THE MECHANISMS BEHIND FEATURE-BASED STEREOTYPING: AN
EMPIRICAL TEST

A thesis submitted in partial fulfillment of the requirements for the
Degree of Master of Arts in Psychology, General-Experimental

By

Heather R. Rees

May 2014
The thesis of Heather R. Rees is approved by:

______________________________________________________________________  __________________
Naveen Khetarpal, Ph.D.  Date

______________________________________________________________________  __________________
Erica L. Wohldmann, Ph.D.  Date

______________________________________________________________________  __________________
Debbie S. Ma, Ph.D., Chair  Date

California State University, Northridge
ACKNOWLEDGEMENTS

I would like to thank my committee members Dr. Naveen Khetarpal, Dr. Erica Wohldmann, and Dr. Debbie Ma for serving on my committee and for their insightful comments throughout the thesis process. Specifically I would like to thank Naveen for taking the time to meet with me on multiple occasions to help improve my methods. In addition, I would like to thank Erica for her design suggestions on both this research project and others. In particular I am thankful to Debbie for being my mentor. Allowing me to join your lab and taking me on as a graduate student completely changed my life for the better. I am grateful for your high expectations and time you spent helping me with writing, statistics, and research skills. I will never forget all of your help, and look forward more conversations about science and collaboration in the future.

I would also like to thank my first mentor Barbara Drescher for setting me on the path of science; the ideals I learned in your methods course have been irreplaceable to me as a person and as a scientist. I am also thankful to my mother, grandmother, and Luffy for supporting me and being understanding of the never-ending nature of my work. In particular I thank my mother, Teresa Rees, who has by far always been the most supportive person in my life and is generally an exemplary human being. Because of your sacrifices I will be the first in our family to go to a doctoral program, and I cannot express how thankful I am for your making that possible. I would also like to thank my friends who help keep me excited about research and prevent me from burning out. In particular, I want to thank Maxim Babush for discussing theory with me and helping me maintain my enthusiasm. I am grateful for all of what you bring to my life both inside and outside of academia.
# TABLE OF CONTENTS

SIGNATURE PAGE

ACKNOWLEDGEMENTS

ABSTRACT

INTRODUCTION

The Role of Categorization in Stereotyping

Feature-Based Stereotyping

Mechanisms Underlying Feature-Based Stereotyping

Direct feature-trait association

Graded categories

Disrupting Categorical Processing

Present Research

METHOD

Participants

Design

Stimuli

Questionnaire

Procedure

RESULTS

Signal Detection Analysis
ABSTRACT

THE MECHANISMS BEHIND FEATURE-BASED STEREOTYPING: AN EMPIRICAL TEST

By

Heather R. Rees

Master of Arts in Psychology

General-Experimental

Previous studies have found evidence for “feature-based stereotyping” (FBS), in which the more “prototypic” an individual’s features are of their social group, the more that individual will be stereotyped. Both categorization and a direct association between features and stereotypes (Blair, 2006) have been proposed to mediate of this process. The current study provides an empirical test of the mechanisms underlying FBS by employing verbal interference, which has been used in previous studies to disrupt categorical perception (the advantage for discriminating between objects from different categories, relative to objects from the same category). We predicted that if categorization mediates FBS, verbal interference will eliminate the use of features in stereotyping. If features are directly associated with stereotypes, categorization should be unnecessary and verbal interference will have no effect on FBS. Although no evidence was found to support the differential effects of interference on FBS, analyses suggest that our interference tasks did not work as intended.
INTRODUCTION

On a daily basis people are exposed to, and attend to, an immense amount of information. Much of the information that people encounter will be novel, which would quickly become overwhelming without a coherent structure to manage it with. To effectively organize this information, people utilize “categories”, which according to Allport: “The human mind must think with the aid of categories....Once formed, categories are the basis for normal prejudgment. We cannot possibly avoid this process. Orderly living depends upon it” (Allport, 1954, p.20). Categories are formed around “prototypes”, exemplars that possess the highest amount of features that are typical of the category, and the fewest amount of features that are atypical of it (Rosch, 2002). The structuring of categories around prototypes facilitates differentiating between targets that belong to different categories, compared to targets from the same category —a phenomenon referred to as “categorical perception” (e.g., Harnad, 1987).

While categories support a predictable and stable existence, they can also lead to overgeneralization. People give a disproportionate amount of attention to how targets differ between categories, while overlooking within-category differences (Harnad, 1987; Stangor, Lynch, Duan, & Glass, 1992; Taylor, Fiske, Etoff, & Ruderman, 1978). In this way, people tend to deemphasize continuous difference in category membership, and instead perceive discrete differences at certain boundary points. Categorical perception also applies to person perception, where, despite variation in fit to a racial category, social group membership is perceived as distinctly categorical. It follows that the ability to discriminate between racial categories is advantaged, while within-race differences become more difficult to discern (Levin & Angelone, 2002). The differential attention
Given to between-category difference has clear implications for stereotyping, where differences within category members may be ignored, and generalizations may be broadly applied to any individual belonging to a category regardless of whether they are correct. In this way, categorical thinking can facilitate the stereotyping of individuals (Lepore & Brown, 1997).

Given the perceptual advantage that people have for distinguishing targets from different categories, it probably is not surprising that the study of stereotyping has long focused on between-group stereotypes. However, the between-group distinction has been somewhat limited, as people are also sensitive to how well the target fits into a category (Locke, Macrae, & Eaton, 2005). Within-category differences have an effect on how much a target is stereotyped. For example, recent research finds that individual differences in facial features and skin tone contribute to stereotyping over and above racial category membership alone—a phenomenon referred to as feature-based stereotyping (FBS; Blair, 2006; Maddox, 2004). Despite the attention directed toward studying stereotyping between social groups, relatively little is known about how FBS operates on the level of mental processes. The goal of the current study is to explore possible mechanisms underlying FBS.

The Role of Categorization in Stereotyping

Social categorization is an important part of the stereotyping process, as category activation increases stereotype accessibility. For example, research conducted by Lepore and Brown (1997) has shown that participants primed with category labels are more likely to stereotype a target than participants primed with nonwords. Additionally,
individual differences in prejudice moderate the valence of the stereotypes that are activated, where more negative stereotypes are associated with higher prejudice, relative to lower. However, being primed by category labels lead to increased stereotyping of targets, regardless of whether participants personally endorsed stereotypic content. These findings suggest that stereotypes can be automatically activated by social categorization alone (see also Arkes & Tetlock, 2004; Devine, 1989; Smith & DeCoster, 2000). For this reason, social categorization and the ease with which it operates is vital to the understanding of stereotype activation.

The categorization of individuals into social groups is both a robust and spontaneous process. People use categorization strategies without being explicitly asked (Stangor, Lynch, Duan, & Glass, 1992; Taylor, Fiske, Etcoff, & Ruderman, 1978) even when attending to categories disrupts performance (Macrae, Quinn, Mason, & Quadflieg, 2005). Evidence for this has been observed in the “who-said-what” task. In this task, participants are presented with faces of varying social group (e.g., Black or White faces) along with various statements that are attributed to the individual targets. Participants are told that this is an impression formation task and no special attention is drawn to the social group membership of the targets. Later, participants are given a surprise recognition task in which they are asked to match statements to the faces which they were previously presented. Participants are more likely to misremember information about targets that belong to the same social group (e.g., misremembering a statement made by a Black man with a statement made by different Black man), than across social groups (e.g., misremembering a statement made by a Black man with a statement made by a White man). The pattern of matching errors suggests that participants spontaneously
attend to social categories to organize information about targets—even when the task does not require social categorization (Stangor et al., 1992; Taylor et al., 1978).

Although people process faces generally in a holistic manner (Farah, Wilson, Drain & Tanaka, 1998; Tanaka & Bukach, 2004), when individuals categorize targets, they will use a relatively more piecemeal processing strategy. When the processing goal is to categorize an individual, “featural processing” is used, through which particular features are attended to discern a target’s category membership. Notably, very little perceptual information and cognitive resources are required to determine category membership in comparison to discerning the identity of a face (Cloutier & Macrae, 2007; Schyns, Bonnar, & Gosselin, 2002), making social categorization a highly efficient process. Social categorization based on visual inputs is so robust that degradation of the visual stimuli (e.g., face inversion, blurring, Martin & Macrae, 2007), has significantly less of a disruptive impact on social categorization relative to individuation (Cloutier, Mason, & Macrae, 2005). For example, other studies have found that participants are able to identify the gender of a face by seeing the left eye and corners of the mouth when the rest of the face is blurred (Schyns et al., 2002). Converging on the notion of features being strongly related to categorization, certain features appear to be attended to and weighted more heavily, such as hair as a cue for gender (Cloutier & Macrae, 2007). It follows that when these features are removed, participants are significantly slower to categorize targets as male or female. Additionally, EEG studies have found that categorization of social groups occurs at the speed of a few hundred milliseconds, during which there is evidence of participants attending stimulus features, providing additional support for how specific features are implicated in social categorization (Ito & Urland,
2003). Taken together, people attend category relevant facial features quickly, and use these features to categorize targets even with minimal visual information and in other perceptually challenging situations.

The ease with which social categories are activated is also meaningful because of the creation of automatic stereotypic associations. When a target (such as a Black face) is presented frequently enough with certain information (such as stereotypes of danger), the association strengthens and activates automatically. Once an automatic association is formed, the mere presence of the target can elicit the associated stereotypes (Devine, 1989). A person does not need to endorse the stereotype in order to be affected by it; merely having knowledge of a stereotype can lead to its activation (Arkes & Tetlock, 2004; Devine, 1989; Smith & DeCoster, 2000). The effects of stereotype activation can manifest in varied ways such as through implicit tasks, for example in how people orient and respond to the presentation of faces. Black stereotypes are often danger and crime-related, sensitizing people to the presence of Black faces, where people are found to more quickly orient to Black faces, relative to White faces (Donders, Correll, & Wittenbrink, 2008; Eberhardt, Goff, Purdie, & Davies, 2004). Black faces are also more strongly associated with crime-related objects, such as guns, compared to White faces, a finding of the Weapon Identification task (WIT; Payne, 2001). In the WIT, participants are primed with either Black or White faces, followed by an object that can be either a gun or a tool. Errors on the WIT reveal a pattern of racial bias. Typically, participants misidentify tools as guns more often following a Black prime, than when shown a White prime. Conversely, when primed with a white face, participants misidentify guns as tools more often than when shown a Black prime (Payne, 2001). Given the ease at which categories
are activated by the presentation of a target, the relationship between categorization and stereotyping is an important one, as it explains the automaticity of stereotype activation.

In sum, social categorization takes place spontaneously, and can manifest with minimal perceptual information. Social categorization is highly relevant to the studying of stereotyping due to the strength of the relationship between categories and stereotypes. Categorization is an automatic process, a quality which generalizes to stereotyping when targets are paired with stereotypic information frequently enough. Understanding the advantage that people have for between-category distinctions and how it has applied to stereotyping has been highly useful to study. However, within these categories, people are also sensitive to where targets fall on the average “fit”, which leads to a more fine-grained form of stereotyping, FBS.

**Feature-Based Stereotyping**

The psychological ease with which individuals attend to categories has led researchers to focus primarily on between-group differences when studying stereotyping. Researchers have historically studied stereotyping effects through the comparison of the average difference in response to Black and White targets (e.g., average reaction time to stereotypic objects after a Black or White prime). However, focusing on between-group differences actively ignores the possible influence of important, systematic within-group differences. Specifically, category activation may be mediated by the features that a target has, where more typical features activate a category more strongly based on its featural similarity to a “prototype.” This sensitivity to within-group differences is evident in a phenomenon referred to as FBS. In general, the more “prototypic” a target’s features
are of their social group, the more that a target will be stereotyped. Black prototypic (i.e. darker skin, broad nose, and fuller lips; Blair, & Judd, 2010) features are often operationalized in order to study FBS; the more of these features that a target has, the more closely the target will be associated with Black stereotypes. For example, when participants are asked to list cultural beliefs about Black men of varying skin tone (one of the factors in Black prototypicality); there is an interaction between skin tone and stereotypicality. Beliefs listed about Black men with darker skin tone are more stereotypic and more negative than beliefs listed about Black men with lighter skin tone (Maddox & Gray, 2002).

These associations also extend to implicit responses. Black faces with darker skin and highly prototypic features have been found to elicit more automatic negative reactions on priming tasks (Hagiwara, Kashy, & Cesario, 2012). Participants were primed with Black faces that varied independently in skin tone and prototypicality, after which they were asked to identify words that varied in valence (negative/positive) as good or bad. Responses to negative words were faster both after darker skinned Black faces, and after more prototypic Black faces, suggesting that there is an independent but additive effect of features and skin tone (Hagiwara, Kashy, & Cesario, 2012; Livingston & Brewer, 2002). Negative responses to highly prototypic targets also influence judgments on implicit tasks, such as decisions to “shoot” or “not shoot” a target. Using the “First Person Shooter” task (Correll, Park, Judd, & Wittenbrink, 2002), Ma and Correll (2011) found that target prototypicality moderated the decision to shoot. Specifically, participants fail to shoot more armed Whites relative to armed Blacks when White prototypicality is high. These findings suggest implicit bias in shooting decisions operate
on a continuum, where participants are more likely to decide to shoot targets that are high in Black prototypicality, and less likely to shoot targets that are high in White prototypicality.

Stereotyping by features also takes place over and above racial categories. Controlling for criminal history and severity of crime, both Black and White inmates with more prototypic features are given harsher criminal sentences than targets who have less prototypic features (Blair, Judd & Chapleau, 2004). Eberhardt, Davies, Purdie-Vaughns and Johnson (2006) report similar findings; they found that in cases where the victims were White, Black defendants with more prototypic features were more likely to get the death sentence relative to Blacks with less prototypic features. Notably, judges have been successfully trained to avoid racial bias in sentencing decisions; that is, court records in both studies showed no racial bias in sentencing, suggesting that practice with suppressing category-based stereotyping (CBS) does not generalize to FBS, and that prototypic features were independently driving the observed stereotyping effect.

One of the other ways FBS seems to differ from CBS is in how it operates under cognitive load and explicit suppression, as observed in impression formation tasks. In these tasks, participants are presented with written descriptions of individuals who vary in Black prototypicality (stereotypic or counter-stereotypic) and in valence (negative or positive). Facial photographs that vary in prototypicality (high or low) are then presented, and participants are asked to estimate the probability that the previous information they were given describes the face. Stereotyping is defined as assigning higher probabilities of Black stereotypic content to faces, than counter-stereotypic content. During this task, some of the participants are given a cognitive load task, while the other group is
instructed not to use race in their decisions (category suppression condition). Blair and colleagues (2004) find that participants are able to avoid CBS when instructed to avoid using race, but are unable to control their use of FBS. Cognitive load also had an interactive effect with type of stereotyping. Participants increased their use of CBS, but their use of FBS remained the same. That is, regardless of instructions to avoid stereotyping by race, or under cognitive load, participants assigned higher probabilities of stereotypic content to faces with more Black prototypic features, compared to faces with less Black prototypic features. In a follow-up study, an additional feature-suppression condition was added; but was similarly ineffective in reducing FBS (Blair, Judd, & Fallman, 2004). The differential response of FBS and CBS to cognitive load and conscious suppression, suggest that FBS operates differently from CBS.

Disrupted processing also has unique effects on FBS relative to CBS. FBS persists through face inversion, a commonly employed method of disrupting face processing (Rossion & Gauthier, 2002). Using the same impression formation task used in Blair, and colleagues (2004), participants were presented with Black faces that varied in spatial orientation (inverted or upright) and prototypicality (Blair, 2006). Because of the increased face processing difficulty, inverted faces had the effect of reducing participant self-presentation concerns about CBS. Specifically, when participants made judgments about inverted faces, they increased their use of racial stereotypes regardless of valence, relative to upright faces which indicated only negatively valenced stereotyping. In contrast, use of FBS remained constant across face orientation, with faces high in Black prototypicality being assigned higher stereotypic probabilities than faces low in Black prototypicality. Given that face orientation reduced self-presentation
concerns for CBS, these results suggest that participants were unaware of using prototypic features in their judgments and thus face orientation had no effect. This study provides converging evidence that in contrast to CBS, which was again increased by restricted access to cognitive resources, FBS appears to operate by a different pathway than CBS. However, what specifically differs between these two forms of stereotyping remains unclear.

**Mechanisms Underlying Feature-Based Stereotyping**

**Direct feature-trait association.** One question that the divergence between CBS and FBS raises is whether the mental processes underlying these two types of stereotyping differs. Clearly the two types of stereotyping differ depending on the task at hand, but do they operate differently at a cognitive level? There are two major accounts of the process behind FBS (Blair, Judd, Sadler, & Jenkins, 2002; Maddox, 2004). One explanation is “direct feature-trait association” (DFTA): wherein certain physical features become directly associated with traits due to associative learning. According to this theory, “if certain physical features are frequently and consistently paired with certain traits- through their shared association to the category- then those features and traits may become directly associated with one another through associative learning” (Blair et al, 2002; Hebb, 1948). For example, Black prototypic features are often paired with Black targets, which activate the category “Black person”, as well as the stereotypes (see Figure 1). This theory posits that if the process takes place frequently and consistently enough, Black prototypic features will be able to activate the stereotypes independent of whether the target is actually Black. Through this framework, DFTA asserts that prototypic features directly activate group stereotypes, bypassing categorization altogether (Blair et
Supporting the notion that FBS can take place independently of categories, FBS operates without regard to racial category membership. Investigations of court records reveal that while there appears to be no racial bias in sentencing decisions, there is evidence of FBS. Specifically, White and Black defendants are given equally harsh sentences, but both White and Black defendants with more Black prototypic features have an increased likelihood to be given longer sentences, and are also more likely to receive the death sentence than defendants with less Black prototypic features (Blair et al., 2004; Eberhardt et al., 2006). These findings suggest that FBS can take place above apparent racial category membership, where even on a White face, Black prototypic features are strongly associated with stereotypes of danger.

Through the comparison of CBS and FBS (at least based on the extant data), there appears to be a difference in how the two operate. For example, participants increase their use of CBS under cognitive load, but use of FBS remains constant regardless of access to cognitive resources. This discrepancy suggests that while participants are aware of employing CBS and can consciously suppress its use, they are unable to do the same with FBS. Another difference between CBS and FBS is in its controllability. When explicitly asked to do so, participants reduce their use of CBS; however, participants cannot control their use of FBS. Similarly, expertise in avoiding CBS does not generalize to the controllability of FBS (Ma & Correll, 2011). On the first person shooter task, participants who were trained and capable of avoiding racial bias in decisions to shoot were unable to avoid FBS. Specifically, participants who did not exhibit a pattern of racial bias in shooting decisions, would fail to shoot more armed White targets with highly prototypic
features, than White targets with less prototypic features. The differential response of CBS and FBS to cognitive load, explicit instructions, and the lack of generalizable expertise in CBS suppression suggest that FBS operates somehow differently from CBS. While both forms of stereotyping are efficient and uncontrolled processes, FBS may use a different cognitive mechanism.

Although FBS is convincingly distinct from CBS in the aforementioned studies, they only offer indirect support for DFTA. In regards to FBS operating over and above racial categories, it is unknown how the people making the sentencing decisions perceived the defendants (Blair et al., 2004; Eberhardt et al., 2006). Specifically, it is impossible to know whether the targets were categorized as White, and not some other non-White race due to their features, confounding the FBS effect with stereotypes of other non-White racial groups. The impression formation studies are clearer in support of FBS operating differently from CBS, but are all reliant on impaired processing leading to a reduction of self-presentation concerns to account for these differences. Specifically, that cognitive load and face inversion increased CBS, but had no effect on FBS was explained as participants being less aware and able to control biased responses. In addition, CBS is well-known and socially undesirable, which leads participants to exert effort to reduce their overt stereotyping. FBS in contrast is virtually unheard of and carries no social stigma. Due to the social unacceptability of racial stereotyping, people become practiced at conscious suppression, sometimes to the point of influencing their implicit use of it (Moskowitz, Gollwitzer, Wasel, & Schall, 1999). Thus, a practice confound is present between the two types of stereotyping, where even when participants are informed of FBS, it is unlikely that they could reduce their use of it on an equal level.
to CBS, which they have consciously subdued for years. While racial stereotyping and FBS seem to function differently, the evidence is indirect in its support of actual processes, leaving space for alternative interpretations.

**Graded categories.** An alternative explanation for how FBS unfolds comes from the current understanding of how categories are structured. To make organization of information coherent, categories rely on categorical perception to minimize within-group difference, while maximizing between-group difference (Harnad, 1987). Such a grouping strategy works toward forming categories at the most general and abstract level where people are able to obtain the greatest amount of information from a category with the least amount of effort (Medin, 1989).

In order to increase the flexibility and distinctiveness of categories, they are defined in terms of a “prototype” or by prototypical examples. Prototypes possess the features most representative of the category, and possess the least amount of features that are representative of things outside of it, meeting the goal of minimizing within-group difference while maximizing between-group difference (Rosch, 2002). To this effect, when Rosh and Mervis (1975) trained participants to categorize nonwords, they found objects with the least overlap in features with objects in a different category were seen as the most prototypic, and objects with the most overlap with other groups were rated the least prototypic. In addition, participants learned the category membership of the most prototypic objects more quickly than less prototypic objects, and responded more quickly to prototypic objects when asked to sort them into categories. Similarly, children learn the category membership of prototypes earlier than other members of a category (Rosch, 1973). Thus, objects that have more prototypic features facilitate category activation more
easily, and are more likely to be automatically associated with such categories than objects that have less prototypic features.

With categories formed around prototypes, there are both good and bad examples of a category based on their similarity to the prototype (Medin, 1989). Because of this, categories have a “graded” structure, where a concentration of certain features can activate a category more strongly based on its featural similarity to the prototype. Supporting this concept, targets that possess a high concentration of prototypic features are learned more easily, categorized more quickly, and are rated more prototypic (Rosch & Mervis, 1975). Likewise, research using brain imaging techniques supports the notion of graded categories. For example, more prototypic category exemplars of faces activate the fusiform face area (an area of the brain that has been implicated in processing faces) more strongly than less prototypic exemplars (Mur et al., 2012).

Taken together, an alternative interpretation of the mechanisms behind FBS emerges. The tendency for faces that are more racially prototypic to be rated more stereotypically can be accounted for by the differential activation of categories based on how representative a target is of its category (Medin, 1989). That is, highly prototypic faces activate racial categories more strongly, and are thus more strongly associated with stereotypes for that group. In addition, prototypic features come from the average of features most common to exemplars as well as the most commonly encountered features, making them associated with categories with high frequency, facilitating automatic activation (Smith & Collins, 2009). The characteristics of categorization may also apply to the automaticity of FBS (Blair et al., 2004), in that the more prototypic a target is, the more quickly it will be identified as a category member (Rosch et al., 1976). Adding to
that, categorization is itself an efficient process (Cloutier & Macrae, 2007; Schyns et al., 2002), and the methods used in previous studies do not prevent categories from being activated when FBS is being employed. Thus, the graded structure of categories could lead to a stronger, quicker, and more efficient activation of stereotypes when a highly prototypic target is present (see Figure 1).

While FBS appears to operate distinctly and differently from CBS, whether it takes place independently from categories is ambiguous. The methods that have been previously employed do not prevent stereotypes from being activated through the use of categories. Specifically, prototypes may be too closely tied to categories for the use of typical category suppressions to be effective. To address this, we intend to disrupt categorical processing itself in order to see if FBS takes place.

**Disrupting Categorical Processing**

To more rigorously test the competing explanations for FBS, the current study seeks to inhibit the use of categorization by manipulating categorical perception. There are many methods that can be used to manipulate categorical perception, such as presenting targets differentially to brain hemispheres (Hellige, 1993; Roberson, Pak, & Hanley, 2008) or by minimizing the value of categories through context (Macrae & Cloutier, 2009). However, it would be difficult for participants to perform an implicit task that was presented to different hemispheres, and a context manipulation would likely influence FBS expression in unintended ways. As such, the approach we use in the present study is a form of verbal interference.

While the phenomenon of categorical perception is robust, previous research has demonstrated that verbal interference can disrupt categorical perception. Roberson and
Davidoff (2000) employed a categorical perception task on color, in which participants were shown a color, followed by two other colors, and were asked to indicate which color matched the color presented previously. Some of the participants completed this task under visual interference (trace a line through a dot pattern with their eyes) while others performed the task under verbal interference (reading both color-related and non-related words aloud). In the visual interference condition, basic categorical perception effects were found; participants were better at discriminating between colors from different categories than colors that came from the same category. However, in the verbal interference condition, the between-category advantage was eliminated – participants performed approximately equal in their discriminating between colors that came from the same category or different one. In a follow-up study, the same effects were demonstrated using faces, where verbal interference disrupted the between-category advantage for discriminating between facial expressions, and spatial interference had no effect (Roberson & Davidoff, 2000; Wigget & Davies, 2008; Pilling, Wigget, Ozgen, & Davies, 2003).

The difference in effect between verbal and spatial interference on categorical perception is attributed to the use of verbal coding. This explanation is further supported by hemisphere lateralized studies, which find that when stimuli is presented first to the left hemisphere, participants categorize targets that cross categories more quickly than targets from within the same category (Ma, Khetarpal, Davis, & Correll, 2013; Roberson et al., 2007; Sanders, McClure, & Zarate, 2004). In contrast, this effect is much weaker for stimuli presented to the right hemisphere. However, when verbal interference is introduced, this between-group advantage is once again eliminated for the right
hemisphere, whereas spatial interference does not suppress categorical perception (Gilbert, Regier, Kay, & Ivry, 2007). The left hemisphere advantage for categorical perception is explained as being due to its specialization in language processing, which speeds categorical perception through category labels. Verbal interference would prevent a labeling strategy, which interferes with the verbal encoding of information (Pilling et al., 2003; Wigget & Davies, 2008). It bears mentioning that there are other accounts of how categories may be organized or come to be, such as embodied cognition.

**Present Research**

The current study employed a verbal interference task in order to test whether categorization mediates FBS. To manipulate categorical perception we used a version of the two-back task developed by Gray (2001), to introduce verbal and spatial interference. This two-back task required participants to attend to verbal or spatial information. In the verbal version of the task, participants were presented with a letter and told to remember its identity, whereas in the spatial version of the task, participants were told to remember the letter’s location. Critically, the two versions of the task disrupt distinct cognitive processes. The verbal version of the task created verbal interference, disrupting categorization, while the spatial version created spatial interference. We chose this task because it can be used to impair categorization, in addition to the fact that the spatial version of the task can be used to provide a comparable control task, shown to be equal in terms of difficulty (e.g., Beilock, Rydell, & McConnell, 2007). In order to measure FBS in this experiment, we used a version of the WIT (Payne, 2001), selected for its lack of reliance on words in order to maintain the approximately equal difficulty between the verbal and spatial tasks. We also used phones rather than tools due to an advantage of
perceptual similarity between phones and guns (Payne, 2001).

While typically participants are primed with either Black or White faces, we added the variable of prototypicality. After being presented with a face with more Black prototypic features, participants should be more likely to identify a phone as a gun than when primed with a face with less prototypic Black features. Conversely, participants should be more likely to identify a gun as a phone when primed with a face with more White prototypic features, relative to following a less White prototypic prime.

The existence of FBS is well supported (Blair, 2006; Blair, Judd, & Chapleau, 2004; Blair et al., 2005; Blair et al., 2004; Blair et al., 2002; Eberhardt et al., 2006; Ma & Correl, 2011; Maddox & Grey, 2002), but the mechanisms behind it remain unclear. Both DFTA (Blair et al., 2002) and graded categories (Lepore & Brown, 1997; Macrae et al., 2005) are viable explanations for the process. If DFTA explains FBS, we predict that verbal interference will have no effect and that the errors will indicate a pattern of FBS, in which trials with highly prototypic primes will have the highest rate of object identification error (phones misidentified as guns, and vice versa) relative to trials with low prototypic primes Spatial interference trials should also show FBS effects, as spatial interference should have had no effect on categorical perception. If categorization explains FBS, we predict that trials with verbal interference will have equal amounts of object identification error across every level of prototypicality, indicating no FBS.
METHOD

Participants

Participants were 92 White, right-handed undergraduates (33 men, 59 women; \(M_{age} = 19.49\)) from the Psychology Department study pool at California State University Northridge. All participants were randomly assigned to one of three conditions, control \((n = 31)\), verbal interference \((n = 33)\), or spatial interference \((n = 28)\). However, data for 4 participants were removed due to below chance accuracy on the WIT (<50%), 3 participants were from the control group \((n = 28)\) and 1 from the verbal condition \((n = 32)\).

Design

The study combined the WIT (Payne, 2001), and interference tasks (Gray, 2001) in a 3 (interference type: verbal/spatial/control) \(\times\) 2 (prototypicality: high/low) \(\times\) 2 (race: Black/White) \(\times\) 2 (object: gun/phone) mixed-group design with interference type varying between-subjects and all other factors within-subjects.

Stimuli

The faces used in both tasks were images of 6 Black and 6 White faces from the Chicago Face Database (Ma et al., 2012) which were paired to create 6 mixed-race pairs. For each pair of original faces (unmorphed), we created a continuum of mixed-race faces that range from 0% Black / 100% White to 100% Black / 0% White using FantaMorph software. We then selected 4 faces from each continuum at 30% intervals: 5% Black / 95% White, 35% Black / 65% White, 65% Black / 35% White, and 95% Black / 5% White (see Figure 3). This selection allowed us a more objective measure of racial
prototypicality while equating for physical difference across within and between-category pairs. Additionally, these faces were pretested to ensure that they were subjectively perceived as intended, where 5% and 35% faces were perceived as Black, and 65% and 95% faces were perceived as White (Ma, Khetarpal, Davis, & Correll, unpublished manuscript; see Figure 4). Faces were also equated for contrast and luminance. These faces were used in both the manipulation check for the categorical perception judgments and in the experimental task as primes for the WIT.

The stimuli used for the interference task was adapted from Gray (2001), made up of 10 letters (b, c, d, f, g, h, j, k, l, m) that appear in black against a white background. Multiple copies of the letter were shown in 48-point Geneva font inside a 5.4-cm box that appeared in one of six possible locations around the center of the screen. Each presentation of the box with letters was shown against a mask of multiple letters.

**Questionnaire.** Upon completion of the task, participants completed a questionnaire concerning demographics and the modern racism scale (MRS; McConahay, 1986). Participants received partial course credit and were fully debriefed upon completion of the study.

**Procedure**

The experimental task was performed on PC computers using the reaction time software E-prime. In the verbal interference version of the task, participants were presented with a box of (target) letters over a letter mask for 1000 ms., followed by an image of a Black or White male that appeared for 200 ms, followed directly by either a phone or gun. After the object presentation a back mask appeared and participants were
given 500 ms to indicate which object they saw. If they answered correctly participants continued the task and saw a second box of (test) letters, to which they indicated if the letter did not match the target letter within 2500 ms. (see Figure 2). If participants answered incorrectly on the WIT, the trial would reset and participants would repeat the process with a new randomly selected letter. In the spatial interference version the procedure was identical with the exception that participants were asked to indicate when the location of the test letter did not match the location of the target letter. The control condition operated identically to the spatial and verbal versions, but participants were asked to do nothing during letter presentations. There were 27 practice trials total, 7 for the interference task, 10 for the WIT, and 10 for the combination of both the interference and WIT tasks. After finishing the practice trials, there were 192 test trials where the interference and WIT were combined. In the test trials each of the 6 faces were presented 32 times, 8 times on each of the four levels of racial prototypicality. In the interference task, approximately 30% of the interference trials presented matching letters, while 70% of the trials were nonmatching.
RESULTS

Signal Detection Analysis

In order to better understand the effect that prototypicality and interference may have on racial bias, we employed signal detection theory on the data (Green & Swets, 1966). SDT estimates 2 orthogonal parameters based on errors. In particular, SDT uses misses (i.e., errors made identifying guns as phones) and false alarms (i.e., errors identifying phones as guns) to estimate $d'$ (sensitivity) and $c$ (criterion). $d'$ is calculated by subtracting hits from false alarms for each trial type, producing scores that represent the ability to discriminate guns from tools. In this case $d'$ is calculated for both race and prototypicality, giving four estimates of $d'$ (i.e., high $d'$ for black = hits on high Black trials – false alarms on high Black trials). Given that accuracy on gun trials on the WIT tend to be fairly high (Payne, 2001), we do not hypothesize any meaningful differences on the $d'$ parameter, making our primary focus $c$. $c$ is calculated by hits (i.e., correct identification of a gun) + false alarms divided by -2, giving an overall average of response tendency for a trial type. In the context of the current study, there are four estimates of $c$ for each combination of race and prototypicality (i.e., high $c$ for Whites = hits on high White trials + false alarms on high White trials / -2). Conceptually, $c$ estimates the hypothetical point at which individuals favor a “gun” response compared to a “phone” response. To the extent that an individual has a prepotent tendency favoring a “gun” response following a Black prime, $c$ should be fairly low and negative. In contrast, $c$ should be comparatively high and positive on White prime trials where the prepotent tendency does not favor a “gun” response. Critically, the difference between $c$ for Black trials versus White trials represents racial bias on the WIT where a lower $c$ for Blacks
compared to Whites indicates greater racial bias. In the current context, \( c \) may also differ between prototypicality, where highly prototypic Whites compared to highly prototypic Blacks may elicit a higher positive \( c \) than will low prototypic Whites.

In order to assess the effect of prototypicality on \( d' \) and \( c \), separate measures were computed for each level of prototypicality. Given that signal detection examines only gun trials, and is calculated from differences for each race, the race and object variables are computed into the model. As such, we employed a 2 (prototypicality: high/low) × 3 (interference type: control/verbal/spatial) mixed-ANOVA on \( d' \) and \( c \). \( d' \) for high trials is calculated from high Black \( d' \) – high White \( d' \), while \( d' \) for low trials = low Black \( d' \) – low White \( d' \). Similarly, \( c \) for high targets is calculated from high Black \( c \) – high White \( c \), and \( c \) for low trials = low Black \( c \) – low White \( c \). In sum, the \( d' \) and \( c \) estimates represent a difference score for each level of prototypicality, without regard to race.

c. To test the hypothesis that interference would have differential effects on the expression of FBS, we conducted a 2 (prototypicality: high/low) × 3 (interference type: control/verbal/spatial) mixed-ANOVA on \( c \) scores. There was no evidence for a main effect of interference type \( F(2, 84) = 0.73, p = .49 \). There was a trending main effect of prototypicality on criterion \( F(2, 84) = 3.40, p = .07 \), where there was a higher criterion for highly prototypic targets (\( M = .06, SD = .35 \)) than low prototypic targets (\( M = -.02, SD = .30 \)), but no prototypicality × interference type interaction \( F(2, 84) = 0.77, p = .47 \). These findings suggest that there was a significant difference in criterion between high and low prototypic targets. However, given that the scores for high prototypic targets come from the difference between high White targets and high Black targets, and low comes from
the difference between low White and Black targets, it is difficult to determine what the prototypicality main effect alone means for racial bias.

In order to test for FBS effects that vary within race and that this effect may vary with interference, we conducted a 3 (interference type: control/verbal/spatial) × 2 (prototypicality: high/low) × 2 (race: Black/White) mixed ANOVA on criterion scores. This analysis found no main effect of prototypicality $F(1, 84) = 1.80, p = .18$, or interference × prototypicality interaction $F(2, 84) = 2.23, p = .11$. There was no main effect of race $F(1, 84) = 1.84, p = .18$, or race × interference interaction $F(2, 84) = 1.18, p = .31$. However, there was a trend toward a prototypicality × race interaction $F(1, 84) = 3.61, p = .06$ (see Figure 5), and no evidence for an interference × prototypicality × race interaction $F(2, 84) = 0.77, p = .47$. Follow-up analyses indicated that this interaction was driven by a main effect of prototypicality on White trials $F(1, 87) = 4.09, p = .05$ where there was higher criterion for highly prototypic White targets ($M = .02, SD = .26$) than low prototypic White targets ($M = -.05, SD = .26$). In contrast, there was no significant difference in criterion between high and low Black prototypic targets $F(1, 86) = 0.30, p = .59$. These findings all suggest that the main effect of prototypicality was driven by a difference in criterion between high and low White prototypic targets, while participants showed no differential criterion for Black targets based on their prototypicality.

$d'$. While $d'$ is a measure of how well participants are able to differentiate between phones and guns and is unlikely to be sensitive to the effects of interest, it was important to investigate any differences in accuracy across conditions. In order to test for differences in participant ability to discriminate between phones and guns as a function of the prototypicality of a prime or interference type we conducted a 2 (prototypicality:
high/low) × 3 (interference type: control/verbal/spatial) mixed-ANOVA. There was no
main effect of interference type \(F(2, 84) = 0.40, p = .67\), main effect of prototypicality
\(F(1, 84) = 0.29, p = .59\), or interference type × prototypicality interaction \(F(2, 84) = 1.82,
p = .17\). Overall suggesting that participant ability to discriminate between guns and tools
did not vary as a function of prototypicality or interference.

**Multilevel modeling**

**Errors.** Because prototypicality was nested within race, we employed multilevel
modeling to look at the effect of prototypicality on error rates. Multilevel models
compute the average slope for each participant on every variable. In this case, we
computed the slopes for each participant on interference type, prototypicality, race, object
and all interactions for error rate. The average of these slopes is then subjected to one-
sample t-tests that compare the means to zero. We first ran a basic race × object model in
order to test for the hypothesis that participants would show racial bias (Payne, 2001). As
such, we conducted a t-test on the slopes of race and object, resulting in a 2 (race:
Black/White) × 2 (object: gun/phone) analysis, looking at the effect of object and race on
error rates. We failed to support for a race × object interaction \(t(87) = 1.22, p = .23\), or a
race main effect \(t(87) = -0.55, p = .58\). There was however a theoretically uninteresting
trend toward an object main effect \(t(87) = 1.95, p = .05\), whereas in prior work we found
worse accuracy for phone trials \((M = .07, SD = .08)\) than for gun trials \((M = .05, SD = .06)\).

In order to test the hypothesis that FBS would vary among levels of interference
in error rates, we ran a model that consisted of 3 (interference type: control/verbal/spatial)
× 2 (prototypicality: high/low) × 2 (race: Black/White) × 2 (object: gun/phone). There was no evidence for a prototypicality × race × object interaction, *t*(87) = 1.23, *p* = .22, nor was there a race × object interaction *t*(87) = -0.44, *p* = .66 or a prototypicality × race interaction *t*(87) = -0.22, *p* = .66. The interference type × prototypicality × race × object interaction also failed to reach significance *t*(87) = 1.00, *p* = .32. There was no main effect of interference *t*(87) = 0.74, *p* = .46, prototypicality *t*(87) = -0.95, *p* = .35 or race *t*(87) = -0.74, *p* = .55. As in the race × object model, there was a theoretically uninteresting main effect of object *t*(87) = 2.00, *p* = .05, replicating prior work where participants were less accurate on phone (*M* = .07, *SD* = .08) trials than on gun trials (*M* = .05, *SD* = .06).

**Reaction time.** While the form of the WIT used in this study had a short response window and was designed to force participants to make errors rather than allow differences in reaction time in response to primes, reaction times could still vary across trial type. Reaction times were trimmed by removing trials that were greater or less than 2.5 standard deviations from a participants’ average response time (approximately 2.6% of trials). In addition, trials where participants failed to respond within the 500ms window were removed (approximately 9.2% of trials). The remaining reaction times were then log transformed.

We first looked at the race × object interaction to determine if we replicated the basic WIT model of racial bias. However, there was no evidence for a race × object interaction *t*(87) = -0.34, *p* = .74, main effect of race *t*(87) = 0.29, *p* = .77, or object *t*(87) = 1.0, *p* = .32. Next we tested for the existence of FBS, using a 2 (prototypicality: high/low) × 2 (race: Black/White) × 2 (object: gun/phone) model on reaction times.
There was a significant race × prototypicality × object interaction \( t(87) = 2.00, p = .05 \) (see Figure 6). To probe the 3-way interaction we looked at the interaction between race and object for high and low prototypic groups separately, breaking the analysis down into two separate 2 (race: Black/White) × 2 (object: gun/phone) models for high and low prototypic trials. In the highly prototypic group, there was no main effect of race \( t(87) = -1.34, p = .18 \), or object \( t(87) = 0.11, p = .91 \). In addition, the race × object interaction was not significant \( t(87) = -0.79, p = .43 \). In the low prototypic group there was no main effect of race \( t(87) = 0.95, p = .34 \), or object \( t(87) = 1.42, p = .16 \). However, there was a trending significant race × object interaction \( t(87) = 1.94, p = .06 \). To understand the interaction we compared the slopes for low Black and White gun trials, and found a trend towards a significant simple effect of race \( t(87) = 1.98, p = .05 \) on gun where participants were quicker to respond to gun trials after a low White prime \((M = 5.41, SD = .26)\) than after a low Black prime \((M = 5.44, SD = .30)\). In contrast, race had no effect on reaction time for phone trials \( t(87) = -0.67, p = .51 \). We also investigated the simple effect of object on low prototypic race by comparing slopes for gun and phone trials. For low White trials there was a significant simple effect of object \( t(87) = 2.47, p = .02 \), where participants were quicker to respond to gun trials \((M = 5.41, SD = .26)\) than to phone trials \((M = 5.45, SD = .23)\), while on low Black trials there was no difference in reaction time between gun and phone trials \( t(87) = -0.10, p = .92 \).

Finally, in order to test the hypothesis that FBS would vary among levels of interference, we employed a 3 (interference type: control/verbal/spatial) × 2 (prototypicality: high/low) × 2 (race: Black/White) × 2 (object: gun/phone) multilevel model looking at reaction time. There was no main effect of interference type \( t(87) = -
1.00, \(p = .32\), prototypicality \(t(87) = -1.29, \ p = .19\), race \(t(87) = 0.11, \ p = .91\), or object \(t(87) = -0.93, \ p = .36\). There was also no interaction between race and object \(t(87) = -0.42, \ p = .68\), race \(\times\) prototypicality \(t(87) = -1.73, \ p = .09\), or object \(\times\) prototypicality \(t(87) = -0.95, \ p = .35\). Finally, there was no 4-way interaction between interference type, prototypicality, race, and object \(t(87) = 0.95, \ p = .33\).
DISCUSSION

Overall, the signal detection analysis found evidence for racial bias on criterion. Specifically, participants showed a higher criterion for high White targets relative to low White targets, suggesting that participants favored a “gun” response after low White primes to a greater extent than after high White primes. However, there was no main effect of prototypicality for Black targets, suggesting that participants did not differentiate between high and low Black targets. We did not hypothesize an exclusive effect of prototypicality on White targets; however these results are not entirely unexpected given that prior literature has found evidence for greater sensitivity to prototypicality in White targets (Ma & Correll, 2011; Ronquillo et al. 2007).

In the multilevel model we found no theoretically interesting effects on error rate; however we discovered an unanticipated significant prototypicality × race × object interaction on reaction time. Follow-up analyses indicated that this was driven by a difference between low White and Black prototypic targets, where low White primes facilitated a gun response to a greater extent than low Black primes. Similar to the SDT results, these findings suggest that low White targets may have been seen as the most threatening, and as such activated danger stereotypes to a greater extent than other targets.
METHOD

Because one of our predictions relies on null effects, and because of our novel use of the interference task, we employed a manipulation check of our interference task. In order to clarify the effect that the interference task had on category accessibility, we combined it with a categorical perception task in which participants discriminated between faces. In this task, participants were presented with two faces simultaneously that varied in similarity (race and prototypicality), which was followed by a back mask on which participants were asked to indicate whether the faces were the same or different. The back mask was used to erase any afterimages of the faces to ensure that their same/different response was based on the 450 ms presentation time. Typically, the categorical perception tasks find that participants are more accurate at judging when faces are from different categories (i.e., discriminating between Black and a White face) than from the same category (i.e., discriminating between two Black faces or two White faces). Using this task in combination with interference, we anticipated a conceptual replication of the differential effect of interference type found in previous studies employing spatial and verbal interference in combination with categorical perception tasks (Gilbert et al., 2007; Roberson et al., 2008; Pilling et al., 2003; Wigget & Davies, 2008). We predicted an effect of interference type on categorical perception. Specifically, we predicted that in the control group, participants would be more accurate at discriminating between faces from different racial categories (between-category) than from the same (within-category). The spatial interference condition was predicted to have no effect on categorical perception and participants were expected to demonstrate an advantage for discriminating between faces that come from different racial categories.
relative to faces from the same category (Levin & Angelone, 2002). However, on trials with verbal interference, we expected categorical perception effects to be eliminated (i.e., equivalent discrimination of between and within race faces). In sum, we anticipated that the spatial and control conditions would show categorical perception for race, while the verbal condition would not show a between-group advantage.

Participants

Participants were 100 White, right-handed undergraduates (70 women, 30 men; $M_{age} = 20.47$) from the Psychology department study pool at California State University, Northridge. All participants were assigned to one of three conditions: control ($n = 37$), verbal interference ($n = 29$), or spatial interference ($n = 34$).

Design

The study combined a categorical perception task with three between-subjects conditions, resulting in a 3 (interference type: control/verbal/spatial) $\times$ 2 (difference type: between-category/within-category) mixed group design with accuracy rate and reaction time as dependent variables of interest. In the manipulation check the interference task was the same as in the experimental task, but the primary task was to determine whether the faces presented were the same or different.

Stimuli

Stimuli used for the categorical perception tasks were the same Black and White morph faces used in study 1. Similarly, the interference stimuli were also identical to study 1.
**Questionnaire.** Participants completed a demographics questionnaire. Upon completion of the task and questionnaire they received partial course credit and were fully debriefed.

**Procedure**

The manipulation check was performed on PC computers using the reaction time software E-prime. In the verbal interference version of the task, participants were presented with a box of (target) letters over a letter mask for 1000 ms. Following the letter presentation, two faces were presented (450 ms) followed by a back mask in which participants were given infinite time to judge whether the two faces were the same or different. After the judgment, participants continued the task and saw a second box of (test) letters, to which they indicated if the letter did not match the target letter within 2500 ms. In the spatial interference version the procedure was identical with the exception that participants were asked to indicate when the location of the test letter did not match the location of the target letter. The control condition operated identically to the spatial and verbal versions, but participants were asked to do nothing during letter presentations. There were 27 practice trials total. First participants went through 7 practice trials with the interference task alone, followed by 10 practice trials with the categorical perception task alone. Following the isolated practice trials, there were 10 practice trials with the combination of both the interference and categorical perception task. Practice trials were followed by 100 test trials where the categorical perception and interference tasks were combined. Each of the 5 faces was presented 18 times, 3 times on each of the four levels of racial prototypicality. In the categorical perception task, faces were the same on 40 presentations, and different on 60. In the interference task,
approximately 30% of the interference trials presented matching letters, while 70% of the trials were nonmatching.
RESULTS

Accuracy Rate

Because the faces used in this study were morphs and high in featural similarity, we were primarily interested in participant accuracy rates on between-category trials compared to within-category trials. Given that the faces were Black-White morphs, faces varied in objective similarity on the 5% to 95% continuum, making the difficulty of the comparisons in the task also vary on a continuum. For example, a highly prototypic White face that is a 95% morph is very different from a highly prototypic Black face that is a 5% morph, as the degree of separation between the two on the continuum is 90%. In contrast, a low Black prototypic face (35% morph) is only 30% different from a low White prototypic face (65% morph). Of primary interest in this study is participant accuracy rate when comparing faces that are relatively similar, and are thus 30% different from each other. All of the following analyses are conducted on the 30% difference trials, as this allowed us to get at the categorical perception effect in the most controlled manner and assess participant sensitivity to fine-grain features such as racial prototypicality.

In order to test the hypothesis that interference would have differential effects on categorical perception we conducted a 3 (interference type: control/verbal/spatial) × 2 (difference type: between-category/within-category) mixed-ANOVA on accuracy rates where difference type was a within-subjects factor, and interference was a between-subjects factor. There was a main effect of difference type $F(1, 97) = 4.78, p = .03, \eta^2 = .05$ (see Figure 7). As anticipated, there was superior accuracy for between-category judgments ($M = .43, SD = .23$) relative to within-category ($M = .39, SD = .20$). There was
a main effect of interference type $F(2, 97) = 4.3, p = .02, \eta^2 = .08$, in which the control condition was the most accurate on average ($M = .48, SD = .03$; see Figure 8) followed by the verbal condition ($M = .38, SD = .03$), and spatial ($M = .36, SD = .03$). Pairwise comparisons revealed that accuracy was significantly different between control and verbal $t(64) = 1.92, p = .029$, and control and spatial $t(69) = 2.45, p = .01$, but verbal and spatial did not significantly differ $t(61) = 0.26, p = .69$. There was no evidence of a difference type × interference type interaction, $F(2, 97) = 0.04, p = .96, \eta^2 < .01$.

**Reaction Time**

We also investigated reaction times in order to test for differences in categorical perception as a function of interference type using a 3 (interference type: control/verbal/spatial) × 2 (difference type: between-category/within-category) mixed ANOVA. Reaction times were trimmed where trials were removed that were below or exceeded 2.5 standard deviations from a participants’ average response time (approximately 29% of the trials). The remaining reaction times were then log transformed. There was no main effect of difference type $F(1, 88) < .01, p = .95, \eta^2 < .01$. There was no main effect of interference type $F(2, 82) = 0.10, p = .91, \eta^2 < .02$. The interaction between difference type and condition was also not significant $F(2, 88) = 0.69, p = .51, \eta^2 = .02$. Contrary to previous reports, we did not find any evidence for categorical perception in reaction times. That is, there was no evidence that participants responded faster when making between-category decisions than within-category decisions. However, given the difficulty of the categorical perception task, we did not expect to observe any effects of difference type or interference type on reaction time.
DISCUSSION

In summary, we replicated a basic categorical perception effect for racial prototypicality where participants were more accurate at judging whether faces that were from different (between) racial categories than the same (within) category. However, there was no evidence that interference type moderated this effect. Despite this null result, we probed the effect of categorical perception within each interference type to better understand the interference manipulation. Here, we observed no significant differences between within-category and between-category accuracy in the control \( t(36) = 1.39, p = .17 \), verbal \( t(28) = 1.19, p = .24 \), or in spatial \( t(33) = 1.28, p = .21 \). There was no evidence of a difference type \( \times \) interference type interaction, \( F(2, 97) = 0.04, p = .964, \eta^2 < .01 \). Although the verbal and spatial interference conditions appeared to weaken categorical perception relative to the control condition, this effect was not statistically significant. Counter to our hypothesis, we did not find support for differential effects of verbal and spatial interference on categorical perception, and seemed to find only differences in accuracy between the control compared to the spatial and verbal conditions. In sum, our spatial and verbal interference conditions reduced average performance for participants, but failed to disrupt category activation in the hypothesized manner.
GENERAL DISCUSSION

Overall, participants did not respond differently to verbal and spatial interference on the categorical perception or the WIT task. That is, counter to the hypothesis, both verbal and spatial interference appeared to harm categorical perception. How the interference may have changed performance relative to the control condition may have varied between experiments. In the categorical perception task there was a main effect of interference on accuracy scores, an effect that was not replicated on the WIT task. These differences may be an artifact of differences in task difficulty. For example, the categorical perception task requires participants to judge between ambiguous stimuli, while the WIT is comparatively straightforward for object (gun versus tool) identification. Attesting to this distinction in difficulty, categorical perception performance was approximately 65% for trials discriminating between 30% faces, while for the WIT participants averaged accuracy of approximately 85%. An additional factor adding to the difficulty for the manipulation check is how similar the instructions for the interference task (secondary task) were to the categorical perception task (primary task). On the secondary task participants were told to keep in mind a letter (location) and determine if the letter (location) was different on a later trial. In-between letter (location) presentations, they were asked to judge if two faces were the same or different. That is, both tasks required participants to judge whether stimuli were different from each other, making the primary task more vulnerable to unintended interference effects from task similarity. Overall, because of primary and secondary task similarity, the manipulation check may have been impacted to a greater extent by cognitive load than was the WIT.
While verbal and spatial interference did not affect FBS differentially in the hypothesized manner, there were no significant differences in accuracy between conditions. This suggests that the interference tasks were similar in difficulty for participants. However, in the manipulation check we found an accuracy difference between the control and verbal as well as control and spatial conditions, suggesting that the control condition was not similarly well matched. In the experimental task, the differences in difficulty between the interference conditions and control were less clear. For example, there were no significant differences in timeouts across interference conditions; however, 5 participants were removed from the control condition due to below chance accuracy on the WIT, which suggests possible engagement differences between conditions. Overall, while verbal and spatial may be well matched for difficulty, it is possible that our control condition was not as comparable to either form of interference.

As noted above, we failed to find the predicted differences between our spatial and verbal conditions. Although load effects appeared to be more pronounced for the categorical perception task, a similar pattern of interference disrupting FBS was observed. These results suggest that while the tasks may have been of similar difficulty, the forms of interference may not have been distinct enough to have differential effects on categorical perception. It is possible, for example, that participants found a way to verbally encode the spatial interference. When a letter box appeared participants could have been mentally rehearsing the location (i.e., “upper left corner”) in a phonological manner rather than mentally visualizing the location while doing the primary task. Additionally, while verbal and spatial conditions were not significantly different from
each other on measures of performance, there was a trend toward participants in spatial interference conditions to perform worse than any other condition. Either due to the utilization of verbal strategies, or subjectively greater difficulty, the spatial condition was potentially more challenging for participants.

On both the manipulation check and experimental task we were able to replicate prior research. Participants showed an advantage for discriminating faces that were from different racial categories relative to within-category. On the WIT there was mixed evidence for replication of racial bias, however criterion and reaction time measures uncovered differential sensitivity within race to high and low prototypic targets. Participants had a higher criterion for highly White prototypic targets relative to low White targets, suggesting that low White targets facilitated the activation of danger-related stereotypes. Similarly, multilevel modeling uncovered a similar bias for White prototypic targets on reaction time where participants were faster to identify objects as guns after a low prototypic White prime than after a low Black prime. These results suggest that our task was successful at eliciting FBS, specifically, that less White-looking targets activated danger stereotypes more strongly than did more White-looking targets or less Black-looking targets.

**FBS Differences for Black and White targets**

Although White-only FBS was not a specific part of our hypothesis, it is consistent with prior studies. For example, Ronquillo et al. (2007) found an interaction between skin tone and race in amygdala activation. Specifically, participants showed approximately the same amygdala response to Black faces that had darker or lighter skin,
but showed greater amygdala activation to White faces with darker skin relative to lighter skin tone. These findings suggest that participants may have a stronger affective response to variations in racial prototypicality on White faces than Black faces.

Similarly, Ma and Correll (2011) found FBS on the first-person shooter task (FPS) that was primarily driven by sensitivity to prototypicality of White faces. Highly prototypic White targets were significantly less likely to be shot when armed, and more likely to be not shot when unarmed compared to low prototypic targets. Although multiple studies (Blair, 2006; Blair, Judd, & Chapleau, 2004; Blair et al., 2005; Blair et al., 2004; Blair et al., 2002) have found greater stereotype activation for highly Black prototypic targets, many of them have been methodologically dissimilar to the present task. Specifically, much of the work on FBS has been obtained from impression formation tasks, rather than implicit tasks. Impression formation tasks allow participants unlimited time to come to a judgment about a target, while implicit tasks such as the WIT or FPS have enforced response times of around 500ms. Given the increased difficulty that comes from a speeded response task, and the well-noted advantage for discriminating between ingroup faces relative to the outgroup (Kelly et al., 2007; Tanaka, Kiefer, Bukach, 2004), it may not be surprising that White participants only show sensitivity to White prototypicality.

Another potential explanation for our study’s failure to find FBS for Black faces may be context. In a threatening context where an individual must judge how threatening a target is, or in the context of threat-related objects White participants may demonstrate a magnified cross-race deficit for processing outgroup faces. That is, threatening situations increase the cognitive load imposed on an individual, draining cognitive
resources that are needed to make more fine-grained judgments of threatening stimuli, such as outgroup faces (Van Dillen, Heslenfield, & Koole, 2009). Comparatively, processing of ingroup faces is an efficient process (Kelly et al., 2007; Tanaka, Kiefer, Bukach, 2004), allowing greater sensitivity to prototypicality for White targets and the observed FBS. Overall, threat judgments, similar to what the shooter task, and the WIT requires participants to do, may impose a greater strain on cognitive resources and eliminate FBS of outgroup targets, but not ingroup ones.

**Future Directions**

In the manipulation check and WIT task we were able to replicate prior literature, suggesting that the primary tasks are appropriate, but that the interference manipulations should be adjusted. The verbal interference condition essentially operated as intended, disrupting categorical perception. In contrast, the spatial condition showed not only worse accuracy than the control condition, but also less use of categories. These findings suggest that any follow-up should focus on the spatial condition, while keeping the verbal task as similar as possible.

A promising solution would be to change the spatial interference task. Specifically, changing the spatial interference stimuli so it may be less easily verbalized, such as by using novel shapes (Makovski, Sussman, & Jiang, 2008), or dot patterns (Roberson & Davidoff, 2000). By changing to a form of interference that is more difficult to put an existing label on, participants would be prevented from unintentionally using verbal encoding to keep track of visual information such as location. In this way we could
ensure that participants are using the intended rehearsal strategies and increase the difference between our verbal and spatial conditions.

Although we were unsuccessful at testing our hypothesis due to manipulation difficulties, pursuing the question of what mechanism underlies FBS remains important. Whether FBS is category mediated or takes place through DFTA has significant implications for the field and the real world. For example, if FBS is a result of a direct association of features to a stereotype, FBS would operate differently from between-group stereotyping, making current approaches (i.e., category suppression) to training individuals out of such stereotyping unlikely to be effective. Given the observed harmful effects of FBS ranging from trait judgments (Blair, Chapleau & Judd, 2005) to court decisions (Eberhardt et al., 2006), gaining a better understanding of FBS and as a result potentially reduce its effects continues to be a valuable endeavor.
REFERENCES


doi:10.1016/j.jesp.2009.04.021


According to embodied cognition, categories are formed around frames of perception from multiple sensory sources (Barsalou et al., 1993, Joyce, Richards, Cangelosi, & Coventry, 2003). If embodied cognition is correct, verbal interference would not have the intended effects of categorization, and would lead to a reliance on less explicit categorization strategies.

In order to access possible engagement differences and load effects across interference type, an ANOVA was conducted on average reaction time in the experimental task. There was no significant difference between all three levels of interference in average reaction time $F(2, 84) = 0.87, p = .424$, suggesting that all levels of interference were approximately equal in difficulty. Average timeouts were also investigated to rule out differences in task engagement across all three interference conditions. There were no significant differences in average timeouts across interference type $F(2, 85) = 0.01, p = .449$. 

Footnotes

1 According to embodied cognition, categories are formed around frames of perception from multiple sensory sources (Barsalou et al., 1993, Joyce, Richards, Cangelosi, & Coventry, 2003). If embodied cognition is correct, verbal interference would not have the intended effects of categorization, and would lead to a reliance on less explicit categorization strategies.

2 In order to access possible engagement differences and load effects across interference type, an ANOVA was conducted on average reaction time in the experimental task. There was no significant difference between all three levels of interference in average reaction time $F(2, 84) = 0.87, p = .424$, suggesting that all levels of interference were approximately equal in difficulty. Average timeouts were also investigated to rule out differences in task engagement across all three interference conditions. There were no significant differences in average timeouts across interference type $F(2, 85) = 0.01, p = .449$. 
Figure 1. Explanations for Feature-Based Stereotyping
APPENDIX B

Figure 2. Schematic of the experimental task.
APPENDIX C

*Figure 3.* Faces used in both the experimental task and manipulation check.
APPENDIX D

Figure 4. Categorization results from Ma and colleagues unpublished manuscript.
Figure 5. Interaction between prototypicality and race on WIT criterion scores.
APPENDIX F

*Figure 6.* Interaction between prototypicality, race and object on WIT reaction time.
Figure 7. Effect of difference type on categorical perception accuracy rate.
Figure 8. Interference type effect on categorical perception accuracy rate.