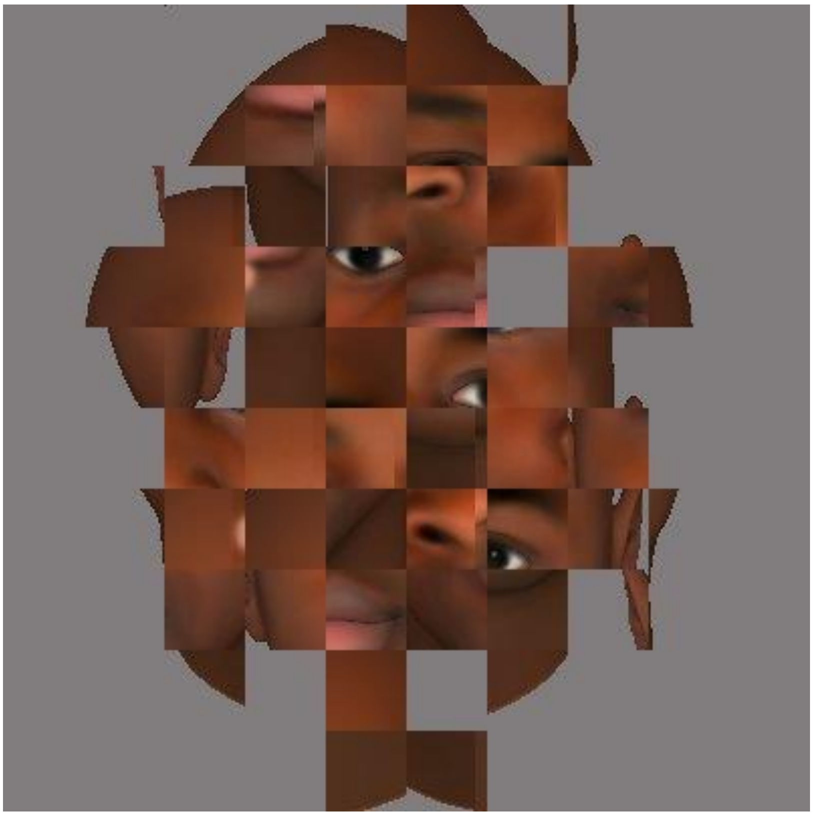
A 2 conditioning treatment (win vs lose) x 2 face type (paired and unpaired face) x 2 time (pre-conditioning/post-extinction) mixed model ANOVA was conducted with face type and time as repeated measures on the zygomatic muscle activity. As zygomatic activity increases with smiling, this analysis investigates whether positive valenced evaluations became associated with an exemplar following conditioning. There were no significant interactions involving face type or time, *p*'s >.329. When the analysis was extended to include gender results remained similar, *p*'s > .431. Together, EMG results failed to detect any significant effects, which suggests positive and negative valence did not become associated with faces. This result should be followed up further as the technical error resulted in a large number of participants being excluded from the analyses.

Skin Conductance Responses: Supplemental Analyses that Suggest Successful Extinction of Negativity/Anxiety

The analysis presented in the main text checked for successful extinction with a 2 conditioning treatment x 2 face type x 2 extinction block (first block and ten trial block) mixed model ANOVA with face type and extinction block as repeated measures on SCR extinction data. The analysis presented in the main text found evidence for successful extinction. I supplemented the analysis carried out in main text with a 2 conditioning treatment (win vs lose) x 2 face type (paired and unpaired face) x 2 extinction block (post-test acquisition and last extinction block) mixed model ANOVA with face type and extinction block as repeated measures on SCR. In this study extinction was methodologically extinguished by the continuously presenting faces throughout extinction until no increases in SCR responses were observed across 4 consecutive trials (minimum number of trials was 10). Hence, participants may have observed more than 10 presentations during extinction and this analysis will ensure emotions were extinguished. As expected, a significant extinction block main effect was detected, $F\ (1, 52) = 23.69$, $p < .001$, $\eta_p^2 = .31$. A paired samples *t*-test confirmed the SCR decreased from the acquisition block to the last block of extinction for the paired face (Acquisition block: M = 1.09, SD = .19; Last extinction block M = 1.00, SD = .03), t (53) = 3.85, p <.001 and unpaired face (Acquisition block: M = 1.07, SD = .12; Last extinction block M = 1.01, SD = .00), t (53) = 4.22, p <.001. All other effects were non-significant and this analysis suggests anxiety was extinguished.

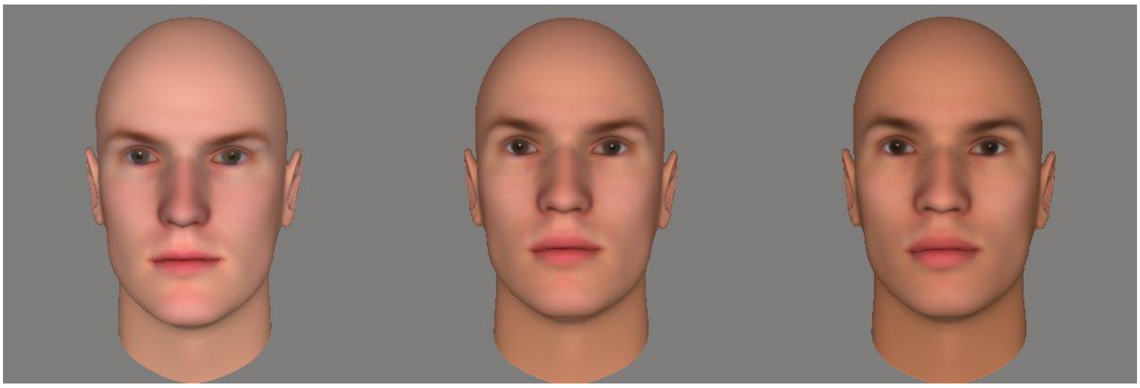
A one samples t-test with a test statistic of 1 (a constant of 1 was added prior to square root transformation) confirmed heightened emotions was brought back to baseline for the paired face, t (53) = 1.31, p = .197 and unpaired face, t (53) = 1.48, p = .146. Together, these analyses suggest responses were brought back to baseline, demonstrating successful extinction.

Appendix Q: Masking Stimulus



Appendix R: White Faces Presented During Studies

Original Face (CS+/CS-)	25% Variation Face	50% Variation Face



New Exemplar Faces