The Automatic Social Categorization Test: Validating a New Measure

Mara P. Sedlins

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Program Authorized to Offer Degree:
Department of Psychology

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Mara P. Sedlins

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the supervisory committee:

Yuichi Shoda

Reading committee:

Miriam Bassok

Anthony G. Greenwald

John M. Miyamoto

Yuichi Shoda

Date: 11/10/2011

University of Washington

Abstract

The Automatic Social Categorization Test: Validating a New Measure

Mara P. Sedlins

Chair of the Supervisory Committee:
Professor Yuichi Shoda
Department of Psychology

Categorization is an essential process in making sense of the world. However, it simplifies perception in a way that distorts reality to some degree. For instance, the color spectrum is a continuum of wavelengths; but colors are grouped into linguistically defined categories (Berlin & Kay, 1969). Although people can tell the difference between different hues within a category, they are better at discriminating colors from different categories than colors from the same category, even if they differ by the same degree in terms of physical wavelength. Harnad (2003) describes this process as "'warping' perceived similarities and differences so as to compress some things into the same category and separate others into different categories." This dissertation describes the Automatic Social Categorization Test (ASCAT), a new way to measure the *extent* to which this process occurs for social categories (e.g., gender, age, race) by examining confusion errors among stimuli that vary continuously between two categories. We operationally define automatic categorical perception as the degree to which confusion rates are higher for a pair of stimuli within the same subjective category than for a pair belonging to different subjective categories. We first present a series of validation studies with non-social stimuli, showing that automatic categorical perception is stronger for categorical stimuli (i.e., with a gap in the middle) than for continuous stimuli. Next, we extend our method to social stimuli (i.e., morphed faces varying in racial composition), demonstrating a categorical trend that is reduced when social information is removed (i.e., when the faces are scrambled). The degree of automatic social categorization also depends on which social category is presented; faces varying in gender and race are perceived more categorically than faces varying in age. Finally, we show that there are reliable individual differences in automatic race categorization that cannot be attributed to working memory ability or differential fatigue effects. Performance on the ASCAT is also relatively stable across multiple testing sessions. These results suggest that the ASCAT is a useful new tool for research on the correlates and consequences of automatic social categorization.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number Page

ACKNOWLEDGEMENTS

I would like to thank my advisor, Yuichi Shoda, for his support throughout the dissertation process. I am also grateful to research collaborators Lori Wu Malahy and Jason Plaks for their important contributions to the development of this new measure. Portions of this dissertation were written as part of a NSF Grant application prepared by the four of us. My committee also provided helpful feedback on earlier versions of this dissertation, and this research has greatly benefitted from many discussions with past and current members of the Shoda Lab over the years. This research would not have been possible without the help of the computer programmers who implemented the ASCAT procedure, or the research assistants who assisted with data collection. Finally, thank you to the NSF Graduate Fellowship program for financial support.

DEDICATION

To my father, who taught me to see the world through creative eyes.
To my mother, who taught me the joys of patience and persistence.

CHAPTER 1. INTRODUCTION

Many aspects of the world vary continuously; for instance, light varies in wavelengths, musical pitches vary in frequency, and speech sounds vary in factors like voice onset time (which makes the difference between the sounds /ba/ and /pa/). Yet, people tend to experience the world in terms of categories (e.g., colors, musical notes, phonemes). Researchers have taken a variety of different theoretical and empirical approaches to understanding this phenomenon. Most generally, the process of categorization is fundamental to perception, cognition, and language (Harnad, 1987; Medin, 1989) and has been a focal point of important debates in cognitive research (e.g., whether categories are represented by prototypes or exemplars; whether similarity judgments are best accounted for by spatial or feature set models). However, the focus of this dissertation is categorical perception, specifically defined as a differential ability to discriminate pairs of stimuli depending on where they are located along a continuum. A poorer ability to tell two items apart is interpreted as an indication that they have been parsed into the same category. In addition to documenting categorical perception (which has already been done in many ways and in many domains; see a brief review below), we are interested in exploring differences in the *degree* to which this phenomenon occurs, focusing on the domain of social categories. Do some social dimensions (e.g. gender) elicit stronger categorical perception than others? Do people differ in their tendencies to categorize others?

In order to answer these questions, we have developed a measurement tool we call the Automatic Social Categorization Test (ASCAT). A detailed description of the task will follow in Chapter 2, but briefly, it is a sequential discrimination task based on the 2-back memory procedure, which was originally developed for use in neuroimaging studies of verbal working memory (Smith, Jonides, Marsheutz, & Koeppe, 1998). Participants are presented with targets one at a time on a computer screen and are asked to indicate whether each one is identical to the last target. In our version of the task, targets are faces that vary along a social continuum (e.g., from a White face to a Black face). The task is designed to be difficult enough that participants will make errors, but not too difficult with stimuli that are difficult to verbalize (i.e., faces; Schooler & Engstler-Schooler, 1990).

One reason we chose a memory performance task to assess categorical perception is that we think it reduces the opportunity for self-presentation bias. Historically, categorical perception research has focused on fairly neutral domains, such as colors or phonemes, and methods often included self-report measures such as direct classifications or similarity ratings. But social categories are laden with evaluative and affective associations, such as racial stereotypes (Allport, 1954; Devine, 1989; Tajfel, 1981). To the extent that participants notice that stimuli vary along a social dimension, they may become sensitized to the fact that they are being evaluated on how they perceive that attribute and may wish to avoid appearing, for instance, as someone who thinks "all X people look alike." However, we think that the indirect and challenging nature of the memory task reduces the likelihood that participants will alter their responses in order to present themselves in a particular way.

As we will discuss in greater detail in Chapter 2, we operationalize automatic categorical perception as a pattern of confusion rates that reflects lower memory errors at a boundary between categories, typically in the middle of a continuum, than within categories. If a participant wanted to manipulate their responses to appear less categorical, they would need to either increase their accuracy for within-category stimuli (typically at the ends of the continuum) or intentionally make errors for stimuli at the boundary (typically in the middle of the continuum). This seems unlikely, given the sophistication and complexity of the required strategy, as well as the difficulty of the task. We require participants to respond fairly quickly, within 1500 milliseconds on average (see Chapter 2 for timing details). This is enough of a challenge that the average confusion rate for neighboring stimuli is approximately 50%. Thus, participants' attention is likely to be focused on performing the task, without room for thoughts of how they are self-presenting, much less how they might go about manipulating their responses in a systematic way.

In addition to making self-presentation bias unlikely, we think the ASCAT taps into the automatic nature of social categorization. We have chosen to refer to the construct that the ASCAT measures as "automatic social categorization" (or "automatic categorical perception," more generally). Automatic processes specifically are thought to differ from controlled processes in that they occur outside of awareness and are largely involuntary, unintentional, and effortless (Macrae & Bodenhausen, 2001). However, we recognize that there are a variety of ways to describe dual processes in cognition (implicit vs. explicit, unaware vs. aware, unconscious vs. conscious, intuitive vs. analytic, direct vs. indirect, and procedural vs. declarative; Greenwald &

Banaji, 1995), and several of these distinctions may be relevant to the ASCAT. The task might be described as an indirect measure in that we infer representational categories from participants' errors. It might also be considered implicit to the degree that participants' performance on the task is influenced by past experience in a way that they are not able to report. For instance, the degree to which a participant confuses faces that have similar racial phenotypes might be affected by the racial diversity they were exposed to growing up.

We have not systematically tested these aspects of ASCAT performance, but research in stereotyping suggests that the activation of social category information occurs automatically (e.g., Brewer, 1988; Devine, 1989; Fiske & Neuberg, 1990). To the extent that social categorization is a prerequisite for stereotype activation, social categorization processes are likely to be automatic as well. In fact, social categories like age, gender, and race are noticed very quickly – in the case of race, as soon as 120 ms after being exposed to a target face (Ito references). Because the ASCAT requires a fairly quick response (participants are given a response window of about 1500 ms per trial), performance is likely to reflect encoding processes on a timescale similar to perception of these categories in the real world.

The studies presented in this dissertation aim to validate the ASCAT as a measure of automatic categorical perception, capable of detecting variation in the strength of this tendency across social categories and individuals. But first, to situate the present research with regard to related literatures, below follows a selective review of categorization, models of similarity, categorical perception, and social categorization.

Categorization (object concepts and general knowledge structures).

Categorization is a fundamental cognitive process. Categories allow people to view the world in terms of discrete objects and ideas rather than an infinitely varying array of stimuli. They allow items belonging to the same category to be treated equivalently, guiding people's behavior. By classifying an item as, say, a chair, then one knows the item can be sat upon. There has been a good deal of theoretical and empirical work on general categorization processes (for reviews, see Rosch, 1978; Medin & Smith, 1984; Smith & Medin, 1981; Medin, 1989), addressing the structure of category knowledge, classification strategies, and the relation between categories and variation in the real world.

The classical view of concept structure held that a category should be definable based on a finite set of rules. For instance, the category “bachelorette” requires being an adult, never married, eligible, living, female, and human. Thus, any given stimulus can be determined as either belonging to a category or not in an all-or-none fashion. This view is problematic because for many categories, it is difficult to come up with an adequate set of defining rules (e.g., the category “games”; Wittgenstein, 1953), and category membership is often not all-or-none. Some instances of a category are better than others (e.g., a robin is seen as a better example of the category “bird” than an ostrich; Rosch, 1973), and there are ambiguous cases which do not clearly fall into or outside of a given category (e.g., is croquet a sport or not?; McCloskey & Glucksberg, 1978).

Because of these issues, the probabilistic view of concept structure is preferred by scholars. This view allows for the existence of “fuzzy categories” which have a graded structure; categories are defined by clusters of correlated attributes rather than by all-or-none rules, so particular items can be more or less typical of a category (Rosch, 1973; 1975; Rips, Shoben, & Smith, 1973; Smith, Shoben, & Rips, 1974; Barsalou, 1985). Thus, although people have a tendency to parse the world into categories, they also perceive variation within categories. This perspective in turn allows for the possibility of variation in the “categoricalness” of categories. That is, since categories do not need to be all-or-none, then some categories can be more discrete than others. Also, people may vary in their tendency to perceive variation within categories, in their sensitivity to the fact that some items are more typical of a category than others.

Given the probabilistic view of categorization, a further debate in this field concerns how people determine how typical of a category a particular item is; is it compared to an idealized prototype of that category (Posner & Keele, 1968; Reed, 1972; Rosch & Mervis, 1975), or to specific exemplars of the category that have been encountered in the past (Medin & Schaffer, 1978; Nosofsky, 1986; 1992)? Although each of these models has empirical support, there is not a general consensus on which one is correct. With regard to the phenomena explored in this dissertation, the tendency to view continuously varying stimuli in a more or less categorical manner could be explained by either the prototype or the exemplar theory of categorization. That is, a particular social target could be compared to either a prototype of a social category (e.g., a "stereotypical Black person") or to exemplars of the category (e.g., "Black people I have met"). Thus, we do not take a stance on this particular debate.

An issue that relates more directly to this dissertation is the relation between conceptual categories and variation in the real world. One possible view is that

> "… the physical and social environment of a young child is perceived as a continuum. It does not contain any intrinsically separate 'things.' The child, in due course, is taught to impose upon this environment a kind of discriminating grid which serves to distinguish the world as being composed of a large number of separate things, each labeled with a name" (Leach, 1964, p. 34, as quoted in Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

However, Rosch et al. claim that variation in the world is not intrinsically unstructured, and the "discriminating grid" we use to make sense of the world is not imposed in an arbitrary way. There seem to be natural ways of grouping items based on patterns of co-occurrence in their attributes. For instance, feathers are more likely to co-occur with wings than is fur, so the category "bird" is more likely than a category representing "furry things that have wings." There are also levels of abstraction that are more informative than others; knowing that something is a bird tells us more about that object than knowing it is an animal, but knowing that it is specifically a robin does not add information that would significantly change predictions about or behavior toward it. Rosch and colleagues argue that the most common and useful categories are these basic objects.

The application of this idea to the social domain is complicated, because while some human attributes do vary in predictable ways (e.g., depending on the particular culture and time period, having long hair might co-occur more frequently with being a woman than with being a man), people often have inaccurate beliefs about the ways in which social attributes vary (e.g., gender stereotypes). However, people may vary in the strength and content of these beliefs, as well as the extent to which social categories guide behavior (e.g., prejudice). Thus, although there are likely to be basic levels at which people tend to categorize others (e.g., gender, race, and age categories), people may vary in the extent to which they find these categories important and useful. We expect these individual differences to be associated with the tendency to automatically group people with ambiguous category-relevant attributes (e.g., a gender-neutral hairstyle) into to one category or another. With the ASCAT, we hope to facilitate the empirical investigation of this type of question.

Models of similarity.

Closely related to categorization is the study of similarity. If category membership is determined by similarity to a category prototype or exemplar, then how is similarity represented? And how can measures of similarity be used to reveal the underlying structure of categories? Two major approaches to these questions are spatial models (e.g., Shepard, 1962a; 1962b; Kruskal & Wish, 1978) and feature set models (Tversky, 1977) of similarity.

Spatial models of similarity typically use multidimensional scaling (MDS) to arrange stimuli in a multidimensional space; similarity is defined as the metric distance between two items. This technique relies on behavioral indicators of perceived similarity (e.g., similarity ratings, sorting tasks, or confusion errors) and uses an iterative mathematical function to generate an optimal arrangement of items such that the distance between each pair reflects their degree of psychological similarity. The dimensions that emerge from this process and the locations of items with respect to each other can then be inspected to reveal the structure of underlying categories. Individual differences in the importance of dimensions can also be examined using individual differences scaling (INDSCAL; Carroll & Chang, 1970), which generates a common solution for all participants and assigns each participant a weight for each dimension, indicating its relative importance for that participant.

One difficulty with spatial models of similarity is that it can be difficult to interpret the dimensions produced by MDS. Researchers may perform post hoc inspections of the way stimuli are arranged (a bottom-up approach) or verify the meanings of dimensions with independent ratings of the stimuli on features they believe to be relevant (a top-down approach), or both. However, the MDS solution in itself does not explain what the resulting dimensions mean. In using INDSCAL, it can also be problematic to assume a common solution across individuals.

Tversky (1977) had two fundamental criticisms of the spatial model, leading him to develop the contrast (or feature set) model as an alternative. First, spatial models assume that variation among stimuli can be represented by a relatively small number of quantitative dimensions. Tversky pointed out that this assumption may be appropriate for some domains, but that many domains vary in more than just a few ways and can vary qualitatively. Secondly, the metric approach assumes three axioms that similarity judgments do not satisfy: minimality, symmetry, and the triangle inequality. The minimality axiom implies that an object must be more similar to itself than to another object, the symmetry axiom implies that the similarity between A and B must be equal to the similarity between B and A, and the triangle inequality implies that if A is similar to B and B is similar to C, then A must be similar to C. However, experimental data have disconfirmed each of these axioms with regard to human perceptions of similarity.

To address these issues with spatial models, the contrast model represents similarity between two items instead as a weighting of their common and distinctive features. Similarity increases with the number of features two items share and decreases with the number of features that are unique to either item. This model also includes parameters for the importance of each feature in the particular context in which similarity judgment occurs.

Although the ASCAT paradigm is not in principle tied to a spatial model, the stimuli discussed in this dissertation were developed with a spatial model of similarity in mind; stimuli are conceptualized as being equally spaced along a social dimension (e.g., race), and the "distance" between two stimuli is interpreted as the objective similarity between them. This objective distance is contrasted with the psychological distance between pairs of stimuli, operationalized as confusion rates. Although MDS is not our primary analytic technique for interpreting ASCAT data, for illustrative purposes we performed MDS with confusion rates (data from Study 4), constraining the solution to one dimension. In Figure 1.1, the resulting psychological distances given by MDS are plotted on the Y axis against the objective distances between faces morphed between a White face and a Black face. The data suggest a threshold function characteristic of categorical perception (to be discussed in more detail below).

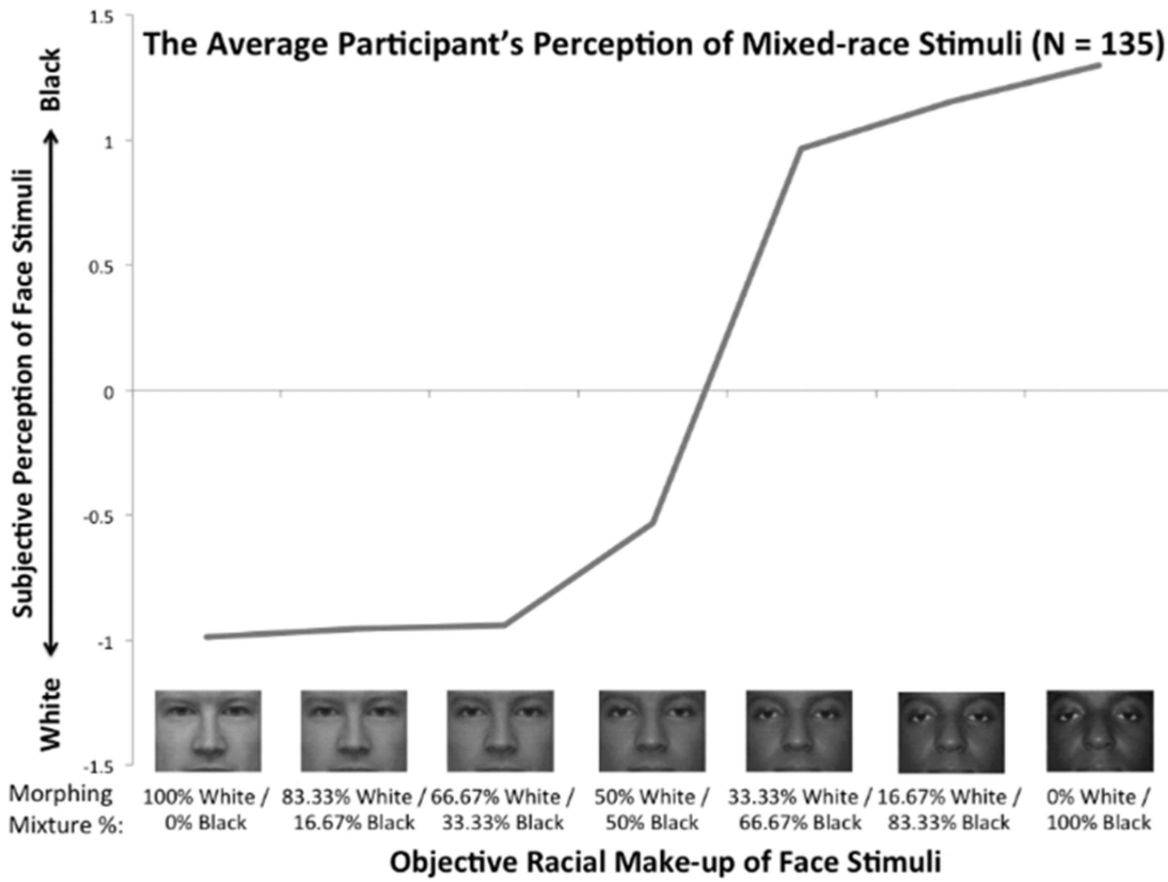


Figure 1.1. Psychological Distance Plotted Against Objective Distance Between Faces Morphed Between a White Face and a Black Face.

Since the stimuli we have used, such as those shown in Figure 1.1, are based on a spatial model of perceived similarity, we must bear in mind the criticisms of spatial models when interpreting our results. Social stimuli vary along many different dimensions, but as it has been used thus far, our paradigm examines only one dimension at a time. This is reflected in the way that stimuli are created, as well as in how the data are analyzed. Also, the perceived similarity between a given pair of faces may vary depending on the context. For instance, if the 100% Black face is seen as more typical of its group, then the 83.33% Black face may be seen as more similar to the 100% Black face than vice versa.

It is also possible that the feature set model could explain data obtained from the ASCAT; rather than representing faces as being closer or farther apart on a social dimension, participants might represent faces in terms of the proportion of features they have in common. In fact, during debriefing sessions, participants have occasionally mentioned that they paid attention to specific features of the faces when determining whether they were identical or not. It is not a goal of the present research to differentiate between spatial and feature set models of similarity; whichever form of representation underlies perceived similarity, it is the evidence of changing degrees of psychological similarity across pairs of faces that we are interested in.

Categorical perception (sensory concepts).

According to Rosch (1977), categorical perception occurs when continuously varying stimuli are perceived as falling into discrete groups and when stimuli within each group are seen as interchangeable (i.e., all group stimuli are perceived as more or less equivalent to each other). For instance, most Japanese speakers do not perceive the difference between English /r/ and /l/ sounds (e.g., the words “rock” and “lock” sound the same; McClelland, Fiez, & McCandliss, 2002). In their minds, /r/ and /l/ belong to the same category of sounds. However, for English speakers, the two sounds are clearly distinguishable, even though they are physically similar (Slawinski, 1999). Thus, stimuli that fall near each other on a physical continuum, such as /r/ and /l/, can be automatically perceived either as the same or different, depending on the categories they are sorted into by an individual’s perceptual system.

Such discrete changes in discriminability at category boundaries, referred to as categorical perception (CP), have been documented in a number of domains, including colors (e.g. Bornstein & Korda, 1984; Rosch, 1975; Winawer, et al., 2007), linguistic phonemes (e.g., Liberman, Harris, Hoffman, & Griffith, 1957; MacKain, Best, & Strange, 1981), and musical intervals (Burns & Ward, 1978; Perlman & Krumhansl, 1996). Categorical perception has also been documented with faces, with morphed continua created from familiar faces showing a stronger CP effect than those created from unfamiliar faces (Beale & Keil, 1995; Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002). There is also evidence that facial expressions are perceived categorically (Calder et al., 1996; Etcoff & Magee, 1992). However, to our knowledge, only two studies have examined CP of morphed faces varying from one race to another (Levin & Angelone, 2002; MacLin et al., 2009). The relation of these studies to the current research will be discussed in greater depth in Chapter 4.

Measuring categorical perception. Categorical perception (CP) has been defined in a number of different ways. The most typical definition of CP involves the relation between identification performance (labeling stimuli as belonging to one category or another) and discrimination performance (the ability to tell two stimuli apart). However, some definitions use only identification (Ehret, 1987) or only discrimination performance (Wilson, 1987; Pastore; 1987). The strong definition of CP, consistent with the classical view of categorization described earlier, holds that items within a category are indistinguishable from one another (e.g., Liberman et al., 1957). However, this position does not hold up to empirical evidence; for instance, people are able to hear differences between speech sounds within a phoneme category (Pisoni & Tash, 1974). Another definition requires that discrimination performance be the same within categories (Eimas, Miller, & Jusczyk, 1987); here, discrimination may be above chance, but it must be worse than performance between categories.

There are also various ways to measure discrimination performance. The ABX paradigm presents two targets (A, then B), followed by a third target (X), which is always the same as either A or B and must be identified as such. A variant of this procedure is the 4IAX paradigm, in which two pairs are shown, one which is identical, and one with two different stimuli; participants must identify the non-identical pair. Other researchers have used a same-different task, in which two stimuli are presented simultaneously, and participants identify whether they are identical or not; the dependent variable is then accuracy (Goldstone, 1994) or reaction time (Bornstein & Korda, 1984). Discrimination performance can be inferred in infants, using a novelty-preference procedure (e.g., Quinn, 2002) or habituation paradigm (e.g., Lasky, Syrdal-Lasky, & Klein, 1975). It can also be measured in animals; for instance, pigeons can be trained to respond when two stimuli match, and then tested with stimuli that differ by varying degrees (e.g, Wright & Cumming, 1971). Non-human primates can be tested for discrimination ability by being rewarded for one response when presented with identical targets and a different response when presented with non-identical targets, or being rewarded for identifying a target that matches a sample (Thompson & Oden, 2000).

In addition to behavioral measures, ERP indicators can be used to measure perceptual discrimination. For auditory stimuli, mismatch negativity (MMN) response occurs when a series of identical sounds is occasionally interrupted by a deviant sound (the oddball paradigm); this response can occur without conscious awareness and is considered a measure of pre-attentive processing of sound differences (Näätänen, 2001). The P300 is examined with the same oddball paradigm but reflects conscious processes (Gaillard & Verduin, 1985), elicited when participants perform a task in which they must discriminate between different stimuli. Because it reflects a general task response rather than auditory processing specifically, the P300 measure can be used with both auditory and visual stimuli. ERP measures of discrimination performance have the advantage of eliminating the possibility of self-report bias, since they directly measure brain activity. They also allow researchers to examine the time course of categorization processes. However, ERP studies are typically more time-consuming and expensive than behavioral studies, and because they are often limited in their number of participants, they are less conducive to examining individual differences.

Categorical perception vs. categorical partitioning. Although categorical perception, defined as better within-category than between-category discrimination, has been documented with a variety of procedures in a number of stimulus domains, research has not supported the strict definition of categorical perception, in which within-category stimuli are not distinguishable from one another at all. For many domains, although participants may be poorer at discriminating items within a category than between categories, they are still able to make within-category distinctions. For instance, in color perception research, Bornstein and Korda (1984) found that reaction times for a same-different judgment task were faster for identical color pairs than for non-identical color pairs, even if the two colors belonged to the same color category.

This type of finding has led to a discussion of whether CP is truly a perceptual phenomenon or whether it is a result of task demands (Massaro, 1987; 1998; Hary & Massaro, 1982). For instance, if a same-different task requires participants to hold items in memory, they may recode the stimulus as a verbal label in order to remember it (Pisoni, 1971, 1973; Pisoni & Lazarus, 1974), producing a categorical pattern of responses even if participants originally perceived the stimuli to vary continuously. Thus, Massaro (1987) advocates a continuous view of perception and conceptualizes CP phenomena as “categorical partitioning” rather than categorical perception.

This criticism of categorical perception research is likely to hold for the ASCAT; the task has high memory demands, so it is a strong possibility that participants use verbal labels as a strategy to perform the task. Thus, high confusion rates between two stimuli may indicate that participants assigned them the same label, rather than literally perceiving them to be the same. The data presented here do not allow us to differentiate between these two possibilities, but even if the phenomenon captured by the ASCAT is categorical partitioning rather than categorical perception, we are still interested in how this tendency varies across domains and individuals. The possibility that individual differences in a partitioning effect could explained instead by individual differences in working memory is addressed in Chapter 5.

Despite the possibility that the term categorical perception may not be fully warranted by what can be measured with the ASCAT, we use the term because of the task's conceptual and methodological ties to the tradition of categorical perception research. However, in the present research we mean perception in a broader sense than is usually used in the cognitive field, to encompass both automatic perceptual processes and categorical vs. continuous representation of stimuli.

Social categorization.

So far, the focus of this introduction has been on general categorization processes and the empirical phenomenon of categorical perception. These topics have been well-explored in the area of cognitive psychology and encompass many interesting and important debates about the structure of knowledge representation and the processes underlying similarity. However, the purpose of this dissertation is not to address those questions, but rather to apply insights and methods gained from research in cognitive psychology to address questions of interest to social psychologists.

The field of social categorization has a long history of doing just that. Cantor and Mischel (1979) drew parallels between object concepts and person concepts such as personality traits, exploring whether a prototype structure could account for personality judgments. Other researchers followed, drawing further comparisons between general knowledge categories and social categories (e.g., Cantor & Kihlstrom, 1981; Srull, 1981; Lingle, Altom, & Medin, 1984). However, an important distinction between social and non-social categories is that social categories have stronger behavioral and affective consequences, particularly for the social category of race.

Race is accorded special significance, particularly in America where a person's race may dictate economic (e.g., job opportunities), political (e.g., power, social policy, resource distribution), social (e.g., interpersonal friendliness), or psychological (e.g., stereotype threat) advantages or disadvantages (e.g., Bertrand & Mullainathan, 2004; Bonilla-Silva, 1999; Eberhardt, Goff, Purdie, & Davies, 2004; Kaiser, Drury, Malahy, & King, 2011; Maddux, Galinsky, Cuddy, & Polifroni, 2008; Sanchez & Bonam, 2009; Steele & Aronson, 1995). One's race can even determine the difference between life and death (Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 1998; Correll, Park, Judd, & Wittenbrink, 2002).

Social psychological research on race has tended to focus on issues such as 1) When or in what contexts do people use social categories (e.g., Devine, 1989; Dovidio, Evans, & Tyler, 1986; Lepore & Brown, 1997; Macrae & Bodenhausen, 2000)? and 2) Which social categories are activated for multiply categorizable individuals (e.g., gender, race, profession; Bodenhausen & Macrae, 1998; Crisp & Hewstone, 2007)? Additionally, researchers examined consequences of grouping people into categories – that categorization often leads to group stereotyping and prejudice (e.g., Tajfel, 1970, 1978; Tajfel, Billig, Bundy, & Flament, 1971). Surprisingly little research has, however, has examined how people mentally represent the concept of race itself. In fact, for the most part, it has been assumed, but not shown, that race perception is categorical. This oversight may be in part due to the lack of an empirical tool to measure the extent to which race is perceived categorically, a lack which we hope to remedy with the development of the ASCAT.

The present research.

This chapter has briefly reviewed prior approaches to categorization, similarity, categorical perception, and social categorization. We hope to build on these approaches and add a new way of conceptualizing social categorization by focusing on the *degree* of automatic categorical vs. continuous perception. We also hope to provide a valid and productive new tool for measuring individual differences in this construct.

In the following chapters, we provide a full description of the ASCAT, as well as statistical approaches to analyzing memory confusion data. In Studies 1, 2, and 3, we demonstrate that the ASCAT detects differences in automatic categorical perception of targets that vary continuously or discontinuously in lightness, saturation, and hue. In Studies 4 and 5, we compare targets varying in gender, age, and race. In Study 6, we find evidence that the results from a ASCAT with faces as targets reflect the processing of social information, likely involving the activation of social concepts. Study 7 extends our findings for automatic race categorization to White-to-Asian face morphs. Finally, in Study 8, we assess the reliability of individual differences in automatic social categorization.

CHAPTER 2. THE AUTOMATIC SOCIAL CATEGORIZATION TEST

In order to measure the ways in which people automatically group items into categories, we have developed the Automatic Social Categorization Test (ASCAT). In this chapter, I will describe the principles underlying the ASCAT, task specifications and procedures, statistical approaches, and general properties of the ASCAT.

Definition and operationalization of automatic categorical perception.

As described in the introduction, categorical perception occurs when two items belonging to the same category are seen as more similar to each other than two items belonging to different categories, even when the objective difference between item pairs is the same. We define automatic categorical perception as the tendency for people to perceive continuously varying targets as falling into discrete groups. We use the term "automatic" because we are interested in examining uncontrolled perceptual processes rather than explicitly held views. The indirect nature of the ASCAT becomes especially important when it is applied in social domains, such as race perception, which could be susceptible to self-presentation concerns. Our focus on confusion errors, rather than explicit categorization, is also consistent with the methodological tradition in categorical perception research, which often relies on discrimination tasks to uncover perceptual processes indicative of categorical representation.

The Automatic Social Categorization Test (ASCAT) is a measure that relies on the following principles: 1) items that are perceived as more similar will be confused in memory more often and 2) people tend to perceive items belonging to the same category as more similar to each other than items belonging to different categories.

Automatic Social Categorization Test. This task capitalizes on confusion patterns using a modified two-back memory task (Smith & Jonides, 1998). Participants view target stimuli that vary continuously along a perceivable dimension (e.g. digitally morphed faces; see Figure 2.1). Targets appear one at a time, interlaced with distractors (e.g. numbers; see Figure 2.2). For each stimulus that appears, participants are asked to identify whether it is the same as (identical to) the last item of that kind that they saw. The computer presents the first target and the first distractor (e.g. a gray block and a number) for 1000 ms each, with a blank screen appearing for 500 ms between stimuli. Each trial begins with a subsequent new stimulus displayed until the participant indicates whether the item is the same as the last item of that kind. After each trial, a blank screen appears for 500 ms. Participants first complete practice trials in which they receive feedback about whether their responses are correct or incorrect. There are 20 untimed practice trials, followed by 20 timed practice trials (this was increased to 25 untimed practice trials and 25 timed practice trials for Study 4). During the timed practice trials, if a participant's reaction time is on average greater than 2000 ms on any five consecutive trials, the participant is prompted to respond faster. After the practice trials, the response window is shortened to an average of 1500 ms on any three consecutive trials.

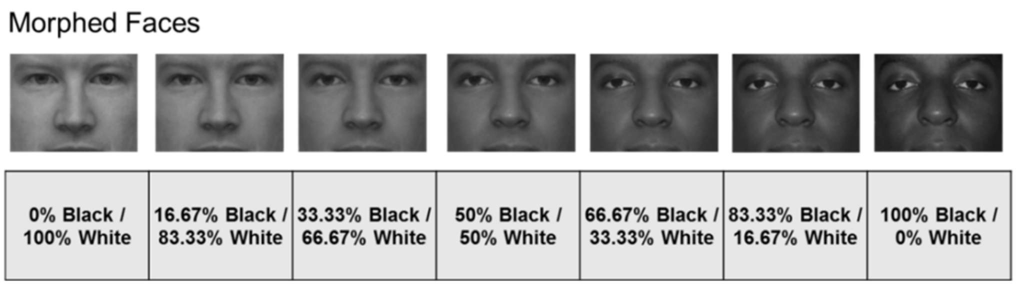


Figure 2.1. Automatic Social Categorization Test Targets. ASCAT targets consist of systematically varying stimuli, such as faces morphed between a prototypically White face and a prototypically Black face.

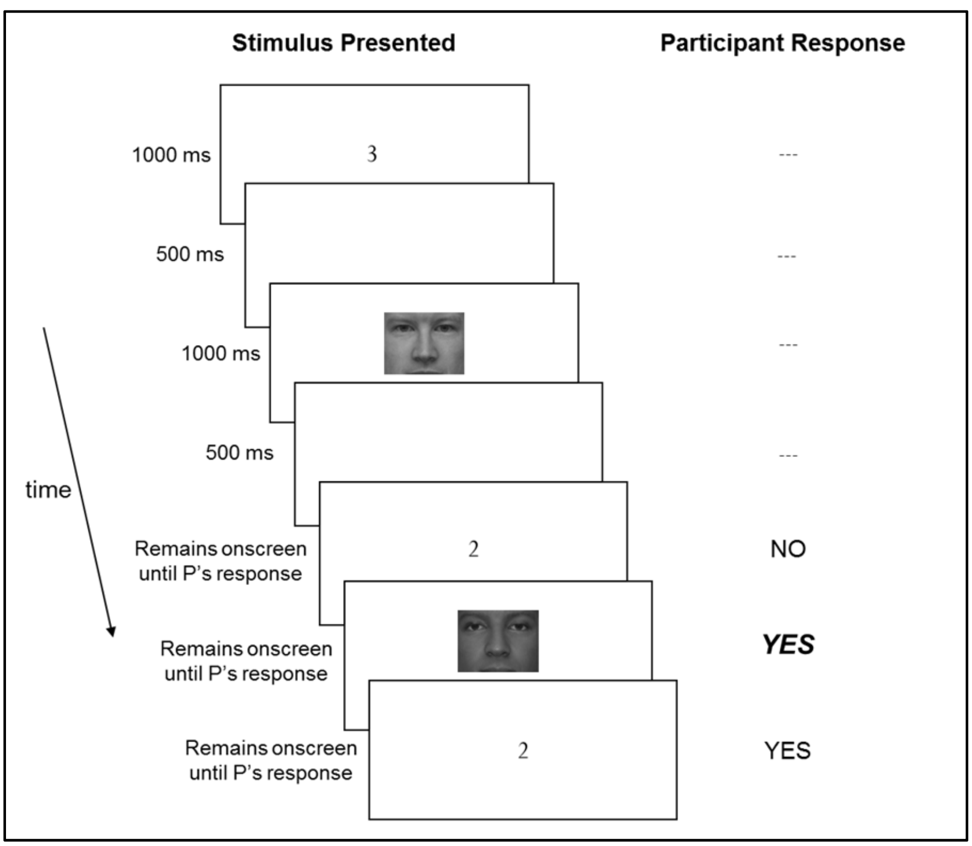


Figure 2.2. The Automatic Social Categorization Test. Stimuli are presented one at a time, with

targets (e.g. faces) interlaced with distractors (e.g. numbers). The italicized response illustrates the type of confusion error that was analyzed.

On each trial during the Automatic Social Categorization Test, an item is randomly selected from the set of targets in such a way that there is a 50% chance that the correct answer is yes on a given trial (i.e. the target is the same as the last one presented 50% of the time). The trials continue until the participant is presented with each of the possible sequences of targets (e.g. face 1 followed by face 1, face 1 followed by face 2, ... face 7 followed by face 6, face 7 followed by face 7) at least once (Studies 1, 2, 3, 5, and 6) or until each possible adjacent sequence is presented at least six times (Studies 4 and 7). The average number of trials ranged from 68.13 (Study 1, in which there were 6 targets) to 119 (Studies 4 and 7, in which there were 7 targets and the maximum number of trials was specified as 120).

The dependent variable is the extent to which participants confuse one target with another. We treat the rate of confusion as indicative of subjective similarity between the two faces. Specifically, by examining how often a given target is confused with its neighbor, one can determine perceptual categories. In theory, strictly categorical perception would have a threshold point on the continuum where all items before the threshold are seen as belonging to one category, and all items after the threshold are seen as belonging to another category. This would result in a "V" shaped graph when confusion rates are plotted against target pairs, as illustrated on the left side of Figure 2.3. By contrast, perfectly continuous perception would result in equal confusion rates for each pair of targets, as shown on the right side of Figure 2.3.

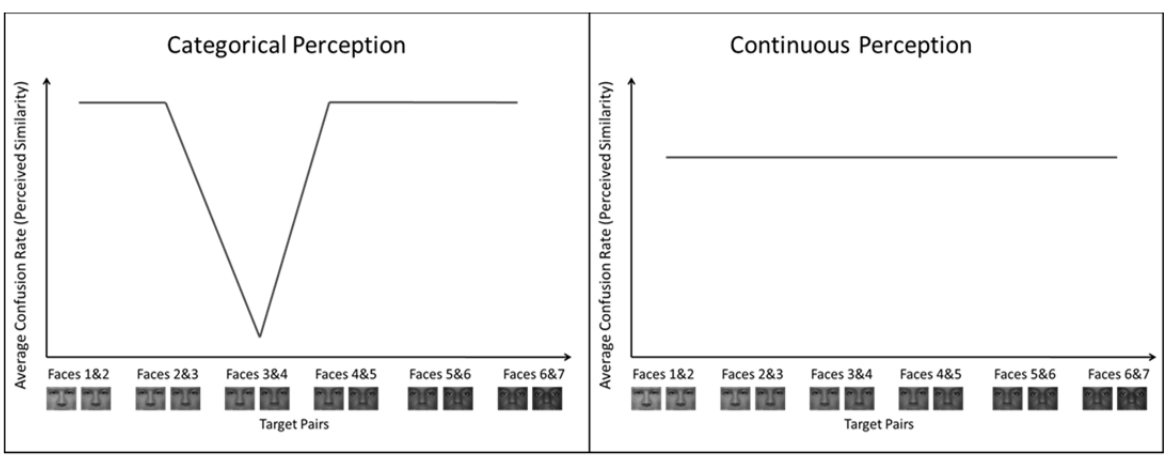


Figure 2.3. Hypothetical Confusion Rate Patterns. Hypothetical confusion rate data for categorical automatic perception and continuous automatic perception.

Although Figure 2.3 illustrates ideal confusion rate patterns for strictly categorical and strictly continuous perception, actual data patterns for the majority of participants and targets have fallen between these two extremes. And in fact, we are less interested in determining whether or not categorical perception occurs than in examining variation in *degrees* of categoricalness. I will now describe the statistical approaches we have employed to capture variation in automatic categorical perception.

Statistical approaches. In order to capture the V-shaped pattern illustrated in Figure 2.3, we have chosen to use the quadratic term from a repeated measures ANOVA for confusion rates across target pairs (see Figure 2.4 for an example of real data). If the quadratic term is statistically significant, this suggests that participants were more likely to confuse targets at the same end of a spectrum than targets in the middle. In order to test for differences between groups, we look for the interaction between the quadratic trend and the current between-subjects variable of interest (or a significant covariate, for continuous individual difference measures).[1]

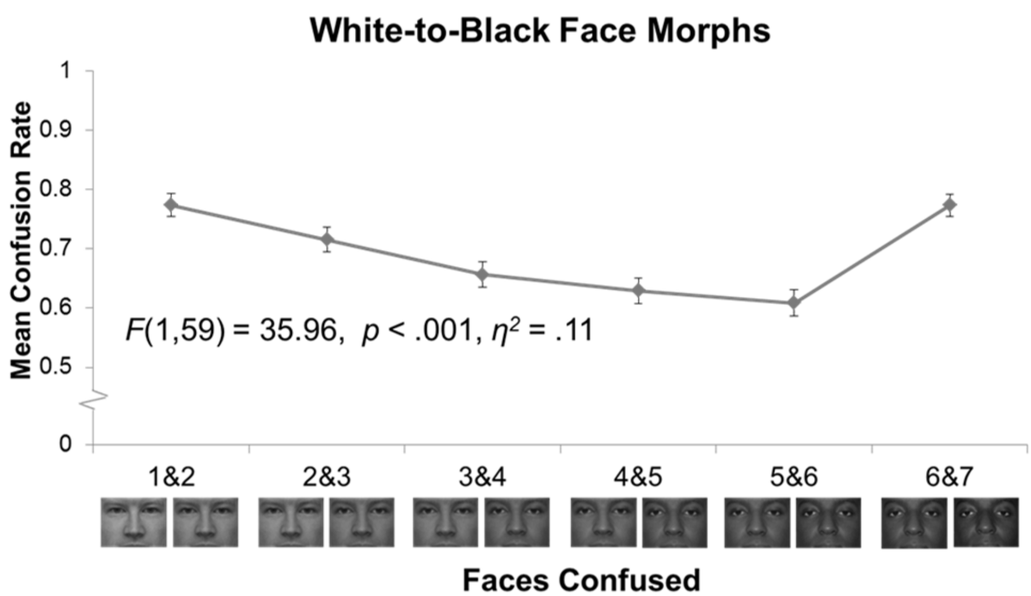


Figure 2.4. Example Confusion Rate Data Plotted Against Adjacent Face Pairs.

[1] Note that because the confusion rates are proportions, standard errors were computed in the following way. First, let p_i be the proportion correct for participant i. Then $p_i * (1 - p_i)/n$ is the unbiased estimate of the variance of the proportion for the i^{th} subject (assuming n trials for the given pair). So $\sum i\ p_i * (1 - p_i)/n$ is the unbiased estimate of the variance of the sum of the proportions, $\sum i\ p_i$, and $(\sum i\ p_i * (1 - p_i)/n\ )/N^2$ is the unbiased estimate of the variance of the mean proportion correct, $(\sum i\ p_i\ )/N$. Thus, the standard error of the mean proportion correct is $\sqrt{(\sum i\ p_i * (1 - p_i)/n\ )}/N$.

A limitation of the repeated-measures ANOVA approach to analyzing ASCAT data is that it may obscure individual-level confusion patterns. For example, if one group has a weaker quadratic trend than another (e.g., due to a shallower V-shape), this could be the result of averaging across individuals who in fact demonstrate strongly categorical confusion patterns but have different category boundaries (see Figure 2.5).

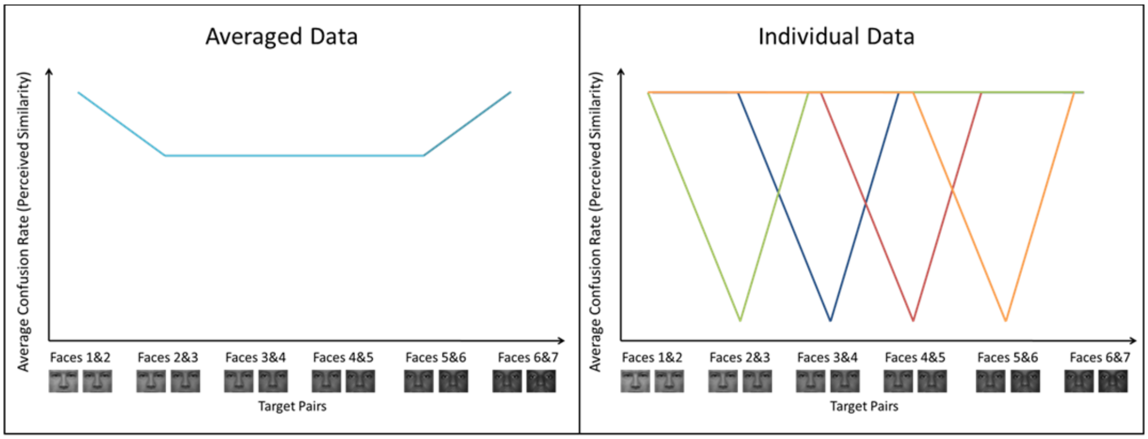


Figure 2.5. Hypothetical Data Illustrating How a Flat Line May be the Result of Averaging Across Individuals.

In order to rule out this possibility, we also use a multilevel modeling approach (using HLM), with confusion rates for adjacent pairs nested within individuals. Because the confusion rates are based on dichotomous judgments ("yes," this target is the same as the last one or "no," it's different), we use a binomial model, with the number of times a pair was confused as the outcome variable. The level 1 model is as follows (with the variable *Pair* being centered):

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(Pair) + \pi_2(Pair^2)$

The level 2 model is as follows:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

The parameter of interest here is β_{20}, the coefficient for the individual-level quadratic term. This approach allows us to determine whether this individual-level indicator of categorical perception is significant, whether there are reliable individual differences, and whether differences can be predicted by level 2 factors (e.g. participant characteristics). Note that for all analyses, the reported results are from the unit-specific model with robust standard errors.

Known limitations of statistical approaches. We recognize that a quadratic model is an imperfect approximation of the ideal shape shown in Figure 2.3. That is, it does not allow for equal confusion rates within categories (represented by the flat sections on either side of the "V" shape). However, a quadratic curve captures the general trend of interest (higher confusion rates at the ends and lower confusion rates in the middle of a spectrum) and has served as a productive dependent variable in the studies described below.

We also recognize that by focusing on a two-category model, we exclude participants who may have more than two categories for a particular set of targets. For instance, the pattern of confusion rates for someone with three categories would look more like a "W" shape, which is not captured by a quadratic equation. Future research would benefit from a model of automatic categorical perception that accounts for varying numbers of categories.

General properties of the ASCAT.

The remaining chapters in this dissertation will present empirical support for the validity and potential usefulness of the ASCAT as a measure of automatic categorical perception. However, as a first check that confusion rates from the task "work" as a measure of perceived similarity, we present a general analysis of the effect of target pair distance on confusion rates. If confusion rates do reflect similarity as we suggest, then they should decrease as pairs of targets become farther apart along a continuum.

To perform an overall test of the relation between target distance and confusion rate, we combined data sets across Studies 1, 2, 4, 5, 6, and 7 (data from Study 3 were excluded because the objective distances between pairs were not easily definable) in an three-level HLM analysis. The outcome variable was confusion rate, predicted by the group-centered objective difference between pairs.[2] The total number of observations per individual was either 15 (for Studies 1 and 2, which used sets of six stimuli) or 21 (for Studies 4-7, which used sets of seven stimuli). These observations were nested within individuals, who were in turn nested within one of nine studies and/or target types: 1) grey blocks in Study 1, 2) colors varying in saturation in Study 2, 3) White-to-Black morphed faces in Study 4, 4) scrambled faces in Study 4, 5) White-to-Black morphed faces in Study 5, 6) female-to-male morphed faces in Study 5, 7) young-to-older morphed faces in Study 5, 8) White-to-Asian faces in Study 6, 9) White-to-Black morphed faces in Study 7. To remove noise from unreliable data, participants with a total number of errors greater than two standard deviations above the mean for their study were removed from the analysis. Also, pair distance was normalized to vary between 0 and 6 for all data sets.

Level 1:

$\textit{Confusion rate} = \pi_0 + \pi_1(\textit{Pair distance}) + \text{error}$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

Level 3:

$\beta_{00} = \gamma_{000} + u_{00}$

$\beta_{10} = \gamma_{100} + u_{10}$

We estimate the coefficient π_1 as $\gamma_{100} = -0.13$, $p < .001$ (df = 8). Thus, the relation between pair difference and confusion rate was in the expected direction, with confusion rates decreasing as the difference between pairs increases.

For the majority of studies, the distractor stimuli are numerals (the numbers 1 through 7 or 1 through 6), with the exception of Study 4 (which used scrambled images as distractors). We did not have any clear hypotheses about automatic categorical perception of numbers; it is

[2] For morphed faces, differences were defined as the number of steps apart the faces were along the morphed continuum, e.g. faces 1 and 2 were 1 apart, faces 1 and 3 were 2 apart, etc. For the grey blocks and scrambles, the difference was defined in terms of lightness, computed using the pixmap image program in R (with 0 = black and 1 = white), and for saturation, the difference was based on saturation values in Paint.NET.

possible that people perceive numbers in terms of low vs. high values, or even vs. odd. Across studies, we found that memory task performance for numbers did not show evidence of categorical perception (the pattern of confusion rates was a flat line). However, the confusion rates tended to be very low (around 5% for adjacent numbers), suggesting that there may be a floor effect for numbers. Because of this, we also do not expect a clear relation between confusion rate and the difference between numbers.

Conclusion.

In sum, the ASCAT is a measure that relies on the following principles: 1) items that are perceived as more similar will be confused in memory more often and 2) people tend to perceive items belonging to the same category as more similar to each other than items belonging to different categories. By focusing on memory confusions rather than direct measures like similarity ratings or explicit classifications, we are able to obtain a measure of categorical perception that is not likely to be subject to the biases that limit the utility of self-report measures. The indirect nature of the measure is especially important for domains that may be sensitive to self-presentation bias, such as perceptions of social categories (e.g., race). We operationalize automatic social categorization as the degree to which the pattern of confusion rates across adjacent face pairs follow a quadratic trend (i.e., with higher confusion rates at either end of the stimulus continuum and lower confusion rates in the middle), with the caveat that this is an imperfect approximation of the ideal V-shaped pattern described above. We treat a stronger quadratic trend as indicative of a more discrete, either/or manner of automatically perceiving the task stimuli, focusing on relative differences in this tendency.

CHAPTER 3.
VALIDATION STUDIES WITH TARGETS VARYING IN LIGHTNESS, SATURATION, AND HUE

We propose that the ASCAT measures automatic categorical perception by detecting changes in confusion rates across continuously varying stimuli. We would like to use the task to identify differences in the degree of automatic categorical perception depending on characteristics of the perceiver and of the target domain. However, before the task can be used in these ways, we need to know that task confusion rates do, in fact, reflect patterns of perceived similarity. In Chapter 2, we presented evidence that as the objective difference between a pair of ASCAT targets increases, confusion rates between those targets decrease, suggesting that the memory task is able to detect differences in perceived similarity between targets. However, the phenomenon we hope to capture with the ASCAT is not this monotonic decrease in memory confusion as targets become more distinguishable, but a drop in confusion rates at subjective category boundaries along the target continuum.

Studies 1, 2, and 3 were designed to test whether the memory task would elicit this pattern of confusion rates when we clearly expected it to – that is, when there was an objective basis for a category boundary. We created sets of stimuli that varied systematically on a physical property of color (lightness, saturation, or hue). The stimulus sets either varied at equal intervals or contained a gap such that one adjacent pair was objectively more different than the others (see Figures 3.1, 3.4, and 3.6). In each study, participants performed the ASCAT twice, once with a continuous set (with equal intervals) and once with a non-continuous, "categorical" set (with a gap). We had three hypotheses about the confusion patterns that would result.

Hypothesis 1: Where there is an objective gap in the target set, we predicted that confusion rates would drop, producing a V-shaped graph. This is a continuation of the analysis presented in Chapter 2, extended to stimuli for which we expect a pattern of confusion rates that resembles categorical perception.

Hypothesis 2: For the continuous target sets, we predicted that confusion rates would be roughly the same across adjacent pairs. As described in Chapter 2, perfectly continuous perception should result in equivalent confusion rates (i.e. a flat line rather than a V-shaped graph). However, there is evidence that categorical perception can occur even for novel continua for which there are no pre-existing mental categories (Campanella et al., 2003; Goldstone 1994;

Levin & Angelone, 2002; Levin and Beale 2000; Livingston et al, 1998). Thus, we might expect to see a slight dip in confusion rates even for stimulus sets that do not contain psychologically meaningful boundaries. But we would expect this pattern to be less pronounced for continuous stimuli than for stimuli with a gap. For ASCAT to serve as a valid measure of CP, it is important that we see a clear difference in confusion rates between continuous and non-continuous sets of stimuli.

Hypothesis 3: Although it was not a central purpose of these studies, they allowed us to test the hypothesis that being exposed to the non-continuous stimulus sets would cause people to perceive those stimuli more categorically than the continuous sets, *beyond* the expected drop in confusion rates at the objective gap. Categorical perception has been described as a "warping" of similarities, such that within-category similarity increases and between-category similarity increases relative to the objective differences between stimuli (Harnad, 1987). The non-continuous stimuli could be considered "warped" in the sense that we increased the similarity of within-category stimuli and decreased the similarity of between-category stimuli in order to create a gap. If exposure to this type of target set encourages categorical perception, then we predict that two targets belonging to the same objective category will be confused more often than two targets that are the same distance apart but presented as part of a continuously varying target set.

Because categorical perception is well-documented in the domain of color perception (e.g. Berlin & Kay, 1969; Bornstein & Korda, 1984; Rosch, 1975), we chose to use blocks of color as targets for our initial validation studies. In order to create systematically varying stimuli that could be presented on a computer screen, we relied on the HSV model of color variation, which represents colors along three dimensions: hue, saturation, and value. In image editing software (e.g., Paint.NET), each of these dimensions can be specified precisely while holding the other two constant. The studies described in this chapter involved targets that varied in value (i.e., lightness; Study 1), saturation (Study 2), or hue (Study 3).

Study 1

In Study 1, we manipulated the spacing of memory task targets varying in lightness. We created two sets of gray blocks such that one set varied in lightness at equally-spaced intervals and the other had a gap in the middle, creating an objective category boundary. We predicted that participants' confusion rates from the ASCAT would show a quadratic trend for the categorical target set, driven by a drop in confusion rates at the gap, and that any quadratic trend for the continuous set, if present, would be much weaker. We also predicted that confusion rates for within-category pairs would be greater than confusion rates for comparable pairs in the continuous set.

Method

Participants.

Participants were 79 students who were recruited from the University of Washington Psychology Department Subject Pool and participated for extra credit. Due to a data collection error, demographic information was not collected for all participants. Information regarding sex and ethnicity was collected for 76 participants, with 38 males and 38 females. In terms of ethnicity, 48% (n = 38) were White, 28% (n = 22) were Asian, 3% (n = 2) were Latino/Hispanic, and 1% (n = 1) were African American, 10% (n = 8) selected more than one ethnicity, and 6% (n = 5) marked "Other." Information regarding age was collected for 75 participants, with age ranging from 16 to 38 (M = 19.77, SD = 2.86). There were no differences in confusion rate patterns based on gender, age, or ethnicity.

ASCAT stimuli.

We used the Paint.NET image editing software program to create two sets of six gray rectangles, varying their lightness using the "value" parameter of the HSV color model. For both sets, hue and saturation were kept constant (at 0). The continuous set varied at equal intervals along the value (lightness) parameter (20, 32, 44, 56, 68, and 80). The categorical set was constructed so that there was a greater difference in lightness between the third and fourth gray rectangles (20, 25, 30, 70, 75, and 80; see Figure 3.1). Note that the darkest gray and the lightest gray were the same, respectively, in both target sets.

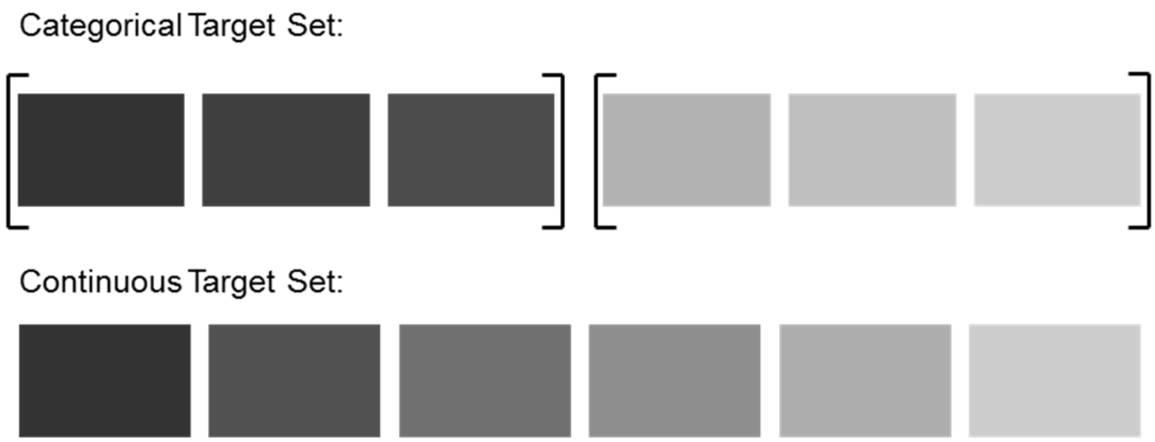


Figure 3.1. Categorical and Continuous Target Sets Varying in Lightness.

Procedure.

Participants completed two memory tasks, one with each target set. The order in which they completed the tasks was randomized. Participants also completed demographic information (age, gender, ethnicity).

Results

In order to remove random noise from participants who may not have given the task their full attention, participants with high confusion rates for targets 1 and 6 (greater than 50%) in either or both conditions were removed. The resulting data set included 56 participants.

Hypothesis 1. Confusion rates for adjacent target pairs in the categorical set showed a significant quadratic trend, $F(1,55) = 247.16$, $p < .001$, $\eta^2 = .31$, with a pronounced drop in confusion rates at the category boundary between targets 3 and 4 (see Figure 3.2).

Hypothesis 2. Confusion rates for adjacent target pairs in the continuous set also showed a significant quadratic trend, $F(1,55) = 8.77$, $p = .005$, $\eta^2 = .04$. However, this trend was significantly weaker than that for the categorical set. A 5 (adjacent target pair) x 2 (target set: categorical vs. continuous) repeated measures ANOVA revealed an interaction between target pair and target set for the quadratic term, $F(1,55) = 43.19$, $p < .001$, $\eta^2 = .14$ (see Figure 3.2).[3] There was also a main effect of target set; the categorical set produced higher overall confusion rates than the continuous set ($F(1,55) = 31.47$, $p < .001$).

[3] For quadratic interactions between two within-subjects factors, eta squared was computed by dividing the sums of squares for the quadratic interaction by the total SS for the quadratic model (omitting the SS for target type).

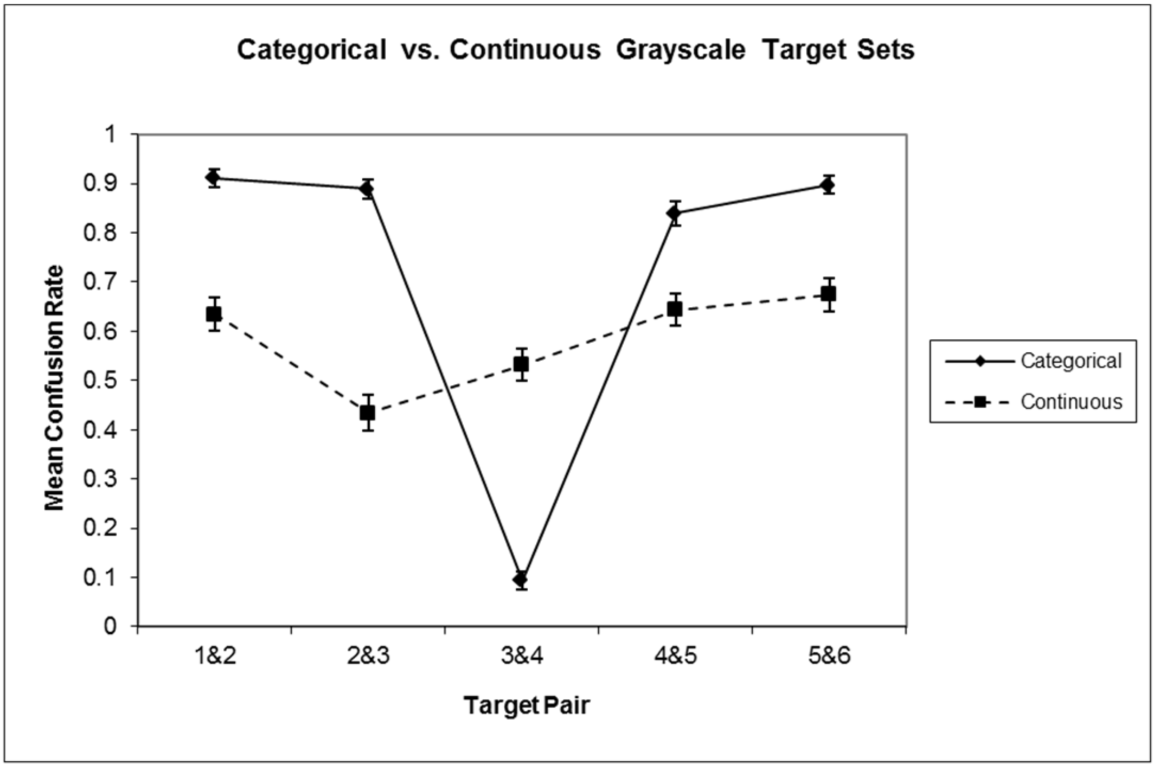


Figure 3.2. Results for Categorical and Continuous Target Sets Varying in Lightness. Categorical stimuli produced a stronger quadratic trend than continuous stimuli for gray targets varying in lightness.

To find out whether this pattern of results held true at the individual level, we conducted a multi-level analysis using a binomial model in HLM, as described in Chapter 2. To test the quadratic interaction between target type and target pair, we used the following models, with *Target Type* coded as 1 (categorical target set) or 2 (continuous target set).

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

Log[φ(1- φ)] = η

$\eta = \pi_0 + \pi_1(\textit{Target Type}) + \pi_2(\textit{Pair}) + \pi_3(\textit{Pair}^2) + \pi_4(\textit{Target Type*Pair}) + \pi_5(\textit{Target Type*Pair}^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

$\pi_3 = \beta_{30} + r_3$

$\pi_4 = \beta_{40} + r_4$

$\pi_5 = \beta_{50} + r_5$

The *target type* X *target pair*^{2} interaction was significant, $\beta_{50} = -2.22$, $p < .001$, such that the confusion rates across target pairs for the continuous target set had a significantly smaller quadratic coefficient than confusion rates for the categorical target set. We also examined each of these target sets separately, using the following model:

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(Pair) + \pi_2(Pair^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

For the categorical target set, the coefficient for the quadratic term was significant, $\beta_{20} = 2.31$, $p < .001$. For the continuous target set, the quadratic coefficient was also significant, but much smaller, $\beta_{20} = 0.15$, $p = .005$.

Hypothesis 3. To test whether merely viewing the categorical target set during the memory task produced categorical perception, we examined confusion rates for target pairs that belonged to the same category in the categorical set (e.g. targets 1 and 3), compared with confusion rates for targets that differed by a similar amount from the continuous target set (e.g. targets 1 and 2). Figure 3.3 illustrates objective differences between such pairs, along with average confusion rates. Targets 1 and 3 from the categorical set were confused at a higher rate than targets 1 and 2 from the continuous set, even though the physical differences are about the same. Likewise, targets 4 and 6 from the categorical set were confused more often than targets 5 and 6 from the continuous set. Each of these differences was significant when tested with a paired-samples t-test (*p*s < .01).

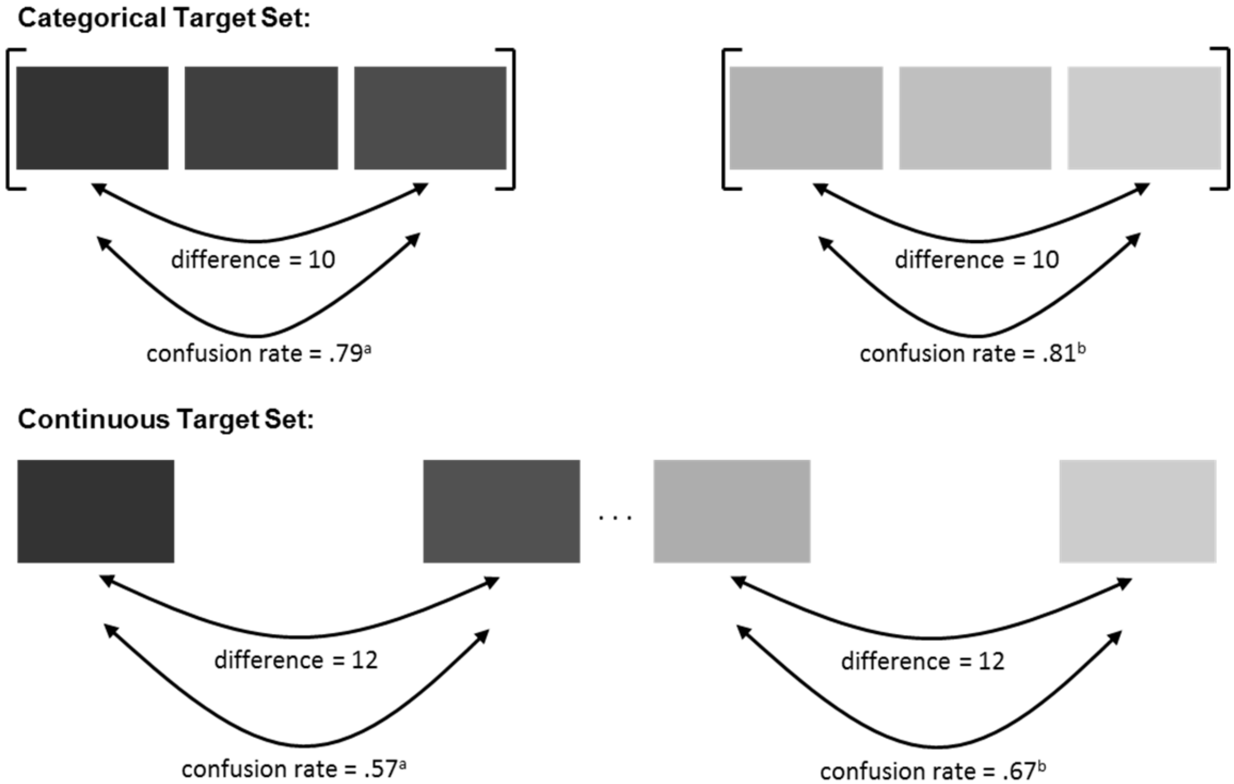


Figure 3.3. Objective Lightness Differences and Confusion Rates for Categorical and Continuous Target Sets. Confusion rates marked with the same superscript are significantly different at $p < .01$.

Study 2

The purpose of Study 2 was to replicate Study 1 with another dimension of color space, varying the saturation of blocks of color rather than the lightness of blocks of gray.

Method

Participants. Participants were 22 students who were recruited from the University of Washington Psychology Department Subject Pool and participated for extra credit. Due to a data collection error, demographic information was not collected for all participants. Information regarding sex, age, and ethnicity was collected for 19 participants, with 7 males and 12 females. In terms of ethnicity, 36% (n = 8) were White, 32% (n = 7) were Asian, 5% (n = 1) selected more than one ethnicity, and 14% (n = 3) marked "Other." Age ranged from 18 to 21 (*M* = 19.58, *SD* = 0.96). There were no differences in confusion rate patterns based on gender, age, or ethnicity.

ASCAT stimuli

Two sets of six green rectangles and two sets of magenta rectangles were created in Paint.NET by varying the saturation parameter of the HSV (hue, saturation, value) color model. For all stimulus sets, hue and value were kept constant (hue was 120 for the green sets, and 300 for the magenta sets; value was 100 for all sets). The continuous sets varied at equal intervals along the saturation parameter (40, 52, 64, 76, 88, 100). The categorical sets were constructed so that there was a greater difference in saturation between the third and fourth rectangles (40, 46, 52, 88, 94, 100; see Figure 3.4). The endpoints were the same in all stimulus sets.

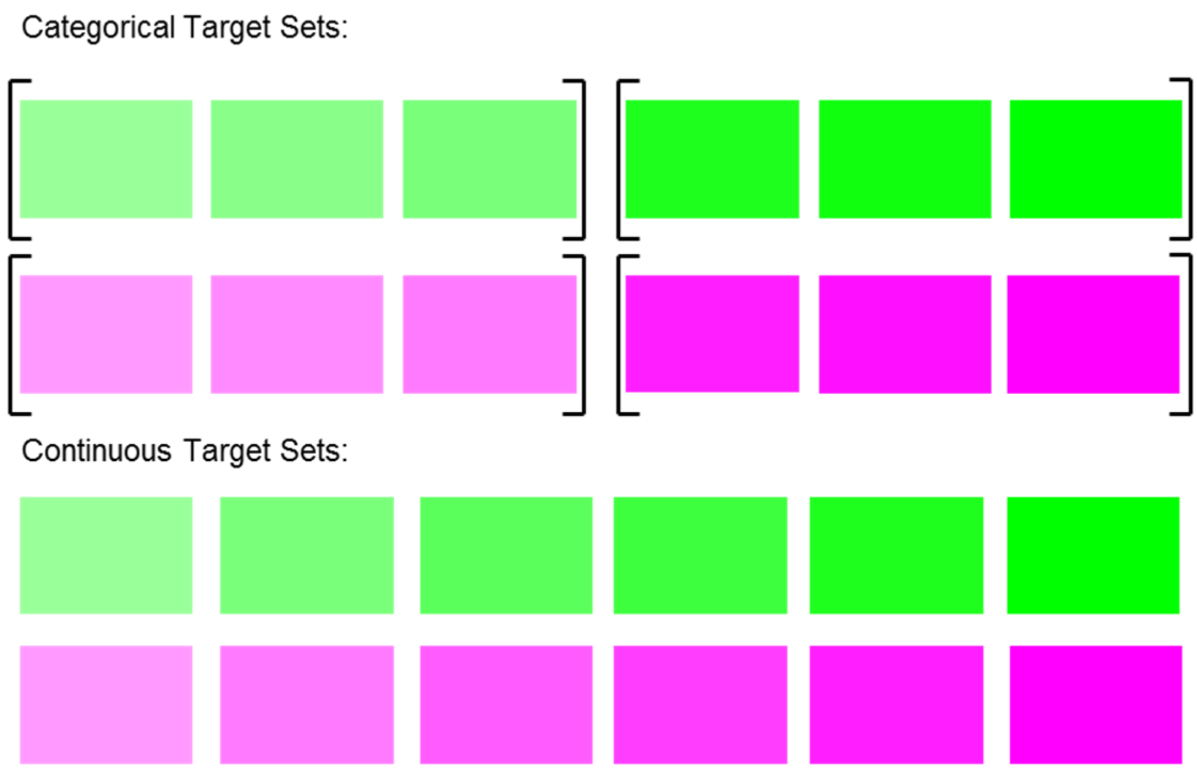


Figure 3.4. Categorical and Continuous Target Sets Varying in Saturation.

Procedure.

Participants completed two memory tasks, one with each type of stimulus set. The order in which they completed the tasks was randomized. Each participant completed the tasks with either green or magenta stimuli. Participants also completed demographic information (age, gender, ethnicity).

Results

After removing data from participants with high confusion rates between targets 1 and 6, there were 14 participants with reliable data (6 assigned to the green target sets, 8 assigned to the magenta target sets). A 5 (adjacent target pair, within-subjects) x 2 (stimulus set: categorical vs. continuous, within-subjects) x 2 (target color: green vs. magenta, between-subjects) ANOVA was conducted. There were no effects of target color, so further analyses were collapsed across green and magenta sets.

Hypothesis 1. Confusion rates for adjacent target pairs in the categorical sets showed a significant quadratic trend, $F(1,13) = 55.60$, $p < .001$, $\eta^2 = .30$, with a pronounced drop in confusion rates at the category boundary between targets 3 and 4 (see Figure 3.5).

Hypothesis 2. Confusion rates for adjacent target pairs in the continuous sets also showed a significant quadratic trend, $F(1,13) = 9.98$, $p = .008$, $\eta^2 = .09$. However, this trend was significantly weaker than that for the categorical sets. A 5 (adjacent target pair) x 2 (target set: categorical vs. continuous) repeated measures ANOVA revealed an interaction between target pair and target set for the quadratic term, $F(1,13) = 11.71$, $p = .005$, $\eta^2 = .08$ (see Figure 3.5). There was no main effect of target set.

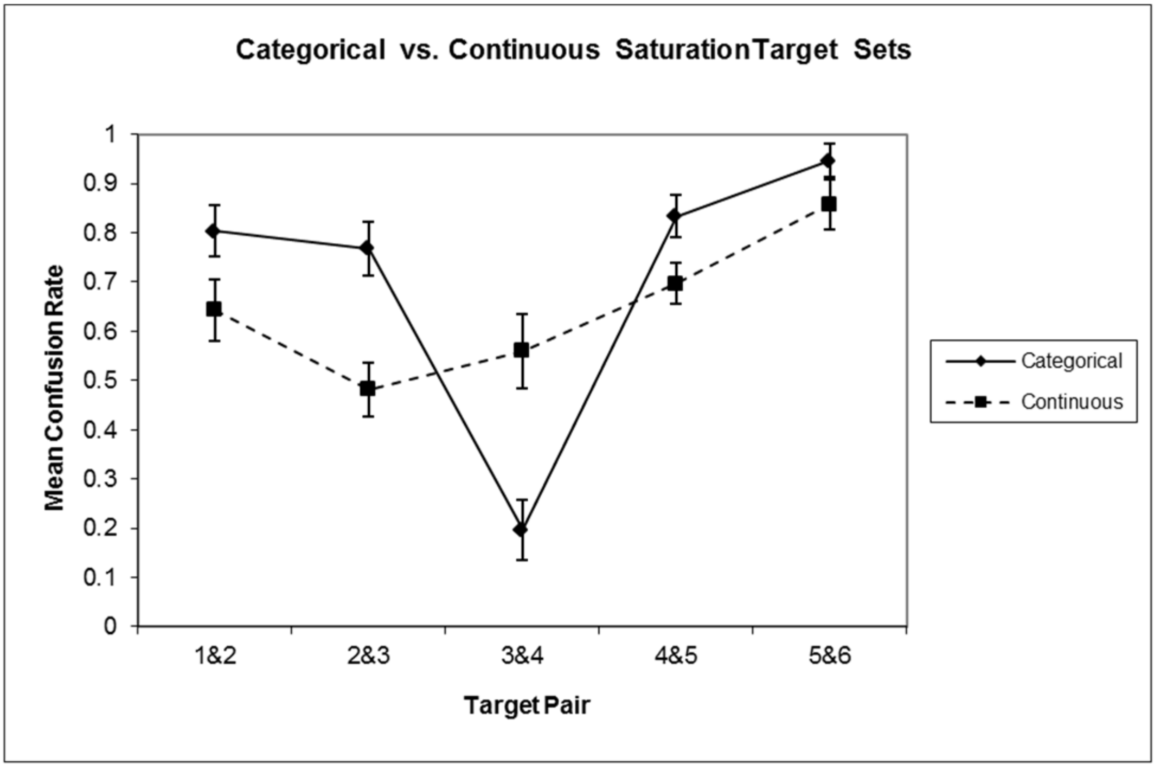


Figure 3.5. Results for Categorical and Continuous Target Sets Varying in Saturation. Categorical stimulus sets produced a stronger quadratic trend than continuous stimulus sets for targets varying in saturation.

To find out whether this pattern of results held true at the individual level, we conducted a multi-level analysis with HLM using the same models as in Study 1. Again, the *target type* X *target pair*^{2} interaction was significant, $\beta_{50} = -0.83$, $p = .001$, such that the confusion rates across target pairs for the continuous target set had a significantly smaller quadratic coefficient than confusion rates for the categorical target set. Color set was not a significant level 2 predictor for the interaction term. We also examined each of these target sets separately. For the categorical target set, the coefficient for the quadratic term was significant, $\beta_{20} = 0.85$, $p < .001$. For the continuous target set, the quadratic coefficient was also significant, but much smaller, $\beta_{20} = 0.20$, $p = .01$.

Hypothesis 3. To test whether merely viewing the categorical target sets produced categorical perception, we examined confusion rates for target pairs that belonged to the same category in the categorical sets (e.g. targets 1 and 3), compared with confusion rates for targets that differed by a similar amount from the continuous target sets (e.g. targets 1 and 2). Note that these pairs differed in saturation by the same amount. Consistent with categorical perception, the average confusion rates for targets 1 and 3 in the categorical sets (*M* = .70, *SD* = .37) were higher than that for targets 1 and 2 in the continuous sets (*M* = .64, *SD* = .36), but these differences were not statistically significant. The same held for the comparisons between targets 4 and 6 in the categorical sets (*M* = .91, *SD* = .23) and targets 5 and 6 in the continuous sets (*M* = .86, *SD* = .23). The results were similar when computed separately for green and magenta target sets.

Discussion: Studies 1 and 2

Studies 1 and 2 provide support for hypothesis 1, demonstrating that the ASCAT can detect a category boundary where it is expected to occur based on the objective distribution of targets along a continuum. This finding increases our confidence that the task will detect category boundaries for stimuli that are objectively continuous but may be automatically grouped into categories, such as faces that vary along a racial continuum but are perceived as being Black or White.

Hypothesis 2 was partially supported; we did not see a flat line for confusion rates between adjacent pairs for the continuous target sets. For targets varying at equal intervals according to lightness or saturation, there was a significant quadratic trend in confusion rates from the ASCAT. This suggests that even for continuously varying targets, participants may have automatically parsed the stimuli into categories. However, this pattern of confusion rates was much weaker (“flatter”) than for the categorical sets. This increases our confidence that the task will detect differences in automatic categorical perception for other types of targets.

In addition to the expected interaction between target pair and target type, in Study 1 there was also a main effect of target type, with higher overall confusion rates for the categorical than the continuous target set. This might be explained by the fact that within-category target pairs in the categorical set were objectively more similar than adjacent pairs in the continuous set. This occurred because the lightness values at the endpoints were the same for both sets, so creating a gap between targets 3 and 4 of the categorical set resulted in target pairs that were objectively more similar to each other. The main effect of target type on confusion rates was not replicated in Study 2, although this may be due to the small sample size.

As for hypothesis 3, in Study 1 we found support for the idea that viewing categorically organized stimuli can lead to categorical thinking beyond what would be expected based on objective differences among the targets. The data from Study 1 suggest that confusion errors during the ASCAT reflected differences in perceived similarity due not only to the actual gap in the categorical set, but also to exaggerated perceptions of similarity within these categories, compared to perceptions of the continuous set. Although this effect was not statistically significant in Study 2, results were consistent with the phenomenon.

Returning to our finding for hypothesis 2, a potential construct validity issue arises when we consider that even for stimuli that vary continuously, the ASCAT produced confusion rates that seem to indicate categorical perception. Could it be that performing the memory task with *any* set of stimuli that vary along a continuum between two endpoints will produce what we are calling a categorical pattern of confusion rates? We test this possibility in Study 3 by using a stimulus set that does not have fixed endpoints but varies along a circle.

Study 3

In Study 3 we extend our investigation to hue, which is different than lightness or saturation in that it varies along a circle, rather than along a dimension with two endpoints. If the dip in confusion rates seen in Studies 1 and 2 was merely an artifact of the target sets' having endpoints, then this pattern should disappear in Study 3, being replaced by a flat line. In Study 3, participants also completed a memory task with one of the targets omitted, creating a gap in the circular stimulus set. Removing a target also results in a set of stimuli that could be described as varying from one endpoint to another. If the confusion rates for the circular set of hues do form a flat line, then we would expect confusion rates for target pairs straddling the gaps to be lower than confusion rates for the other adjacent target pairs. We would also be able to test whether creating endpoints results in a dip in confusion rates between those endpoints.

Method

Participants.

Participants were 35 students who were recruited from the University of Washington Psychology Department Subject Pool and participated for extra credit. Due to a data collection error, demographic information was not collected for all participants. Information regarding sex and age was collected for 33 participants, with 12 males and 21 females. Age ranged from 18 to 35 (M = 19.64, SD = 2.88). Information regarding ethnicity was collected for 32 participants;

31% (n = 11) were White, 20% (n = 7) were Asian, 11% (n = 4) were Latino/Hispanic, 3% (n = 1) were African American, 20% (n = 7) selected more than one ethnicity, and 6% (n = 2) marked "Other." There were no differences in confusion rate patterns based on gender, age, or ethnicity.

ASCAT stimuli.

A set of seven targets was created in Paint.NET by varying the hue parameter of the HSV (hue, saturation, value) color model at equal intervals (hue = 0, 51, 103, 154, 206, 257, and 309). Saturation was kept constant at 100, and value (luminance) was kept constant at 24. Two additional sets of six targets each were created by removing one of the hues (see Figure 3.6). Because these two sets each have a gap in them, they contain endpoints, whereas set with seven targets does not.

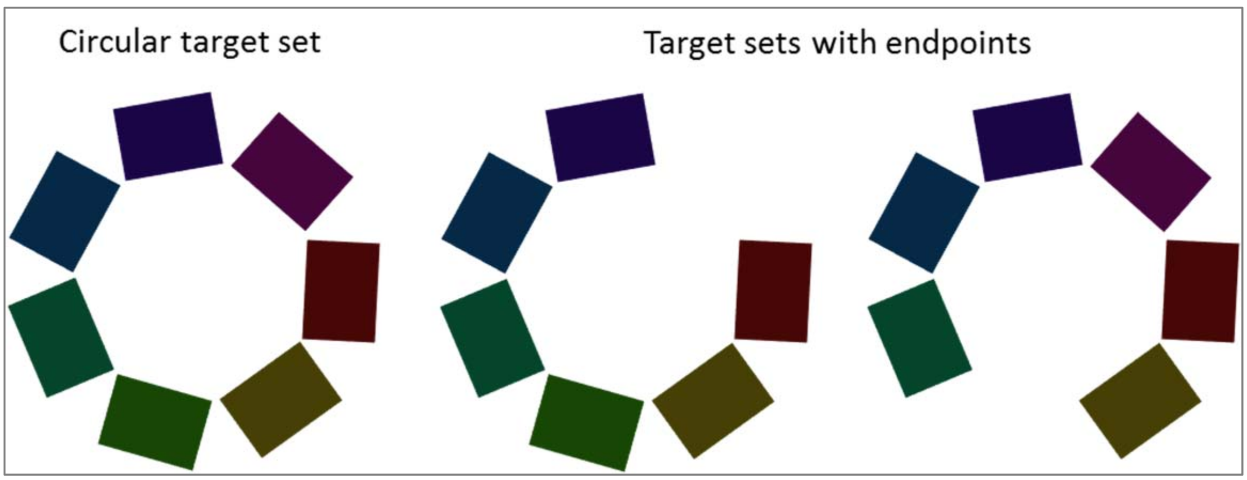


Figure 3.6. Targets Sets Varying in Hue.

Procedure.

Participants completed two memory tasks, one with the circular target set, and one with a target set with endpoints. The order in which participants completed the tasks, as well as which target set with endpoints they were assigned, were randomized. Participants also completed demographic information (age, gender, ethnicity).

Results

Due to a program error, data for the continuous target set were not collected for 5 participants. For the remaining 30 participants, Figure 3.7 shows their average confusion rates across adjacent target pairs.

Hypothesis 1. Because the pattern of confusion rates for the circular target set do not form a flat line (see Figure 3.7), it does not make sense to look for a quadratic trend to indicate a drop in confusion rates at gaps in the continuum, the way we did for Studies 1 and 2.

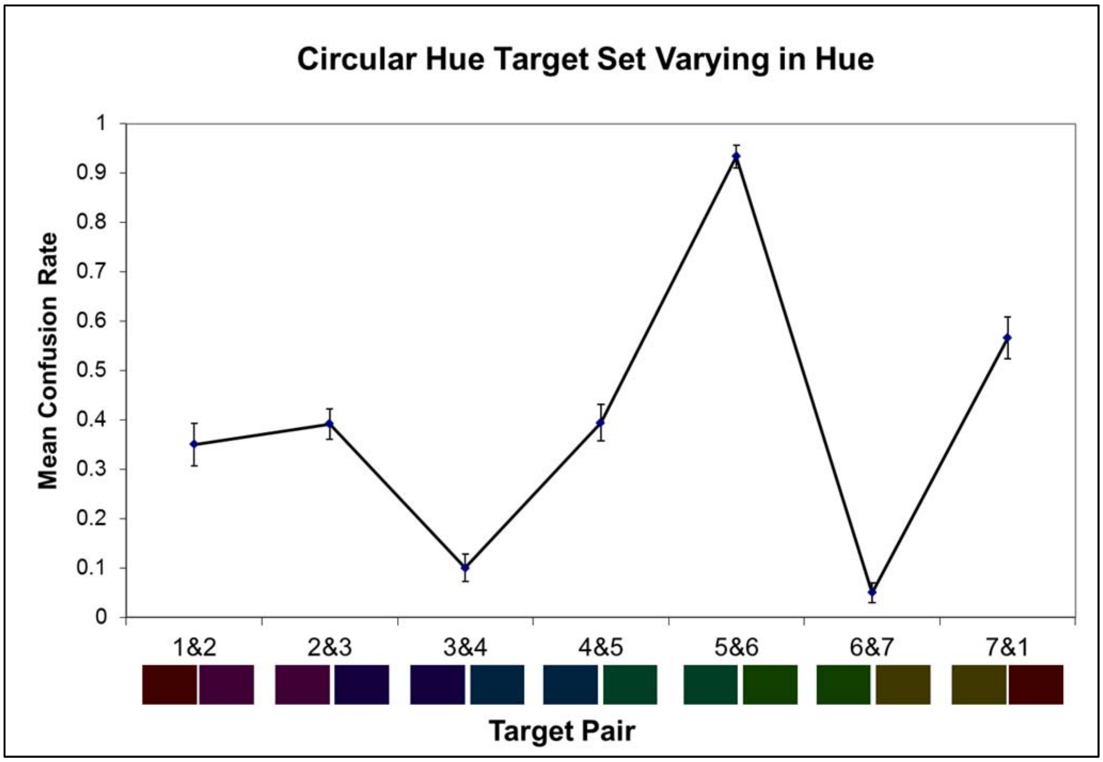


Figure 3.7. Confusion Rates for Targets Varying Continuously in Hue.

Hypothesis 2. A one-way ANOVA for target pairs allows us to reject the null hypothesis that confusion rates are the same across target pairs, $F(6, 174) = 30.63$, $p < .001$. In particular, there is a spike in confusion rates between targets 5 and 6. Pairwise comparisons with Bonferroni's adjustment for multiple comparisons reveal that this confusion rate is significantly higher than confusion rates for all other adjacent pairs (ps < .001). There are also significant drops in the average confusion rates between pairs 3 and 4, and between pairs 6 and 7. These confusion rates are significantly different than those for their neighboring pairs (ps < .05 for targets 3 & 4, ps < .001 for targets 6 & 7).

To explore this pattern further, we performed multidimensional scaling on the average confusion rates for each adjacent pair, limiting the solution to two dimensions. The targets in the resulting graph form a rough circle, with targets that were confused more often clustered together creating gaps between targets that were confused less often (see Figure 3.8).

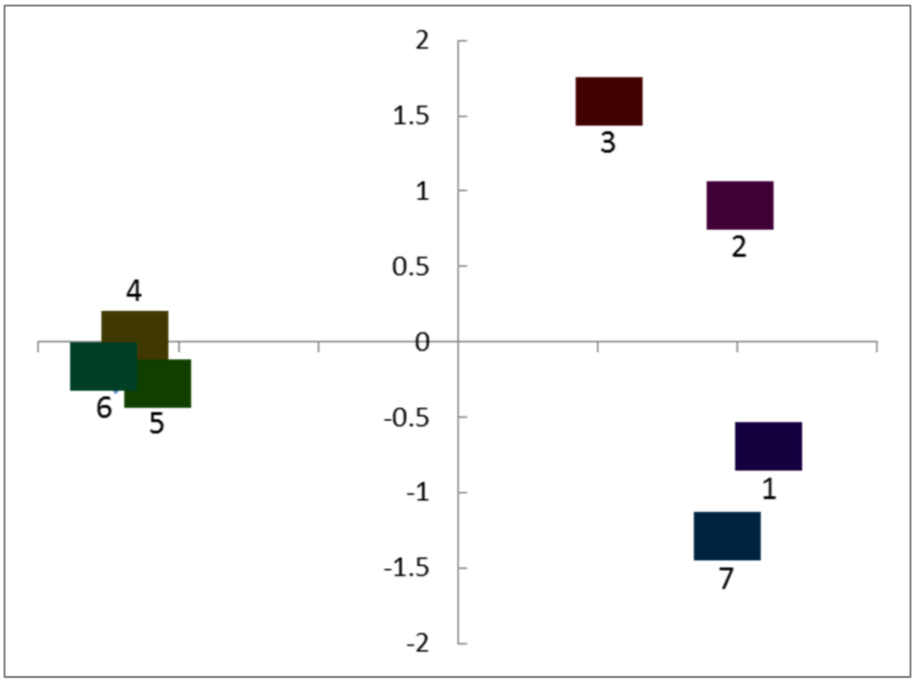


Figure 3.8. Two-Dimensional MDS Solution for Complete Target Set Varying in Hue.

Hypothesis 3. In Study 3, we see little evidence that performing the ASCAT with a target set that is missing a color (creating a gap relative to the circular target set) causes participants to confuse within-category targets more often than they do for the complete target set. Figure 3.9 shows the patterns of confusion rates when one of the targets is removed. Fifteen participants completed the memory task with target 2 omitted, and 17 participants completed the memory task with target 5 omitted. These confusion rates are plotted along with the corresponding confusion rates from the complete target set. The pattern of errors did not change when a target was omitted, with the exception of targets 7 and 1 in the set with target 5 omitted; a paired-samples t-test revealed a significant difference between the confusion rate for 7 and 1 in the complete ($M = .56$, $SD = 0.38$) and incomplete ($M = .34$, $SD = 0.40$) conditions, $t(16) = 2.37$, $p = .03$.

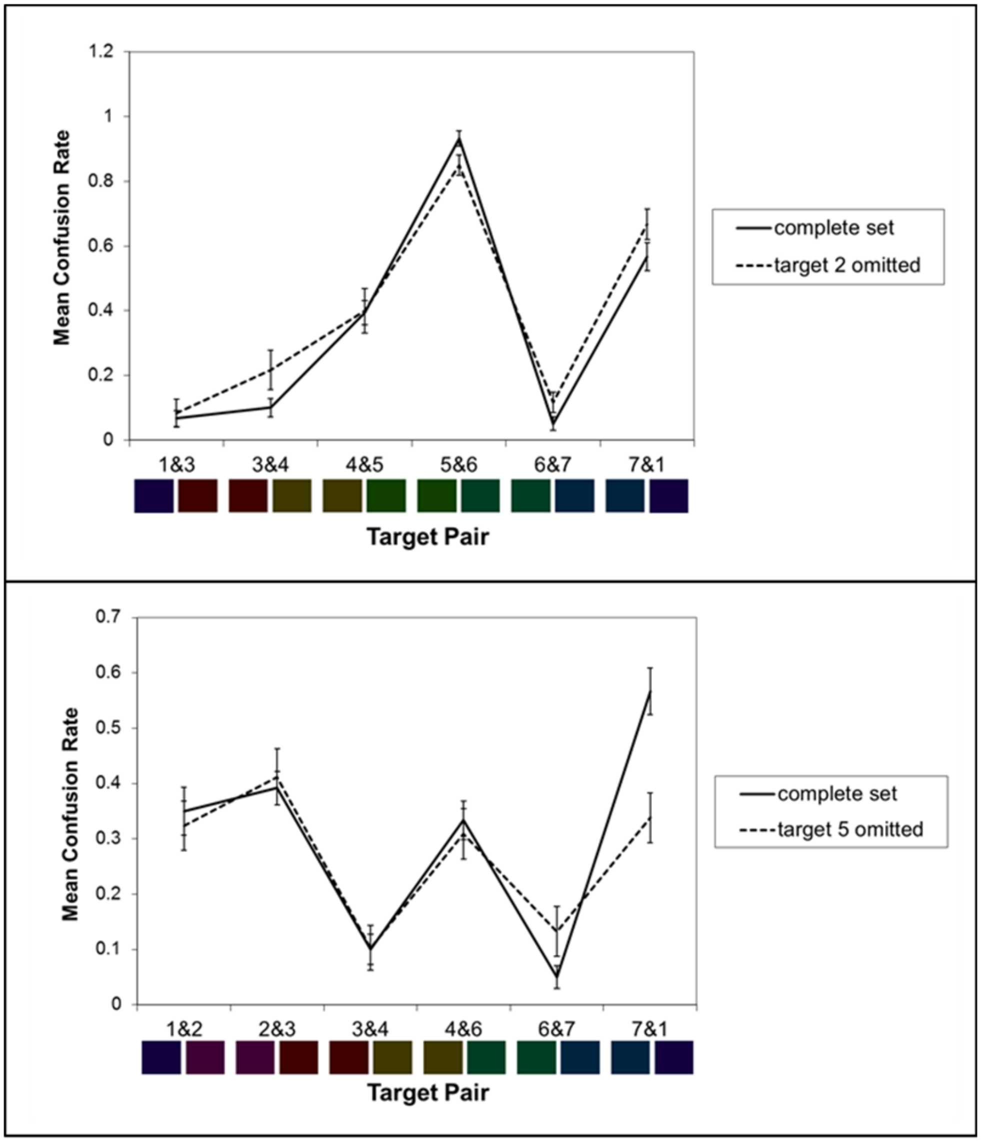


Figure 3.9. Confusion Rates for Complete vs. Incomplete Hue Sets.

Discussion

The pattern of confusion rates for stimuli varying in hue was different than the patterns for both categorical and continuous sets from Studies 1 and 2. If the V-shaped results from the first two studies were due to the fact that the stimuli varied between two endpoints, then we would have expected confusion rates for hue to form a flat line, since hue varies along a circle without endpoints. However, we rejected the hypothesis that there are no differences among confusion rates for adjacent target pairs, and the pattern of confusion rates for targets varying in hue seem to align with natural color naming categories. Although the targets vary at equal intervals along the hue dimension of the HSV color space, the two colors that were confused the most often (targets 5 and 6) might be considered two shades of green. Also, the lowest confusion rates were at locations along the continuum that could be considered boundaries between red and yellow (targets 3 and 4) and between green and blue (targets 6 and 7). To further examine this interpretation of the data, it would have been informative to ask participants to explicitly identify the targets' colors. Unfortunately, we did not collect that information from participants, so interpretation must remain speculative. However, if we interpret these apparent category boundaries as being consistent with color names, then our findings are consistent with past research on color perception showing that color naming behavior is a good predictor of the ability to discriminate between color stimuli (Bornstein & Korda, 1984; Rosch, 1975; Winawer, et al., 2007).

In Study 3, we were not able to induce categorical perception by showing a "gapped" version of the target set. Since the confusion rates were not equivalent across adjacent target pairs in the complete, circular target set, it was not clear what to expect for the incomplete target sets. Removing a color from the spectrum shown to participants might have resulted in a shift in the color categories they perceived (for example, the category "green" may have been less salient in the non-continuous set that omitted target 5). However, the statistically significant difference that did occur is difficult to make sense of. It may that since basic color terms are learned early in life, it is difficult to influence categorical perception of hue, compared to lightness or saturation.

General Discussion

Studies 1 and 2 show that the ASCAT elicits a drop in confusion rates at an objective gap in target stimuli, and that this V-shaped pattern is stronger for these categorical sets than for continuously varying stimuli. Individual-level analyses using HLM support the hypothesis that these patterns holds for individual participants, rather than being a function of averaging across participants with varying category thresholds. However, given that there are few observations (2) per target pair per participant, future studies will collect more data so that individual-level analyses are more reliable. Study 3 provides evidence that the quadratic trend in confusion rates is not simply due to the nature of the task – that is, we do not see this pattern for any set of stimuli varying between two endpoints. In fact, confusion rates for targets varying in hue appear to reflect color naming categories, increasing our confidence that the task taps into genuine psychological categories.

We interpret the high confusion rates for within-category target pairs as reflecting perceived similarity between those pairs. However, an alternative explanation is that these targets were not physically distinguishable from one another. To test this possibility, we presented seven pilot subjects with each stimulus set from this chapter and asked them to arrange the stimuli in order (from light to dark, most saturated to least saturated, or in order based on hue). Everyone was able to sort the stimuli from Study 1 (varying in lightness) with 100% accuracy. Five out of seven subjects were able to sort the stimuli from Study 3 (varying in hue) with 100% accuracy, and the remaining two subjects made only one error each. However, subjects had difficulty sorting the stimuli from Study 2 (varying in saturation), particularly the most saturated stimuli from the categorical target sets.[4] Thus, results for Study 2 may be partially explained by an inability to distinguish among these targets.

These studies also address the question of what “continuous perception” looks like when measured by the ASCAT, since the continuous stimulus sets used in Studies 1 and 2 are unlikely to contain pre-existing perceptual category boundaries. Since the ASCAT is a new measure, it will be informative to treat the results for the continuous stimuli in those studies as a baseline against which other results may be compared. For example, if we found a quadratic trend in confusion rates for faces varying in race (e.g., morphed from a Black face to a White face) that was only as strong as that for a continuum of gray blocks, then we would be hesitant to claim evidence for automatic race categorization. Chapters 4 and 5 present data suggesting that automatic categorization of continuously varying social stimuli (i.e., faces) is, in fact, stronger than for non-social stimuli. In Chapter 6, we will provide a summary of effect sizes for ASCAT results across studies to illustrate the range and relative strength of automatic categorization for different types of targets.

[4] With regard to the categorical sets, only one subject sorted the green targets correctly, and two out of five subjects sorted the magenta targets correctly; all errors were made within the more saturated end of these spectrums

CHAPTER 4.
AUTOMATIC SOCIAL CATEGORIZATION OF FACES THAT VARY IN AGE, GENDER, OR RACE

In Chapter 3, we showed that the ASCAT is sensitive to variation in the objective similarity of targets varying in lightness, saturation, and hue; when there is a gap in the target set, confusion rates drop. If this pattern of confusion rates is observed even though there is no gap in the objective similarity, then it suggests that the psychological processes, such as perception and mental representations, are more categorical in nature. The V-shaped pattern did not occur for targets varying in hue, but the pattern of confusion rates was consistent with color categories, suggesting that ASCAT errors reflect meaningful psychological labels. To further test the discriminant validity of the ASCAT and to examine its utility as a measure of social categorization, we extend our investigation to faces varying in age, gender, and race.

Just as colors vary along various physical dimensions (e.g., hue, saturation, value), faces also vary continuously along dimensions such as eye shape, hair color, and skin tone. However, people quickly and automatically identify others in terms of social dimensions such as age, gender, and race (Brewer, 1988; Bruner, 1957; Fiske, 1998; Fiske & Neuberg, 1990; Kite, Deaux, & Miele, 1991; Hewstone, Hantzi, & Johnston, 1991; Stangor, Lynch, Duan, & Glass, 1992). In fact, race is noticed as early as approximately 120 ms after encountering a face (Ito & Urland, 2003; 2005), followed quickly by gender, about 50 ms later (Ito & Urland,). Research in this chapter examined the extent to which social perception, with regard to age, gender, and race, is categorical, rather than continuous, using the ASCAT.

Although the idea of objective variation in social dimensions is more complex than in physical categories like colors, in order to explore categorical vs. continuous perception of these dimensions, we approximate objective variation by using a face morphing procedure. For instance, to produce target faces varying in race, we chose two faces rated as being highly stereotypic of their respective racial category and created morphs with varying percentages of each face (see further details and examples below).

We first seek evidence for discriminant validity of ASCAT by asking the question, do all social attributes produce similarly categorical memory confusion patterns? We expected that gender will produce the most pronounced categorical pattern, and age, the least pronounced one, because gender is typically defined in terms of two discrete categories, male and female, whereas age can vary along a continuum from young to old. Next, we then apply the ASCAT to race, to find out how it compares to gender and age in terms of the strength of the categorical memory confusion pattern. Human physical features (e.g., skin tone) typically associated with specific racial groups vary continuously and can combine with other features in infinite combinations (e.g., blue eyes with light skin, blue eyes with dark skin). Yet, in everyday language people typically refer to discrete "races" (e.g., Black, White). Do people, in fact, *perceive* race in terms of discrete categories, or do their mental representations of people reflect the continuous variation of these physical traits?

Past Race Research, Viewed through the Lens of Continuous vs. Categorical Perception

Past research on race perception is consistent with, but does not provide a critical test for the hypothesis that race is perceived categorically, as opposed to continuously. Typical paradigms in social psychology examine category activation using the ends of the racial continuum, such as prototypically White or Black faces (Taylor, Fiske, Etcoff, & Ruderman, 1978; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald & Banaji, 1995; Macrae & Bodenhausen, 2000; Martin & Macrae, 2007; Crisp, Hewstone, & Rubin, 2001; Crisp & Hewstone, 2007; Hall & Crisp, 2005). For example, prototypically Black faces automatically activate more negative stereotypes than prototypically White faces (e.g., Fazio et al., 1995). But such a phenomenon can occur even if race perception is continuous, as long as people associate greater negative stereotypes with the "Blacker" end of a continuum ranging from Black to White. Another example comes from the Who-Said-What task (Taylor et al., 1978), in which two Black individuals are more likely to be confused than a Black individual and a White individual. However, such interchangeability can easily occur even when race perception is continuous, because two Black individuals are likely to be much more closely located on a Black-White continuum than a White individual and a Black individual who are on the opposite ends of the continuum. Without examining perceptions of a spectrum of faces that vary along the continuum from one race to another, it is impossible to test for categorical race perception.

In addition, the tasks used in past research may have primed categorical perception. In some of the studies that included ambiguous-race targets, participants were presented with race category labels (e.g., "Black", "White", "Not Black"; Freeman, Pauker, Apfelbaum, & Ambady, 2010; Peery & Bodenhausen, 2008). Such practices may lead some participants to view the targets as consistent with these racial categories (Eberhardt, Dasgupta, & Banaszynski, 2003), whether they would have spontaneously seen them that way or not. Additionally, providing category labels is restrictive. Using two either/or categories (e.g., Black, White) removes the possibility that ambiguous-race faces may fall into some perceived third category (e.g., multiracial, or see research regarding an 'emergent third race'; MacLin & Maclin, 2011; Maclin, Peterson, Hashman, & Flach, 2009) or along a continuum between the two categories.

Despite these limitations, past research using ambiguous-race faces suggests that race is likely to be perceived categorically. For example, Peery and Bodenhausen (2008) showed that even when people know about a target person's multiracial heritage, when asked to quickly categorize a person's race, they will indicate that the individual is monoracial at least 50% of the time. However, even these studies do not include faces ranging along a full continuum of race-relevant features (some studies use only faces composed of 50% of one race and 50% of another; Maclin & Malpass, 2001; Pauker & Ambady, 2009; Willadsen-Jensen & Ito, 2006). By limiting the range of stimuli presented, researchers may obscure category boundaries that do exist (e.g. if faces are viewed according to the "one drop" rule, then the boundary would be closer to the "White" end of the spectrum).

In sum, research on social categories has not adequately addressed the question, is there evidence that race perception is categorical or continuous? One rare exception is a study in cognitive psychology by Levin & Angelone (2002), who found that race perception is categorical using discrimination and classification tasks. Additionally, MacLin and colleagues (2009) found evidence for categorical race perception using a psychophysics approach with morphed White-to-Black continuums. However, their method involved asking participants to explicitly indicate whether a face belonged to a given race category (e.g., "Is this face Caucasian?"). Neither study examined how race perception varies across perceivers without asking the participants to make explicit race judgments, nor did they investigate the social implications of this question.

The research in this chapter builds on these social and cognitive psychology perspectives by systematically examining automatic race categorization using the ASCAT. First, in Study 4 we compare categorization of face morphs varying in gender or age, social attributes that we expect to be perceived more (gender) or less (age) categorically. Next, in Study 5 we show how race categorization compares to gender and age categorization for White-to-Black face morphs. In Study 6 we verify that performance on the ASCAT reflects *social* information processing rather than simply reflecting physical characteristics of the images such as their luminosity, by comparing ASCAT performance for White-to-Black morphed faces with performance for targets from which social information has been removed (i.e., scrambled-pixel faces). To the extent that participants view race categorically and that ASCAT performance reflects categorical perception of faces, the quadratic trend in confusion rates should be noticeably larger for White-to-Black morphed faces than for scrambled-pixel images that can no longer be identified as faces. Finally, in Study 7 we extend our investigation of race categorization to White-to-Asian face morphs.

Study 4

For Studies 4 and 5, we collected data from a community sample over the course of six study sessions during the 2008 US Presidential Election. Due to the fact that one of the candidates, Barack Obama, is biracial, the aim of the longitudinal study was to track changes in categorical race perception over the course of the campaign; we did not find significant changes over time. However, at each session, participants completed the ASCAT with both a White-to-Black face morph set and a "control" morph set with faces that varied in age or gender. Here we report the data for the female-to-male and young-to-old face morphs to examine differences between automatic social categorization of age and gender. We predict that confusion rates will reflect a stronger categorical trend for gender than for age.

Method

Participants.

Eighty-nine participants (M_{age} = 30.51 years, SD_{age} =10.73; 23 M, 58 F; 53.1% White, 29.6% Asian, 17.2% Other; 8 participants did not provide this information) from the Seattle area were recruited via flyers and online advertisements for a longitudinal study of social relations. Participants were paid $10 for a preliminary questionnaire session, $18 for each study session, and a $50 bonus if they attended the questionnaire session and all five subsequent study sessions. Seventy-seven subjects continued through the five sessions. Six months later, participants were invited to complete a sixth follow-up session, for which they were compensated $45. Sixty participants returned for the follow-up session. Due to subject error, missed sessions, and computer malfunction, our final sample consisted of sixty participants (16 M, 43 F; 53.3% White, 21.7% Asian, 23.3% Other; 1 participant did not provide this information) with useable data from at least four sessions. There were no differences in confusion rate patterns based on participants' age, gender or ethnicity.

Procedure.

All study sessions were conducted in the Psychology Department of the University of Washington. Participants were run in groups of up to eight people. Every participant completed one questionnaire session which assessed demographic information and other measures which will not be the focus of this study. In the first study session, participants were randomly assigned to one of four young-to-old face morph sets or one of four female-to-male face morph sets. Participants were also randomly assigned to one of nine groups of White-to-Black face morph sets (to be discussed in Study 5). For all five sessions, each participant completed the ASCAT twice, once with each assigned stimulus set, as well as an evaluative priming task (Fazio, Jackson, Dunton, & Williams, 1995).[5] At the sixth follow-up session, participants completed three memory tasks, two with their assigned face morph sets, as well as a young-to-old set or a female-to-male set, whichever condition they had not yet completed. The order in which the memory tasks were completed was randomized at the first session but kept the same across sessions. At the sixth session, the new young-to-old or female-to-male face morph set was completed last.

ASCAT Stimuli.

To create female-to-male and young-to-old face morph sets, we chose four pairs of high-resolution infant and adult (or female and male) faces that were facing forward. These faces were morphed together to form a continuum ranging from female (young) to male (old) faces. This resulted in a seven-face continuum made from each pair of faces: face 1 (100% female), face 2 (83.3% female/16.7% male), face 3 (66.7% female/33.3% male), face 4 (50% female/50% male), face 5 (66.7% female/33.3% male), face 6 (83.3% female/16.7% male), and face 7 (100% male). The faces were cropped from the eyebrows to the upper lip (see Figure 4.1 for sample morphed stimuli).

[5] The evaluative priming task was included to investigate possible changes in racial bias over time, as well as the relation between automatic categorical race perception and bias; however, these questions are not the focus of the current study.

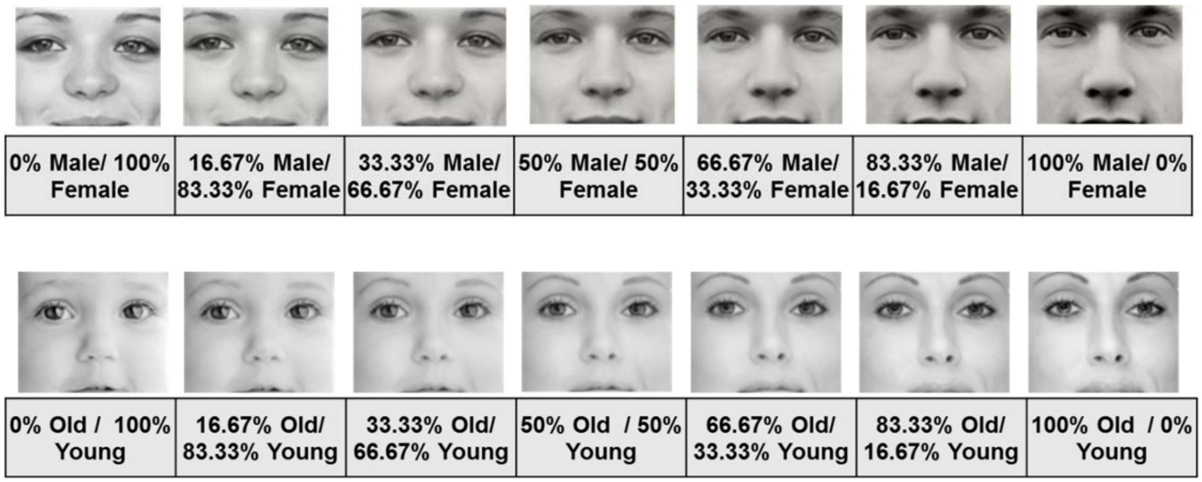


Figure 4.1. Sample Sets of Face Stimuli Morphed from Female to Male and Young to Old Faces.

Pilot Study: Face stimuli ratings.

To find out whether the faces within each stimulus set were approximately equal in perceived attractiveness, age (for the White-to-Black and female-to-male sets), baby-facedness (for the White-to-Black and female-to-male sets), and realism, a separate group of 63 Psychology Subject Pool participants ((M_{age} = 19.40 years, SD_{age} = 1.29; 21 M, 42 F; 54.0% Asian, 25.4% White, 20.6% Other) rated the faces on these characteristics. Each participant was assigned to rate one type of morphed face set (White-to-Black, female-to-male, or young-to-old). The order in which faces were presented was randomized. Ratings for the White-to-Black faces are reported in Study 5.

Attractiveness. All faces were rated on attractiveness ("How attractive does this face look?") on a scale of 1 to 7 (1 = Extremely, 4 = Moderately, 7 = Not at all).

For female-to-male faces, a 7 (faces) by 4 (face sets) repeated measures ANOVA did not show a significant effect of face on attractiveness ratings, $F(6, 114) = 0.95$, $p = .47$.

For young-to-old faces, a 7 (faces) by 4 (face sets) repeated measures ANOVA showed that the faces did significantly differ in perceived attractiveness, $F(6, 120) = 5.53$, $p < .001$. For three of the face sets, attractiveness ratings increased as the faces approached the older end of the face morphs. For the fourth face set, attractiveness ratings did not differ across faces.

Age. All faces were rated on age ("What age does this person look?"). For the female-to-male faces, the rating scale was as follows: 1 = 16-20 years, 2 = 21-25 years, 3 = 26-30 years, 4 = 31-35 years, 5 = 36-40 years, 6 = 41-45 years, 7 = 46-50 years, 8 = 51-55 years, 9 = 56-60 years, 10 = 60+ years. A 7 (faces) by 4 (face sets) repeated measures ANOVA also showed a significant effect of face on age ratings, $F(6,114) = 11.43$, $p < .001$. Specifically, there was a significant linear trend, $F(1,19) = 19.25$, $p < .001$, such that, for three of the face sets, faces with a higher proportion of the male face were rated as appearing older. There was also a significant quadratic trend, $F(1,19) = 18.53$, $p < .001$, which appears to be driven by the fourth face set, for which perceived age decreased as the faces became more male, until the 66.67% male face, after which perceived age increased.

For the young-to-old faces, the rating scale was adjusted as follows: 1 = 1 to 6 mos, 2 = 7 to 11 mos, 3 = 1 to 5 years, 4 = 6 to 10 years, 5 = 11 to 15 years, 6 = 16 to 20 years, 7 = 21 to 25 years, 8 = 26 to 30 years, 9 = 31 to 35 years, 10 = 36 to 40 years, 11 = 41 to 45 years, 12 = 45+ years. As expected, the 7 (face) by 4 (face set) repeated measures ANOVA showed a significant linear trend, $F(1,20) = 799.88$, $p < .001$, such that perceived age increased as the proportion of the older face increased.

Babyfacedness. Female-to-male faces were rated on babyfacedness ("How babyfaced does this person look?") on a scale of 1 to 7 (1 = Extremely, 4 = Moderately, 7 = Not at all). For the female-to-male faces, there was a significant effect of face on babyfacedness ratings, $F(6,114) = 14.47$, $p < .001$. Specifically, there was a linear trend, $F(1,19) = 21.60$, $p < .001$, such that, for three of the face sets, the male faces were rated as less babyfaced ($M = 5.22$, $SD = 0.99$) than the female faces ($M = 3.73$, $SD = 0.91$). The remaining face set showed a quadratic trend, with the 50% female/50% male face being rated as the most babyfaced.

Realism. All faces were rated on realism ("How realistic does this face look?") on a scale of 1 to 7 (1 = Extremely, 4 = Moderately, 7 = Not at all). For the female-to-male faces, there was no significant effect of face on realism ratings, $F(6,114) = 1.20$, $p = .31$. However, for the young-to-old face morphs, faces in the middle of the morphed spectrum were rated as less realistic than faces on either end, as demonstrated by a significant quadratic trend, $F(1,20) = 17.87$, $p < .001$.

Results

Female-to-male face morphs.

After removing data from participants with high confusion rates between faces 1 and 7, there were 25 participants with reliable data from at least four sessions. A 6-level (adjacent target pairs) repeated measures ANOVA revealed a significant quadratic trend in confusion rates, $F(1,24) = 58.30$, $p < .001$, $\eta^2 = .30$, as well as a small order 4 trend, $F(1,24) = 5.01$, $p = .04$, $\eta^2 = .03$ (see Figure 4.2).

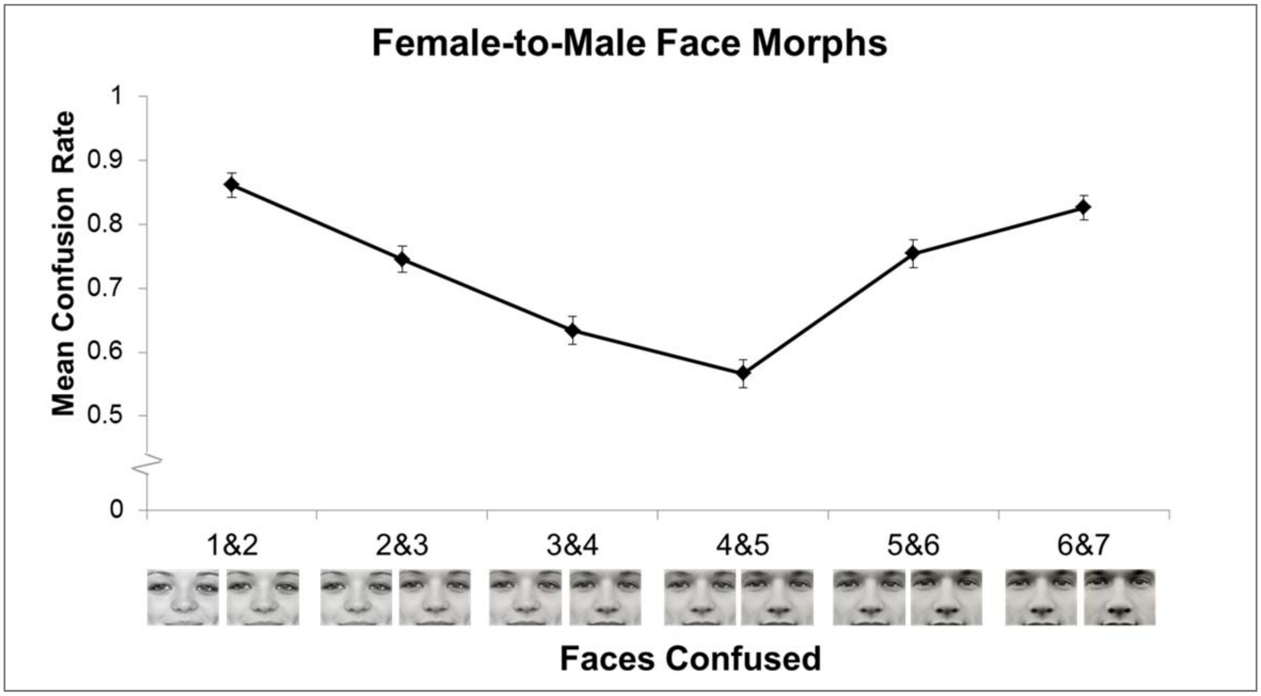


Figure 4.2. Confusion Rates for Female-to-Male Face Morphs.

To find out whether this trend holds true at the individual level, we conducted a multi-level analysis using a binomial model in HLM. The coefficient for the quadratic term was significant, $\beta_{20} = 0.22$, $p < .001$. The order 4 coefficient was not significant.

Young-to-old face morphs.

After removing data from participants with high confusion rates between faces 1 and 7, there were 31 participants with reliable data from at least four sessions. A 6-level (adjacent target pairs) repeated measures ANOVA revealed a significant cubic trend, $F(1,30) = 8.98$, $p = .005$, $\eta^2 = .05$, but the quadratic trend was nonsignificant, $F(1,30) = .007$, $p = .93$ (see Figure 4.3).

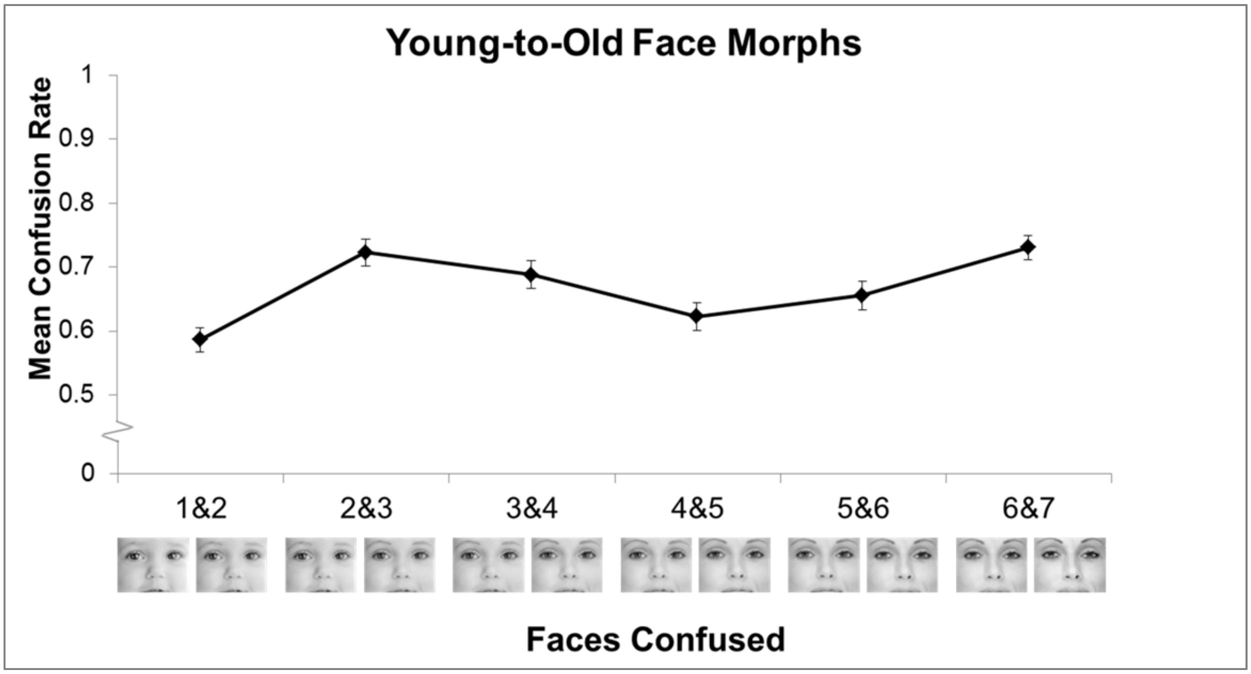


Figure 4.3. Confusion Rates for Young-to-Old Face Morphs.

To find out whether this trend holds true at the individual level, we conducted a multi-level analysis using a binomial model in HLM. The coefficient for the quadratic term was nonsignificant, $\beta_{20} = 0.009$, $p = .82$, but the cubic coefficient was, $\beta_{30} = 0.06$, $p = .02$.

Discussion

Our data support the hypotheses that faces varying in gender would be perceived categorically, but that age would be perceived less categorically. The confusion rates for gender show a strong quadratic trend, but the confusion rates for faces morphed between a young face and an older face did not. However, given that there was a significant cubic trend for the young-to-old faces, it is possible that participants automatically parsed them into three categories rather than two. We must be hesitant in interpreting the data for these face sets, however, given that they differed in ratings of attractiveness and realism. They were also limited in age range. Many stereotypes about age are associated with 'the elderly' (e.g., Cuddy & Fiske, 1998; Kite & Johnson, 1988), but the faces we used ranged from less than one year to the mid-thirties, according to participants' age ratings. If the range of faces had included a larger range of ages, there might have been a stronger tendency to automatically group faces into "younger" and "older" categories. The age range of stimulus faces was limited in order to prevent a visual artifact produced by wrinkle lines in older faces. However, the morphed faces in our target sets were perceived to be less realistic than the endpoint faces, so we may not have escaped other visual artifacts produced by morphing.

Despite hesitations about interpreting the results for the young-to-old face morphs in isolation, the comparison of results for young-to-old and female-to-male faces shows evidence of the ASCAT's discriminant validity. We predicted that female-to-male face morphs would produce a trend of confusion rates consistent with two-category automatic categorization, and that this trend would not occur for young-to-old face morphs. In fact, this is what we found: there was a clear quadratic trend for gender, but not for age. These results suggest that the ASCAT is able to detect differences in the degree of automatic categorization for different social categories. Next, we explore a third social category for which we did not have a prediction about its degree of automatic categorization.

Study 5

Study 5 involves the same participants and procedures as Study 4. However, here we examine the results for the White-to-Black face morphs. As discussed above, there is reason to believe that automatic race perception is categorical; however, since this has not been directly tested before, we did not have a strong a priori hypothesis about the degree of automatic race categorization we would see.

Method

ASCAT Stimuli.

The White-to-Black face morph sets were created from Black and White faces collected at Stanford University (Goff, Eberhardt, Williams & Jackson, 2008). Nine Black and White face pairs were chosen from this set based on pre-testing data. Each chosen face was rated as highly stereotypic of its racial category to ensure a full racial range for each morphed-race continuum. The faces were of approximately equal age and attractiveness. The same morphing procedure as described in Study 4 was used to create face morph sets for White-to-Black faces (see Figure 4.4 for sample morphed stimuli).

0% Black / 100% White	16.67% Black / 83.33% White	33.33% Black / 66.67% White	50% Black / 50% White	66.67% Black / 33.33% White	83.33% Black / 16.67% White	100% Black / 0% White

Figure 4.4. Sample Set of Face Stimuli Morphed from a White Face to a Black Face.

Pilot Study: Face stimuli ratings.

Procedures for this pilot study are reported above in Study 4. Here, we report the results for the White-to-Black faces. For these faces, participants also rated the faces on how stereotypically White and stereotypically Black they looked.

Attractiveness. A 7 (faces) x 9 (face sets) repeated measures ANOVA showed that the White-to-Black face morphs did significantly differ in perceived attractiveness, $F(6, 126) = 9.28$, $p < .001$. Specifically, there were significant linear, $F(2,21) = 5.34$, $p = .03$, and quadratic, $F(1,21) = 40.49$, $p < .001$, trends. On average, faces at the White end of the morphed face sets were rated as more attractive ($M = 4.88$, $SD = 0.89$) than faces at the Black end of the morphed face sets ($M = 5.47$, $SD = 0.87$), despite the faces having been chosen to be approximately equal in attractiveness. The stronger effect was the linear trend, such that faces in the middle were judged to be more attractive than faces at either end.

Age. For the White-to-Black faces, the rating scale for age was the same as for female-to-male faces: 1 = 16-20 years, 2 = 21-25 years, 3 = 26-30 years, 4 = 31-35 years, 5 = 36-40 years, 6 = 41-45 years, 7 = 46-50 years, 8 = 51-55 years, 9 = 56-60 years, 10 = 60+ years. A 7 (faces) by 9 (face sets) repeated measures ANOVA showed a significant effect of face on age ratings, $F(6,126) = 2.51$, $p = .03$. Specifically, there was a significant quadratic trend, $F(1,21) = 29.44$, $p < .001$, such that faces in the middle of face sets were rated as appearing younger than faces on either end of the White-to-Black face sets.

Babyfacedness. A 7 (face) by 9 (face set) repeated measures ANOVA showed no significant effect of face on how babyfaced the White-to-Black morphs looked, $F(6,126) = 1.60$, $p = .15$.

Realism. A 7 (face) by 9 (face set) repeated measures ANOVA showed no significant differences in realism ratings comparing across White-to-Black face morphs, $F(6,126) = 1.23$, $p = .30$.

Stereotypical Whiteness/Blackness. White-to-Black faces were rated on stereotypical Whiteness ("How stereotypically White does this face look?") and stereotypical Blackness ("How stereotypically Black does this face look?") on a scale of 1 to 7 (1 = Extremely, 4 = Moderately, 7 = Not at all). As expected, stereotypical Whiteness ratings decreased as the proportion of the White face decreased, with a significant linear trend across faces, $F(1,21) = 416.98$, $p < .001$. Likewise, stereotypical Blackness ratings increased as the proportion of the Black face increased, with a significant linear trend, $F(1,21) = 173.34$, $p < .001$.

Results

After removing data from participants with high confusion rates between faces 1 and 7, there were 60 participants with reliable data from at least four sessions. A 6-level (adjacent target pairs) repeated measures ANOVA revealed a significant quadratic trend in confusion rates, $F(1,59) = 35.96$, $p < .001$, $\eta^2 = 0.11$, as well as a cubic trend, $F(1,59) = 10.15$, $p = .002$, $\eta^2 = 0.02$ (see Figure 4.5).

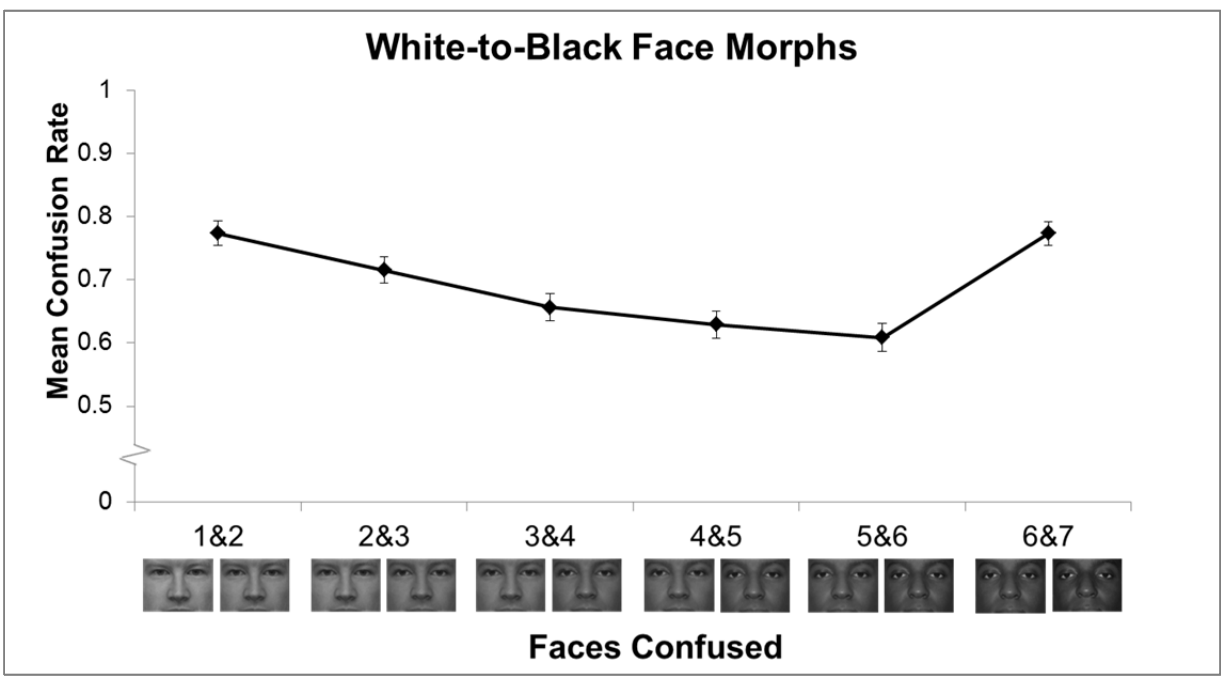


Figure 4.5. Confusion Rates for White-to-Black Face Morphs.

To find out whether these trends hold true at the individual level, we conducted a multi-level analysis using a binomial model in HLM. The coefficient for the quadratic term was significant, $\beta_{20} = 0.11$, $p < .001$. The coefficient for the cubic term was also significant, $\beta_{30} = 0.04$, $p = .005$.

Discussion

Our data suggest that the White-to-Black racial distinction falls somewhere between gender and age in term of the degree of continuous vs. categorical perception (see Figure 4.6 for a comparison of the three types of face morphs). We found a significant quadratic trend in confusion rates for White-to-Black morphed faces, suggesting that people tended to group ambiguous-race faces into one category or the other. The boundary between Black and White categories occurred on average closer to the Black end of the target spectrum; faces with any percentage of the White face were confused more with other Whiter faces than with Blacker faces. This is consistent with Willadsen-Jensen and Ito's (2006) finding that ambiguous-race faces are viewed as more similar to White than to Black or Asian faces.

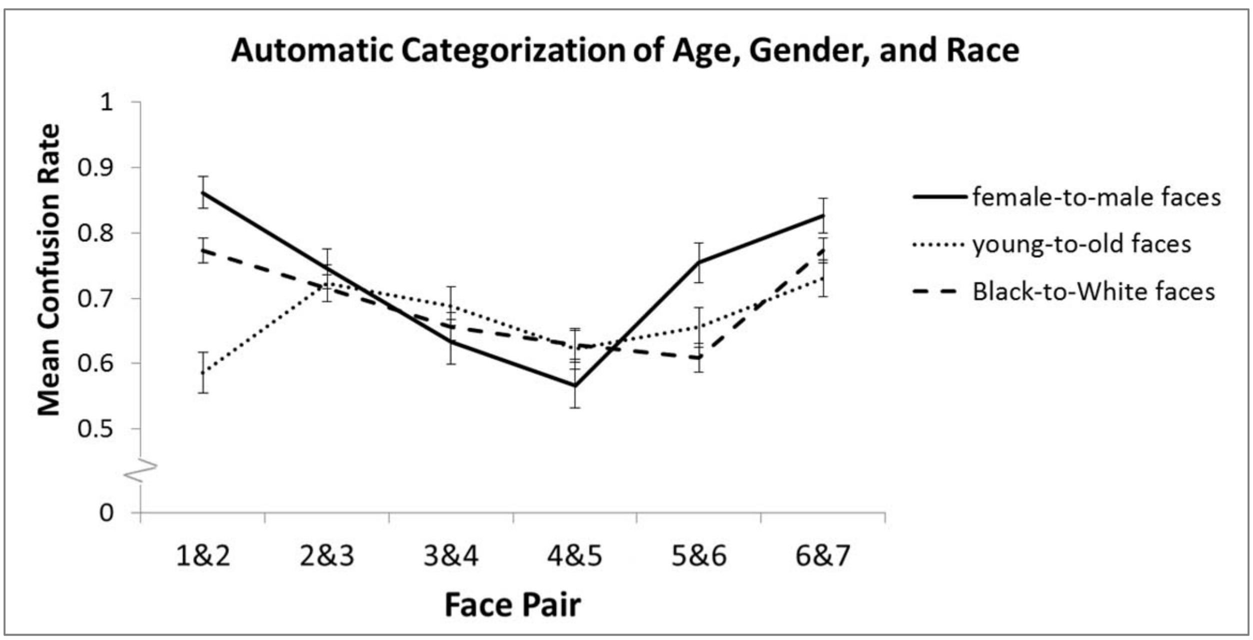


Figure 4.6. Automatic Social Categorization for Age, Gender, and Race.

We must also be cautious in interpreting these data, however, because the White-to-Black morphed faces showed a quadratic trend for perceived attractiveness, with faces in the middle of the spectrum rated as more attractive than faces at the ends. This is consistent with research demonstrating that averaging two or more faces produces faces that are perceived as more attractive than the originals (Langlois & Roggman, 1990). It works against the phenomenon we are interested in capturing, since faces at the ends of the spectrum are *more* different in attractiveness than faces in the middle; but nevertheless, we see evidence of greater perceived similarity for faces at the ends and less perceived similarity in the middle. This interpretation would make our finding all the more compelling. An alternative view of the attractiveness ratings, however, might be that attractive faces are more salient than less attractive faces and thus more memorable, and less confusable with other faces.

Study 6

In this study we test whether, when we apply the ASCAT to social stimuli, it is important that the images are perceivable as faces. If this were not the case – that is, if the quadratic trend in confusion rates occurred to the same degree for faces and non-faces – it would be problematic for our interpretation of the results as being a reflection of social categorization. For instance, the quadratic trend could reflect changes in the luminosity of the photos rather than changes in the perceived racial makeup of the faces. Thus, to perform a direct test of faces vs. non-faces, we created several stimulus sets by scrambling the pixels in White-to-Black morphed faces images. If it is important that participants view the images as faces in order to produce the categorical effect in confusion rates, then the quadratic trend we have seen thus far should be much weaker for the scrambled images than for the faces.

With this study we also introduced a more efficient version of the algorithm that produces the sequence of stimuli during the memory task, such that we were able to collect more data per participant for adjacent target pairs. We also collected data over the course of up to five study sessions. Thus, the individual-level data should be more reliable for this study than for previous studies. The issue of individual differences in automatic social categorization will be further explored in Chapter 5.

Method

Participants.

Participants were 53 students who were recruited from the University of Washington

Psychology Department Subject Pool and participated for extra credit (demographic information was not collected for this study). Participants signed up for two study sessions at a time; after completing sessions one and two, they became eligible to sign up for sessions three and four. After completing four study sessions, participants could sign up for a final fifth session.[6] Our final sample for this study consisted of 46 participants who completed at least two sessions.

Procedure.

Participants were run in groups of up to six people. In the first study session, participants were randomly assigned to one of three groups of White-to-Black face morph sets and its corresponding scrambled image set. They completed a variant of the ASCAT in which these White-to-Black morphed faces were interlaced with the scrambled images (rather than numbers). Participants were allowed to complete no more than two study sessions in a row. The memory task in this study used a more efficient algorithm than in previous studies to create the order in which stimuli were presented. Instead of presenting every possible face pair at least once (i.e., including non-adjacent face pairs), the new algorithm created a sequence of stimuli in which each adjacent pair (i.e., faces 1 and 2, faces 2 and 3, etc.) was presented at least six times (compared to a minimum of two trials per adjacent face pair in Studies 1 – 3, as well as Studies 4, 5, and 7, which were conducted before Study 6).

ASCAT Stimuli.

To reduce variability due to differences among face sets, for this study we chose three White-to-Black face sets from the original nine created for Study 5.[7] In order to create scrambled versions of these images, each photo was read into R using the 'pixmap' package (Bivand, Leish, and Maechler, 2001). The photos were each converted into a matrix in which each entry represented the grey intensity of one pixel (where 0 = black and 1 = white). The ordering of each matrix was then randomized and converted back to an image. See Figure 4.7 for an example of a White-to-Black face set with its corresponding scrambled images.

[6] We will investigate session-to-session stability in Chapter 5.

[7] These three face sets were chosen because they showed confusion rate patterns that were similar to one another and representative of the aggregate results.

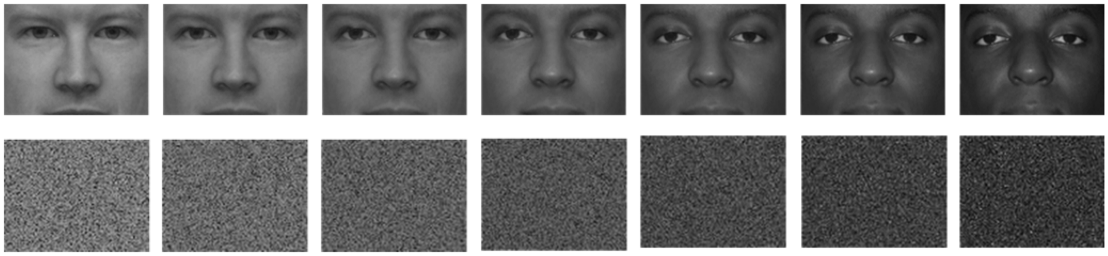


Figure 4.7. White-to-Black Face Morphs and Scrambled Images.

Results

After removing data from sessions at which the participant confused faces 1 and 7 50% of the time or more, there were 43 participants with reliable data from at least two sessions. Confusion rates were averaged across sessions for each participant. To test the hypothesis that ASCAT performance would result in a stronger quadratic trend in confusion rates for the morphed face stimuli than the scrambled stimuli, a 6 (adjacent target pair) x 2 (target type: faces vs. scrambles) repeated measures ANOVA was conducted. There was a main effect of target pair, $F(5,210) = 31.82$, $p < .001$; specifically, there was a significant overall quadratic trend, $F(1,42) = 137.53$, $p < .001$, as well as a cubic trend, $F(1,42) = 11.12$, $p = .002$. However, these effects were qualified by an interaction between target type and the quadratic trend for target pairs, $F(1,42) = 52.07$, $p < .001$. The quadratic trend was stronger for the morphed faces (η^2 = .43) than for the scrambled versions of these images (η^2 = .07; see Figure 4.8).

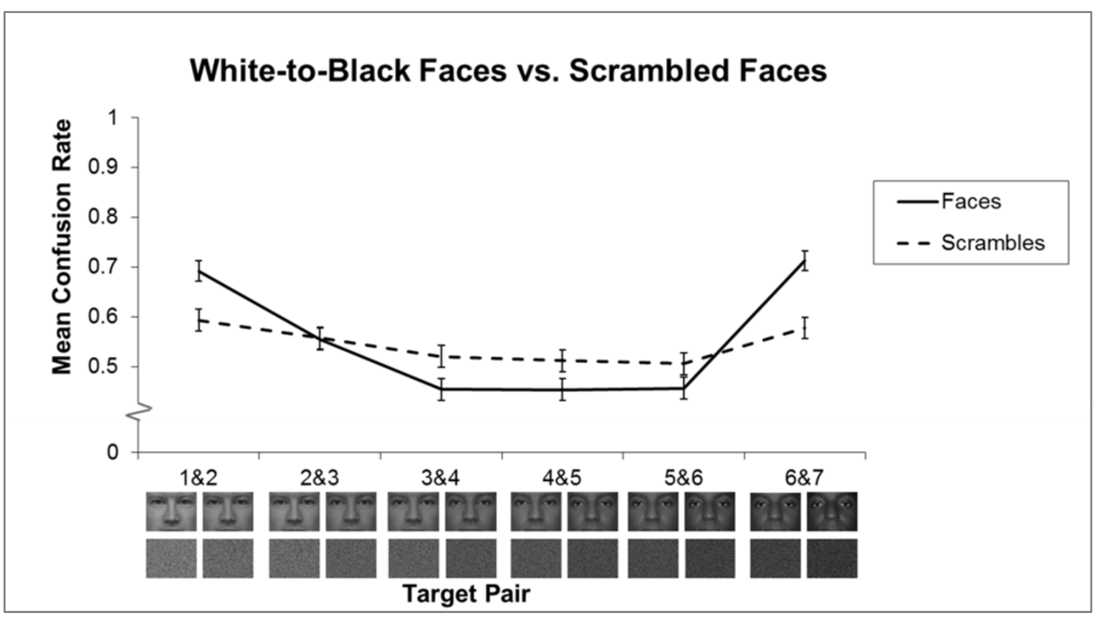


Figure 4.8. Confusion Rates for White-to-Black Morphed Faces and Scrambled Faces.

To find out whether this pattern of results held true at the individual level, we conducted a multi-level analysis using a binomial model in HLM, as we did for Study 1. To test the quadratic interaction between target type and target pair, we used the following models, with *Target Type* coded as 1 (White-to-Black morphed faces) or 2 (scrambled versions of these faces).

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(\textit{Target Type}) + \pi_2(\textit{Pair}) + \pi_3(\textit{Pair}^2) + \pi_4(\textit{Target Type*Pair}) + \pi_5(\textit{Target Type*Pair}^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

$\pi_3 = \beta_{30} + r_3$

$\pi_4 = \beta_{40} + r_4$

$\pi_5 = \beta_{50} + r_5$

The quadratic coefficient was significant, $\beta_{30} = 0.33$, $p < .001$, suggesting that regardless of target type, participants tended to confuse targets at the ends of the target sets more often than those in the middle. However, the *target type* X *target pair*^{2} interaction was also significant, $\beta_{50} = -0.14$, $p < .001$, such that the confusion rates across target pairs for the scrambled targets had a significantly smaller quadratic coefficient than confusion rates for the faces. We also examined each of these target sets separately, using the following model:

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(\textit{Pair}) + \pi_2(\textit{Pair}^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

For the faces, the coefficient for the quadratic term was significant, $\beta_{20} = 0.19$, $p < .001$. For the scrambles, the quadratic coefficient was also significant, but much smaller, $\beta_{20} = 0.05$, $p < .001$.

Discussion

These results support the argument that confusion rates for White-to-Black morphed faces are due to the fact that they are perceived as faces. Although the scrambled images also produced a quadratic pattern of confusion rates, this pattern was weaker. The difference between faces and scrambles was less pronounced than the difference between categorical gray blocks and continuous gray blocks in Study 1, however. Since the scrambled images were interlaced with the White-to-Black morphed faces, then to the extent that race is viewed categorically, it is possible that categorical perception was primed in participants, leading to a stronger quadratic pattern of confusion rates for the scrambled images than would otherwise be the case.

The results of this study, along with those of Study 5, suggest that automatic race perception is categorical. However, in examining race perception, thus far we have only examined one possible category distinction, White vs. Black faces. Historically, this is an important racial distinction in the US, but it is not representative of the racial makeup of Washington. According to the 2010 US Census, Asians are the largest minority racial group in Washington State, making up 7.2% of the population, nearly twice as large as the Black population (3.6%). Thus, in Study 7, we extend our investigation of categorical race perception to White-to-Asian morphed faces.

Study 7

There has been relatively little research on perception of racially ambiguous Asian and White faces (cf. Willadsen-Jensen & Ito, 2006; 2008). However, according to the 2010 US Census, a large portion of the growing multiracial population is made up of individuals who identify as both White and Asian (18%; Humes, Jones, & Ramirez, 2011). Also, given the racial make-up of Washington, it is likely that participants in the University of Washington Psychology Subject pool are more familiar with Asian than with Black faces and that an Asian-White category distinction may be more salient that a Black-White distinction. We were also interested in comparing Asian and White participants' patterns of confusion rates for White-to-Asian morphed faces. Willadsen-Jensen and Ito (2006) found that perceptions of ambiguous Asian-White faces by White participants were similar to perceptions of White faces. However, perceptions of ambiguous Asian-White faces by Asian participants depended on whether the majority of other faces presented were White or Asian (Willadsen-Jensen & Ito, 2008), suggesting that Asian perceivers' racial categorization of Asian-White faces may be more flexible.

Method

Participants.

Participants were 168 students who were recruited from the University of Washington Psychology Department Subject Pool and participated for extra credit (M_{age} = 19.20 years, SD_{age} = 1.76; 56 M, 101 F; 47.6% Asian, 22.0% White, 22.6% Other; 11 participants did not provide this information). Data were collected over the course of three academic quarters. There were no differences in confusion rate patterns based on which quarter data were collected, gender, age, or ethnicity.

Procedure.

Participants completed a memory task with one of two sets of White-to-Asian morphed faces. The specific face morph set was randomly assigned to participants. After the memory task, participants completed several questionnaires that were not the focus of this study, as well as demographic information.

ASCAT Stimuli.

The same procedure as in Studies 4 and 5 was used to create two face morph sets for White-to-Asian faces (see Figure 4.9 for sample morphed stimuli[8]). The faces at the endpoints were from the same face database as those chosen for the White-to-Black morphed faces sets.

[8] Although the sample White-to-Asian morphed faces shown in Figure 4.9 are female, the stimuli used in this study were male faces.

Each face was rated as highly stereotypic of its racial category to ensure a full racial range for each morphed-race continuum. The faces were of approximately equal age and attractiveness.

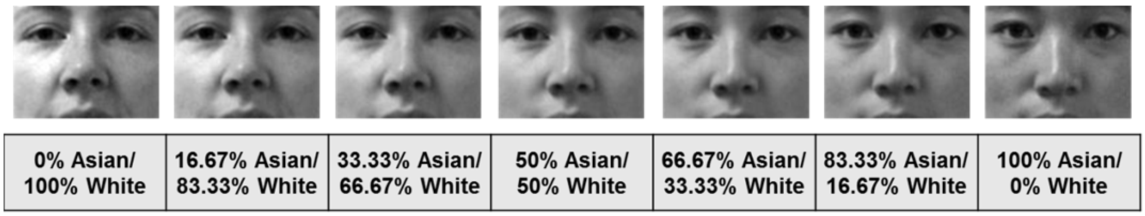


Figure 4.9. Sample White-to-Asian Face Morphs.

Results

After removing data for participants who confused the 100% White face and the 100% Asian face more than 50% of the time, there remained data from 135 participants. There was a main effect of face set, such that one face set elicited higher overall confusion rates than the other face set, $F(1,133) = 4.52$, $p = .04$, $\eta^2 = .03$. However, face set did not interact with face pair, so the relative pattern of confusion rates across adjacent face pairs was the same. Combining across face sets, there was a linear trend, $F(1,134) = 9.44$, $p = .003$, $\eta^2 = .01$, such that confusion rates were higher on average for the faces toward the White end of the spectrum. There was also a quadratic trend, $F(1,134) = 44.81$, $p < .001$, $\eta^2 = .07$ (see Figure 4.10). We compared this trend for White and Asian participants, but the interaction with ethnicity was not significant, $F(1,115) = 0.48$, $p = .49$.

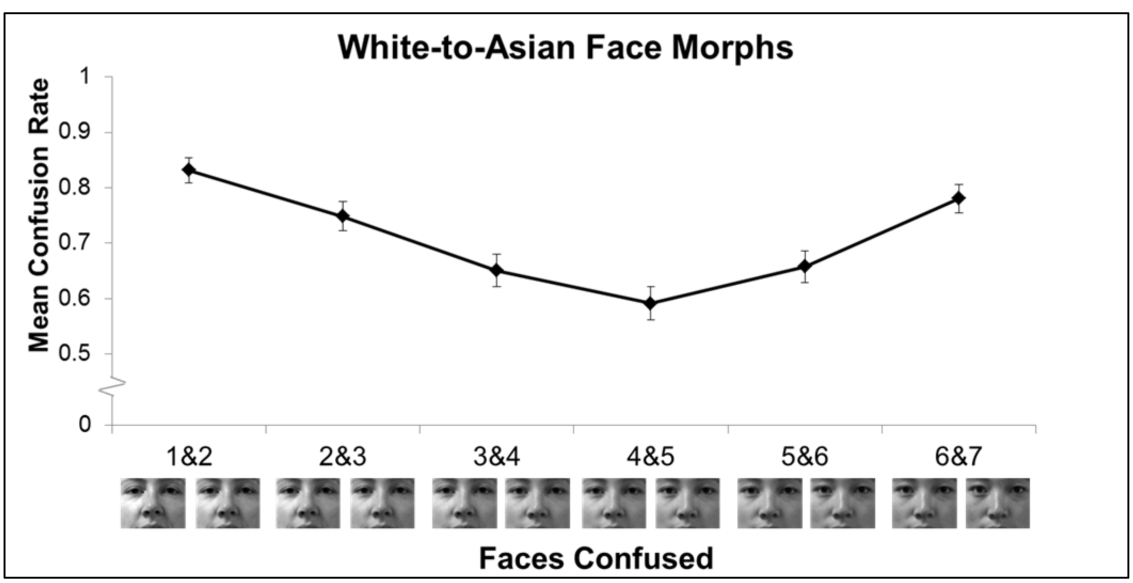


Figure 4.10. Confusion Rates for White-to-Asian Face Morphs.

When these trends were tested using HLM, the linear coefficient was significant, $\beta_{10} = -0.11$, $p = .002$. The quadratic coefficient was also significant, $\beta_{20} = 0.17$, $p < .001$. To test whether this effect varied depending on participant ethnicity, we added ethnicity (Asian or White) as a level 2 predictor. However, consistent with the ANOVA results, participant ethnicity was not a significant predictor of the quadratic term, $\beta_{21} = 0.01$, $p = .81$.

Discussion

Study 7 replicated the findings in Studies 5 and 6 that automatic perception of race tends to be categorical. There was also a significant linear trend, such that the faces closer to the 100% White face were confused more often than faces closer to the 100% Asian face, although the effect size for this trend was small. As with the White-to-Black face morphs, the category boundary appears to be closer to the Asian side of the morphed spectrum. We expected that Asian participants might show a weaker tendency to group the ambiguous faces with one category or another, but they did not differ from White participants in this respect. There were also no differences between Asian and White participants in the linear trend or in their overall confusion rates.

General Discussion

In sum, Studies 4 – 7 demonstrate the use of the ASCAT with social stimuli (i.e., faces), showing that the measure is sensitive to automatic categorical perception of faces (as opposed to non-faces) and that it detects differences in the degree of automatic categorical perception for social categories we expected to differ (gender and age). As predicted, faces varying from a female face to a male face were automatically grouped into one of two categories to a stronger degree than faces varying from a young face to an older face. Although the quadratic term for young-to-old morphed faces was nonsignificant, there was a small cubic effect, suggesting that perhaps faces were actually grouped according to three categories. HLM analyses revealed that this effect held at the individual level, so the less-dichotomous pattern for young-to-old faces was not a result of averaging across participants with different thresholds.

We also applied the ASCAT to a question that had not yet been adequately addressed empirically: is race perception categorical? By examining faces along a full continuum from a Black or Asian face to a White face, we allowed for the possibility of finding evidence for continuous perception of race. However, confusion rates for neighboring faces along these continuums suggested that people do tend to automatically group ambiguous faces into categories. This finding is especially important today, because according to the 2000 US Census, 6.8 million people identify with more than one racial category. As this population increases, greater numbers of multiracial individuals will enter the nation's consciousness. The difference between a more or less automatic race categorization could have implications for everyday interracial interactions.

Although automatic social categorization may shed light on real-world behavior, we do not necessarily expect our lab results with the ASCAT to represent everyday settings. In this research, we used morphed faces rather than photographs of people who vary along the dimensions of interest. This method ensured that the faces were evenly spaced along an objective continuum, but this strategy may raise questions about external validity (e.g., possibly the morphed faces are unusual and atypical relative to faces one might encounter in a natural environment). To address this, participants rated the morphed White-to-Black, female-to-male, and young-to-old faces on realism (i.e., “How realistic is this face?”). The White-to-Black and female-to-male faces did not differ significantly with regard to perceived realism, suggesting that the morphed faces were not perceived as less realistic than the endpoint faces. For the young-to-old faces, however, the morphed faces were rated as less realistic than the endpoints; this may explain the cubic trend in confusion rates, if the morphed faces formed a third category of faces that were noticeably strange-looking. Future research could further address these issues by using real faces as stimuli.

CHAPTER 5.
ARE THERE RELIABLE INDIVIDUAL DIFFERENCES IN AUTOMATIC SOCIAL CATEGORIZATION?

Some individuals may chronically automatically perceive social categories like race more or less categorically. Some may see the social world more in terms of "Black and White," while others implicitly perceive social categories as more continuous. For example, people who spontaneously used the term "multiracial" or "mixed-race" to describe Barack Obama have a less categorical automatic perception of race than others (Malahy, Sedlins, Plaks, & Shoda, 2010). Furthermore, racial categories may be particularly important to some individuals more than others. For example, some people are chronically more threatened by ambiguity than others (Kruglanski & Webster, 1996) and thus may perceive race more categorically. By contrast, racial heritage or social contact may make some people more chronically aware of the continuity of race. For example, research has shown that relative to monoracial individuals, those with mixed-race heritage view race as less essentialized and are less likely to show the cross-race effect (Pauker & Ambady, 2009). Thus, multiracial individuals may have more continuous race perception than others.

To provide empirical support for the viability of examining individual differences as predictors of automatic social categorization, we would like to show that there are reliable individual differences in the quadratic term for confusion rates from the ASCAT. To do this, we first model confusion rates nested within individuals in HLM and examine the reliability estimates for the level-1 quadratic term. These estimates indicate how reliably the quadratic term varies from individual to individual (Raudenbush & Bryk, 2002), and thus may potentially be accounted for by a level-2 predictor (e.g., participant ethnicity). The chi-square test for the estimated variance of the quadratic term gives similar information, testing the null hypothesis that the quadratic term does not differ across individuals (Raudenbush & Bryk). We will examine these indicators of reliable individual differences for data sets described thus far, as well as for Study 8, a multi-session study with White-to-Black morphed faces. The reliability estimate and chi-square test are informative to the extent that there are a large number of observations per individual, so we will focus on data sets for which we have an adequate number of individual-level trials (i.e. studies for which we used the more efficient sequencing algorithm, and/or for which data were collected across several study sessions).

We would also like to examine the stability across multiple testing sessions of the quadratic trend in how confusion rates produced by the ASCAT vary across adjacent face pairs. Thus, for each data set that was collected across several study sessions (Studies 4, 5, 6, and 8), we examine the interaction between session and the quadratic trend in a repeated measures ANOVA, followed by a three-level HLM analysis, with adjacent face pairs nested within sessions nested within individuals.

Finally, we would like to rule out an alternative explanation for individual variability in the strength of the quadratic trend for confusion rates. Because ASCAT uses n-back and task-switching paradigms, which have been used to measure working memory (e.g. Friedman et al., 2008); when working memory is reduced, participants tend to rely on simpler strategies including category-based processing (Macrae, Milne, & Bodenhausen, 1994). Thus, observed differences in confusion rate patterns may be due to either individual differences in category use *or* individual differences in working memory. To rule out this confound, we estimate participants' working memory by computing their overall tendency to make confusion errors during the memory task (i.e. not just confusions between adjacent faces). If overall confusion rate is correlated with the quadratic trend, then ASCAT performance may reflect working memory instead of the tendency to categorize.

Study 8

To obtain more reliable ASCAT data at the individual level and to assess the stability of the measure across time, we recruited participants for a five-session study examining automatic categorical perception of White-to-Black morphed faces. This study also further tests the replicability of the finding that race perception tends to be categorical. Before coming to the lab, participants also completed a battery of questionnaires; however, for the purposes of this study, we focus on the stability of the ASCAT over multiple sessions and the reliability of individual differences in ASCAT performance.

Method

Participants.

One hundred three participants (M_{age} = 19.41 years, SD_{age} = 3.50; 29 M, 66 F; 43.7% White, 35.9% Asian, 12.6% Other; 8 participants did not provide this information) were recruited over the course of two academic quarters, through the University of Washington Psychology Subject Pool (n = 36) or with flyers posted around the University of Washington campus (n = 67). Participants completed an online battery of questionnaires (for which subject pool participants received extra credit), after which they were invited to participate in a paid five-session lab study. Participants were paid $5 per half-hour study session, as well as a $5 bonus if they attended all five sessions. Our final sample for this study consisted of 83 participants who completed at least three sessions. There were no significant differences in confusion patterns based on recruitment method, quarter, gender, or ethnicity. There was an interaction with age that was no longer significant after an outlier was removed.[9]

Procedure.

Participants were run in groups of up to six people. In the first study session, participants were randomly assigned to one of three groups of White-to-Black face morph sets.[10] They completed the memory task with this same face set at each subsequent session. Participants were allowed to complete no more than two study sessions in a row. The memory task in this study used the more efficient algorithm introduced in Study 6, allowing us to collect data for 12 trials per adjacent face pair.

ASCAT Stimuli.

This study used the same nine White-to-Black face sets as in Study 5, interlaced with the numbers 1 to 7.

Results

After removing data for participants who produced high confusion rates (greater than 50% for a given session) between the 100% White face and the 100% Black face, our sample consisted of 82 participants with reliable data from at least three sessions. Confusion rates for adjacent faces were averaged across sessions for each participant. Replicating our findings with to White-to-Black morphed faces from Studies 4 and 5, a one-way repeated measures ANOVA

---

[9] All participants ranged from 18 to 24 except for one participant, who was 51 years old; this person's pattern of confusion rates was markedly different than average, showing a negative quadratic trend.

[10] In the first quarter of data collection, participants were assigned to one of the nine face sets originally created for Study 4, but in the second quarter they were assigned to one of the three face sets used in Study 6. In the end, 72% of participants were assigned one of these more reliable face sets.

(6 adjacent face pairs) revealed a significant quadratic trend in confusion rates, $F(1,81) = 257.40$, $p < .001$, $\eta^2 = .42$ (see Figure 5.1).

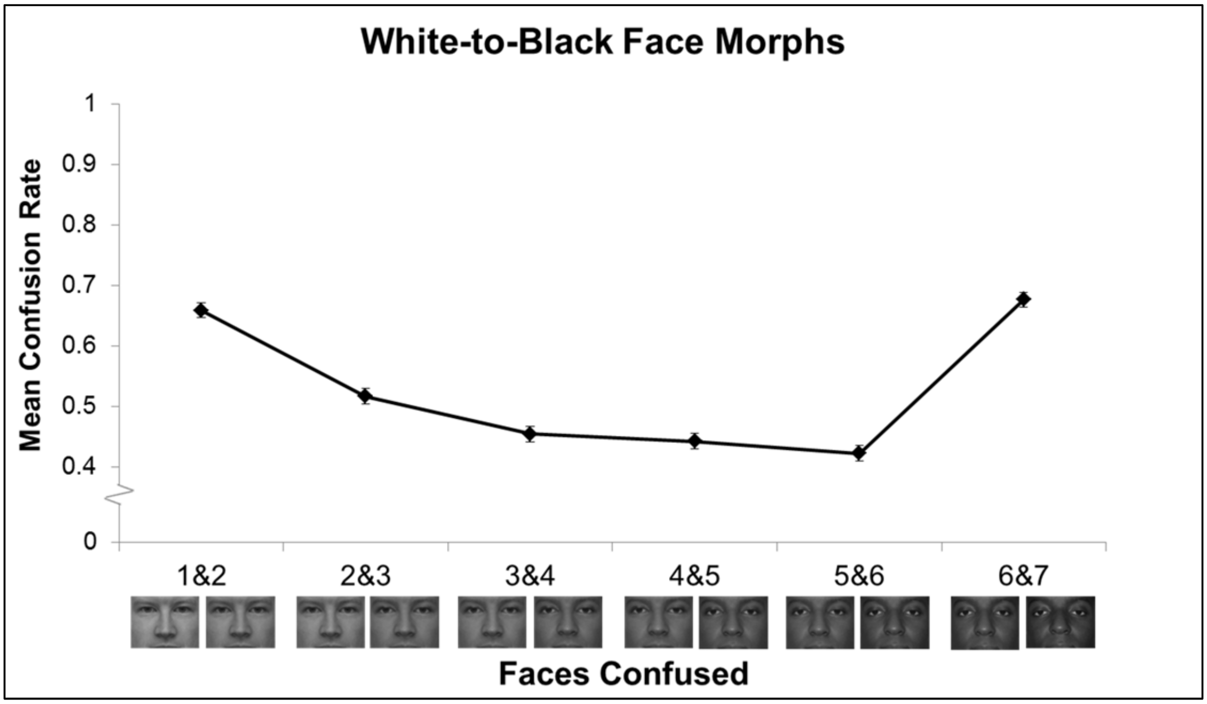


Figure 5.1. Confusion Rates for White-to-Black Face Morphs Averaged Across at Least Three Study Sessions.

When these trends were tested using HLM, the quadratic coefficient was significant, $\beta_{20} = 0.17$, $p < .001$.

Effect of session.

To examine the stability of ASCAT confusion rates across study sessions, we first performed a 5 (session) by 6 (face pair) repeated measures ANOVA (there were 32 participants who completed all five study sessions). We found a significant overall interaction between session and the quadratic trend, $F(20,620) = 1.86$, $p = .01$. There was also a significant interaction between the linear trend of session and the quadratic trend of face pairs, $F(1,31) = 6.49$, $p = .02$, suggesting that the quadratic trend changed systematically over the course of the five sessions. Inspection of Figure 5.2 suggests that the quadratic trend became smaller.

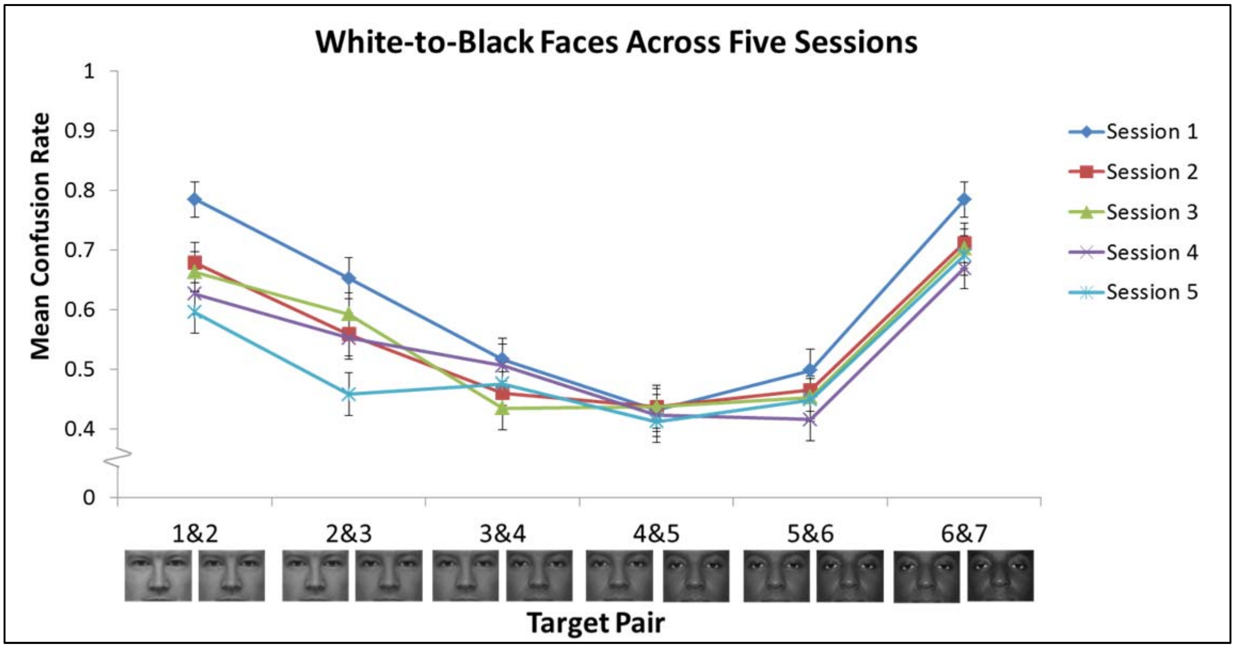


Figure 5.2. Confusion Rates for White-to-Black Face Morphs Across Five Sessions.

To test the effect of session at the level of the individual participant, we conducted 3-level HLM analyses, with confusion rates nested within sessions nested within individuals:[11]

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\mathrm{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(Pair) + \pi_2(Pair^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

Level 3:

$\beta_{00} = \gamma_{000} + u_{00}$

$\beta_{10} = \gamma_{100} + u_{10}$

$\beta_{20} = \gamma_{200} + u_{20}$

[11] Note that since not every participant completed all five sessions, there was missing data at level 1.

To test for a significant session effect for the quadratic term, we look at the estimation of r_2, the random effect for the quadratic term at level 2 (Raudenbush & Bryk, 2002, p. 229). A chi-square test showed that the variance in the quadratic term across sessions was not significantly different from zero, $\chi^2 = 173.05$ (254, N = 82), $p > .50$. However, to follow up on the significant interaction between the linear trend of session and the quadratic trend of face pairs described above, we also entered session as a level 2 predictor:

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1-\varphi)] = \eta$

$\eta = \pi_0 + \pi_1(Pair) + \pi_2(Pair^2)$

Level 2:

$\pi_0 = \beta_{00} + \beta_{01}(Session) + r_0$

$\pi_1 = \beta_{10} + \beta_{11}(Session) + r_1$

$\pi_2 = \beta_{20} + \beta_{21}(Session) + r_2$

Level 3:

$\beta_{00} = \gamma_{000} + u_{00}$

$\beta_{01} = \gamma_{010} + u_{01}$

$\beta_{10} = \gamma_{100} + u_{10}$

$\beta_{11} = \gamma_{110} + u_{11}$

$\beta_{20} = \gamma_{200} + u_{20}$

$\beta_{21} = \gamma_{210} + u_{21}$

The coefficient for the linear effect of session on the quadratic trend, γ_{210}, was estimated as -0.01, $p = .003$. The negative coefficient confirms the trend suggested by Figure 5.2; the quadratic trend became smaller over the course of the five sessions. We repeated this analysis for the other studies involving multiple sessions (see Table 5.1).

Table 5.1. Chi-Square Tests for r_2 and Estimates of the Effect of Session on the Quadratic Trend.

Study	Target type	Number of sessions	Chi-square test for r_2	Session effect
8	White-to-Black faces	5	$\chi^2 = 173.05$ (254, N = 82), $p > .50$	$\gamma_{210} = -0.01$, $p < .01$
4	female-to-male faces	6	$\chi^2 = 96.29$ (112, N = 64), $p > .50$	$\gamma_{210} = -0.02$, $p = .14$
4	young-to-old faces	6	$\chi^2 = 98.99$ (119, N = 62), $p > .50$	$\gamma_{210} = 0.01$, $p = .44$
5	White-to-Black faces	6	$\chi^2 = 204.19$ (251, N = 75), $p > .50$	$\gamma_{210} = -0.02$, $p = .04$
6	White-to-Black faces	5	$\chi^2 = 54.70$ (74, N = 49), $p > .50$	$\gamma_{210} = -0.01$, $p = .54$
6	scrambled images	5	$\chi^2 = 45.27$ (74, N = 49), $p > .50$	$\gamma_{210} < 0.01$, $p = .98$

For Studies 4 and 6, the results from 3-level HLM analyses did not show a significant effect of session on patterns of confusion rates.[12] However, Study 5 showed the same pattern as in Study 8; the quadratic trend decreased over the course of six sessions, $\gamma_{210} = -0.02$, $p = .04$, odds ratio = 0.99.

Individual differences in the quadratic coefficients.

To examine whether there are reliable individual differences in automatic categorical perception, we conducted an analysis with HLM. For the purpose of looking at individual differences, we first averaged confusion rates across sessions for each individual.[13] Therefore, here we use a two-level model with confusion rates nested within individuals.

Level 1:

E(*[Number of times a pair was confused]* = 1 | π) = φ**[Number of times each pair appeared]*

$\text{Log}[\varphi(1- \varphi)] = \eta$

$\eta = \pi_0 + \pi_1(Pair) + \pi_2(Pair^2)$

Level 2:

$\pi_0 = \beta_{00} + r_0$

$\pi_1 = \beta_{10} + r_1$

$\pi_2 = \beta_{20} + r_2$

---

[12] We also examined the effect of session on the cubic trend for young-to-old faces from Study 4, but the effect was nonsignificant: $\chi^2 = 89.01$ (119, N = 62), $p > .50$; $\gamma_{310} = -0.01$, $p = .51$.

[13] For Study 8, we analyzed data for participants who had completed at least 3 sessions; for Studies 4 and 5, for participants who had completed at least 4 sessions; and for Study 6, for participants who had completed at least 2 sessions.

The reliability estimate and chi-square test for r_2 suggest that there are moderately reliable individual differences in the strength of the quadratic trend in confusion rates (reliability = .29 and χ^2 = 112.03 (81, N = 82), p = .01). To further explore the reliability of individual differences in ASCAT performance, we repeated this analysis for other data sets for which we were able to collect confusion rates for more than two trials per adjacent target pair (see Table 5.2).

Table 5.2. HLM Reliability Estimates and Chi-Square Tests for the Level 1 Quadratic Term.

Study	**Target type**	**Number of trials per target pair***	**Reliability estimate**	**Chi-square test for r_2**
8	White-to-Black faces	18	.29	χ^2 = 112.03 (81, N = 82), p = .01
4	female-to-male faces	8	.01	χ^2 = 24.07 (24, N = 25), p = .46
4	young-to-old faces	8	.61	χ^2 = 78.34 (30, N = 31), p < .001
5	White-to-Black faces	8	.18	χ^2 = 72.51 (59, N = 60), p = .11
6	White-to-Black faces	12	.23	χ^2 = 43.02 (42, N = 43), p = .43
6	scrambled images	12	.05	χ^2 = 28.28 (42, N = 43), p > .50

* This is based on the minimum number of trials completed by each participant included in the analysis.

The results do not show evidence of reliable individual differences in the quadratic trend in confusion rates for Studies 5 or 6, or for female-to-male faces in Study 4. However, there were reliable individual differences in the quadratic trend for young-to-old faces. Upon examination of the individual-level data, we found that the quadratic coefficients ranged from -0.09 to 0.08, with a median of 0.004 (this explains why in the aggregate, the quadratic trend for young-to-old faces is nonsignificant). We also examined the reliability of individual differences in the cubic trend for young-to-old faces in Study 4 and the chi-square test for r^2 was significant, χ^2 = 52.98 (30, N = 31), p < .01, reliability estimate = .40.

Relation between overall confusion rates and strength of the quadratic trend.

To rule out the potential confound that individual differences are due to differences in working memory ability rather than automatic social categorization, we use overall confusion rate as a proxy for working memory ability. That is, for each participant, we calculated the average proportion of time that each possible pair of faces was confused with one another. A one-way (6 adjacent face pairs) repeated measures ANOVA with the individual's average confusion rate entered as a covariate did not show a significant interaction between average confusion rate and the quadratic trend across adjacent face pairs, F(1,80) = 0.66, p = .42. Figure

5.3 shows the pattern of confusion rates for participants with low vs. high confusion rates (based on a median split), and the quadratic trend appears the same. A follow-up analysis using HLM showed that average confusion rate is a not a significant level 2 predictor of the quadratic trend, $\beta_{21} = 0.09$, $p = .16$. Although nonsignificant, the direction of the difference is that higher average confusion rates are associated with stronger quadratic trends.

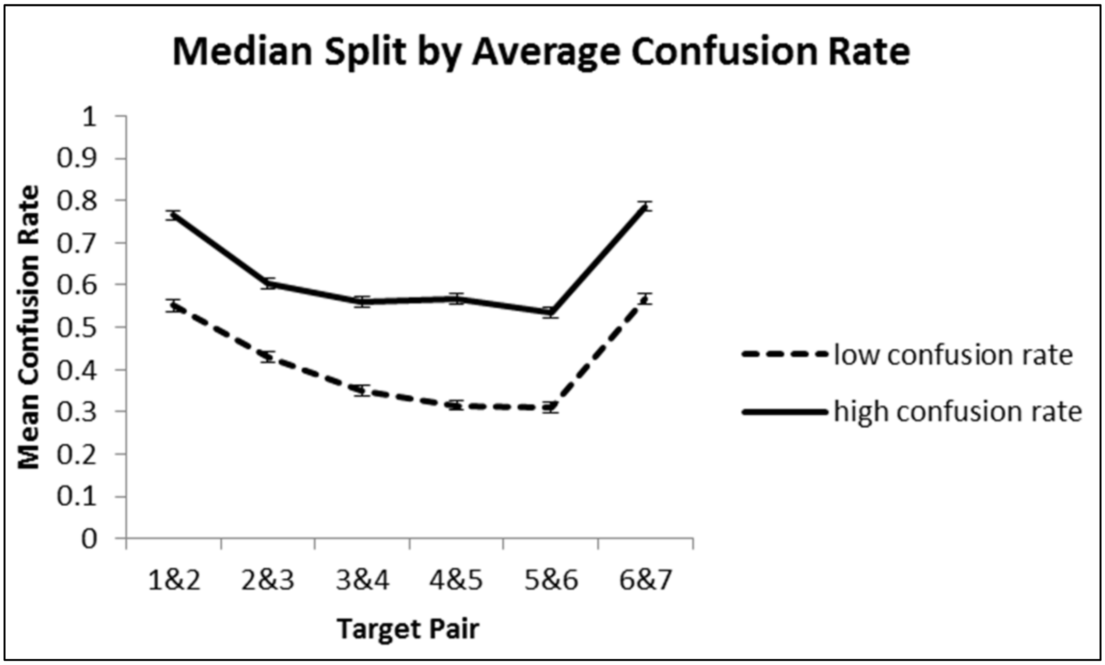


Figure 5.3. Average Confusion Rate as a Predictor of Automatic Race Categorization.

Because it only makes sense to rule out a confound for individual differences where differences occur, we repeat this analysis for the young-to-old faces from Study 4 only. A one-way (6 adjacent face pairs) repeated measures ANOVA with average confusion rate for young-to-old faces entered as a covariate did not show a significant interaction between average confusion rate and the cubic trend across adjacent face pairs, $F(1,29) = 1.29$, $p = .27$.

Relation between fatigue and strength of the quadratic trend.

To examine whether the individual differences in automatic race categorization reflect differential participant fatigue, we examined the effect of having recently completed the ASCAT. When participants had completed a memory task within the last hour, on average the quadratic coefficient was significantly weaker ($p < .001$ level) than when their last session was longer than one day ago (or when it was their first session). Among participants who had completed a memory task session within the same day, there was no significant interaction between the quadratic trend and the number of minutes since their last session, $F(1,101) = 0.36$, $p = .55$.

We also examined potential fatigue effects in a study in which participants completed the ASCAT and another task (evaluative priming). The results showed a stronger quadratic trend when participants did the ASCAT first, rather than after the priming task. Those who did the ASCAT first showed stronger automatic race categorization than those who did the evaluative priming task first. The interaction between face pairs and order was significant, $F(5, 380) = 2.66$, $p = .02$.

Taken together, these results suggest that the categorical pattern of confusion rates observed using the MCT cannot be explained by diminishing working memory capacities or by participant fatigue.

Discussion

On the whole, the results suggest that in order to examine individual differences in automatic social categorization using the ASCAT, it is worthwhile to collect more data per participant per face pair than we did in Studies 1 – 7. In Study 8, however, we obtained evidence of reliable individual differences with a minimum of 18 total memory confusion trials per adjacent face pair (6 per study session, averaged across at least 3 sessions). There may also be reliable individual differences in the strength of the cubic trend for young-to-old morphed faces in Study 4. However, since the young-to-old morphed faces varied in perceived realism, individual differences in the cubic trend may not reflect differences in the degree of automatic categorization of age but differing responses to the less-realistic morphed faces.

The variation in degree of automatic categorical perception of White-to-Black faces in Study 8 cannot be explained by variation in working memory ability, approximated by participants’ overall performance on the ASCAT. Thus, differences in the tendency to confuse faces on the ends of the morphed face spectrum more than those in the middle are not likely to be due to differential reliance on categorization as a memory strategy. This could be tested more systematically in the future by examining the relation between other measures of working memory capacity and automatic social categorization as measured by the ASCAT.

Other studies have suggested that there are individual differences in performance on the ASCAT by showing it to be a predictor of other conceptually related individual differences. Although HLM analyses for White-to-Black morphed faces in Study 5 did not suggest that there were reliable individual differences in automatic categorical race perception, other data from that study showed that individual differences in the tendency to use a multiracial label was a significant predictor of categorical race perception (Malahy, et al., 2010). Participants who generated multiracial terms to describe Barack Obama and other biracial celebrities (as opposed to monoracial terms like Black or African American) showed less categorical automatic race perception during the ASCAT. Another study found that genetic beliefs about race were a significant predictor of automatic categorical race perception as measured by the ASCAT (Plaks, Malahy, Sedlins, & Shoda, 2012). Political ideology may also be associated with automatic race categorization, with more liberal participants showing less categorical perception of race (Malahy, in prep).

To improve the reliability of the ASCAT as a measure of individual differences in automatic social categorization, future versions of the task may add more adjacent-pair trials per memory task session. However, there is a practical limit to the number of trials participants can complete, since the memory task is taxing and may be subject to fatigue or practice effects if allowed to become too long. Also, the analysis of session effects suggests that the automatic categorization effect, as measured by the quadratic trend in confusion rates, may become weaker with repeated sessions. The precise limits of task length could be explored in the future to determine the optimal number of sessions for producing reliable individual-level data with minimal distortion of the categorization effect from repeated testing.

CHAPTER 6. CONCLUSION

The studies presented in this dissertation introduce a new theoretical and empirical approach to social categorization, implementing methods drawn from the categorical perception literature to examine *degrees* of automatic categorical vs. continuous perception of social categories. We assess the construct validity of the Automatic Social Categorization Test (ASCAT) as a measure of automatic categorical perception, showing that confusion rates reflect objective differences between targets, as well as culturally learned categories. We provide evidence of the ASCAT's discriminant validity for social stimuli (i.e., faces varying in age or gender) and then use the task to measure race categorization in a way that has not been done before. Finally, we demonstrate the task's viability as a measure of individual differences in automatic race categorization.

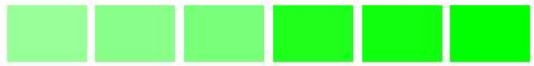
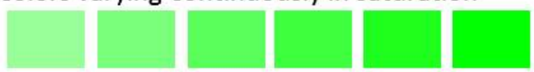
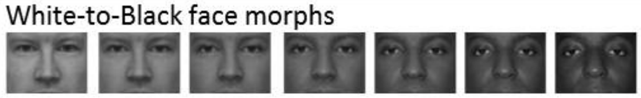
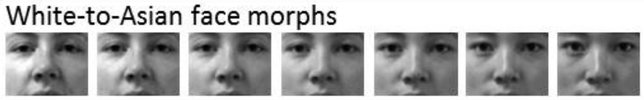
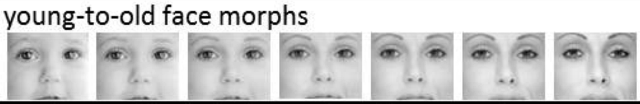
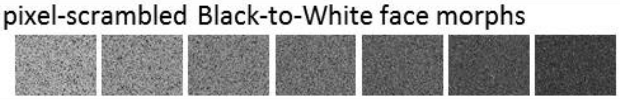
Confusion rates reflect objective target distributions.

The ASCAT is based on the phenomenon that pairs of stimuli become more confusable the more similar they are in a participant's mind. Similarity can be defined either objectively (in terms of the physical properties of the targets) or subjectively (in terms of a person's mental representation of the targets). We began by making sure that confusion rates obtained from the ASCAT reflect differences in the objective similarity between pairs of stimuli. The meta-analysis presented in Chapter 2 showed that across all studies, as the physical difference between target pairs increase, confusion rates decrease.

Relatedly, we also wanted to confirm that the ASCAT produces a categorical pattern of confusion rates (a drop at category boundaries) when stimuli are objectively categorical (varying with a gap). Studies 1 and 2 showed this to be true. Table 6.1 summarizes the quadratic effects we saw across studies and target types (except for Study 3, for which a quadratic trend did not make sense theoretically). The strongest quadratic trends were found for target sets that were constructed to contain an objective gap in the middle.[14] This suggests that when stimuli are continuously spaced but elicit a categorical pattern of confusion rates similar to that for objectively categorical stimuli, the pattern reflects a subjective representation of the stimuli that is categorical.

[14] Note that the quadratic coefficients for these studies cannot be compared directly to those for Studies 4-8, since the number of targets (and thus, the number of adjacent target pairs modeled along the x-axis) is slightly different (6 targets for studies 1 and 2, and 7 for other studies).

Table 6.1. Summary of Automatic Categorical Perception Effects

	Study	Target type	Quadratic coefficient (β_{20})	N
with a gap	1	greys varying in lightness with a gap 	2.31***	56
	2	colors varying in saturation with a gap 	0.85***	14
evenly spaced	2	colors varying continuously in saturation 	0.20*	14
	1	greys varying continuously in lightness 	0.15**	56
gender	4	female-to-male face morphs 	0.22***	25
race	6	White-to-Black face morphs 	0.19***	43
	8	White-to-Black face morphs 	0.17***	82
	5	White-to-Black face morphs 	0.11***	60
	7	White-to-Asian face morphs 	0.17***	135
age	4	young-to-old face morphs 	0	31
luminosity	6	pixel-scrambled Black-to-White face morphs 	0.05***	43

* *p* < .05 ** *p* < .01 *** *p* < .001

Confusion rates reflect culturally learned categories.

There are many domains for which there are culturally determined categories that we would expect to be present for most individuals. One of these is color. People within a culture tend to parse colors in the same way, and there are even consistencies in color categories across cultures (Berlin & Kay, 1969; Kay & Regier, 2006). In Study 3, we showed that the ASCAT detects category boundaries for color that are consistent with English color names.

The ASCAT also detects predicted differences in the degree of categorization for two different social categories. Gender is usually defined in terms of a male-female dichotomy; although there are exceptions (e.g. the hijra of India and Pakistan), we expected our participants to think of gender as falling into two discrete categories. By contrast, age varies in a more continuous way. Although there are culturally significant ages (18, 21, decades), we expected participants to have a less categorical conception of age than of gender. Table 6.2 shows that among morphed face stimuli, the strongest quadratic trend was for female-to-male face morphs, but the quadratic trend for age was nonsignificant.

Race, particularly the Black-White distinction, is another culturally learned category that was reflected in the ASCAT confusion data. For White-to-Black face morphs, the quadratic coefficient ranged from 0.11 (Study 5) to 0.19 (Study 6); for White-to-Asian face morphs, the quadratic coefficient was in the same range, at 0.17 (Study 7). These data indicate that, in the aggregate, people tend to automatically perceive faces according to discrete race categories (Studies 4, 6, and 7), although the categorical trend is not as strong as that for gender.

We were also able to rule out the possibility that categorical perception of White-to-Black face morphs was due, not to the social information contained in the faces, but to differences in luminance. The pixel-scrambled faces produced the weakest quadratic effect except for the young-to-old face morphs. Note that the pixel-scrambled images were interlaced with their corresponding White-to-Black face morphs, so they are not directly comparable to the other stimulus sets. If anything, we may be overestimating the quadratic effect for pixel-scrambled images relative to the morphed faces, since presenting them with the White-to-Black face morphs may have primed the idea of categorization.

Our demonstration that automatic race perception is categorical is an important contribution to the social categorization literature. To our knowledge, this has not been demonstrated before in a way that allows for testing across a full range of ambiguous faces. As discussed in Chapter 4, past research on race categorization has tended to use target faces that are prototypical of a racial category, or to require participants to use monoracial category labels (e.g.

Black, White). The ASCAT avoids these limitations by using target faces that vary along a racial spectrum, and by relying on spontaneous confusion errors to illuminate participants' underlying mental representations of faces. We show that the phenomenon of categorical perception, as demonstrated with non-social stimuli like colors and phonemes, extends to faces varying in racial composition. People tend to confuse two faces with *different physical features* if their features fall into the same perceived race category (e.g. Black); in contrast, two people who are *physically quite similar* but fall into different subjective categories (e.g., Black and White) are not as frequently confused with each other.

Individual differences in automatic social categorization.

The studies just described increase our confidence that the ASCAT is a valid measure of automatic social categorization. That is, it detects varying degrees of categorical vs. continuous perception of faces that vary along social dimensions such as gender, age, and race. By relying on confusion errors rather than more direct indicators of categorization, the task accesses automatic, perceptual processes rather than explicitly held views on social categories. Study 8 showed that with adequate individual-level data, the ASCAT can be used to examine individual differences in automatic race categorization.

Our emphasis on individual differences in the *degree* of categorical perception of race is a novel contribution to research on race perception. Past research has focused on when the concept of race is likely to be activated. For example, researchers have examined: 1) when or in what contexts people use social categories (i.e., when is race salient?; Devine, 1989; Dovidio, Evans, & Tyler, 1986; Lepore & Brown, 1997; Macrae & Bodenhusen, 2000) and 2) which social categories are activated for multiply categorizable individuals (e.g., gender, race, profession; Bodenhausen & Macrae, 1998; Crisp & Hewstone, 2007). Researchers have also focused on the consequences of grouping people into categories – that categorization often leads to group stereotyping and prejudice (e.g., Tajfel, 1970, 1978; Tajfel, Billig, Bundy, & Flament, 1971). However, for the most part, it has been assumed that race perception is categorical. The fact that we find evidence of individual variation in the strength of automatic race categorization is notable in itself; now that we have a tool for measuring individual differences in this tendency, we can explore the correlates and consequences of having a more categorical vs. continuous perception of race.

Individual difference predictors of automatic race categorization.

Multiracial heritage. Some research suggests that race is perceived differently by monoracial and multiracial individuals (Pauker & Ambady, 2009; Willadsen-Jenson & Ito, 2008). Multiracial individuals are less likely to use race category labels and tend to perceive race as less essentialized (Pauker & Ambady). Thus, multiracial individuals themselves may view race less categorically than monoracial individuals.

Multiracial salience. Malahy, et al. (2010) examined the effect of the accessibility of the concept of "mixed-race." For example, in one study, we asked people to generate explicit classifications of various mixed-race public figures (e.g., Halle Berry, Barack Obama). Participants responded to prompts such as, "I would racially categorize Barack Obama as _____." Responses were then manually coded as monoracial or multiracial. We posited that if an individual used a multiracial term at least once to describe a multiracial public figure, one could infer that the concept of "multiracial" is available to some extent to this individual. In contrast, never using the term "multiracial" for any of these public figures suggested the concept was not very accessible to that individual. Participants also completed the ASCAT with morphed White-to-Black race continuums. People who used multiracial terms (e.g., 'multiracial,' 'biracial,' 'Black and White') to describe mixed-race celebrities showed weaker automatic race categorization than people who solely use monoracial labels (e.g., 'White,' 'Latino') to classify these celebrities $F(1, 64)=8.19$, $p = .006$. These results give some credence to the idea that the availability of racial language may determine racial perception. However, we do not know whether giving people multiracial language or making it more accessible *causes* less categorical (or three-category) race perception. For example, if race is discussed in less discrete terms in the media, on governmental forms, in schools, etc. would people show weaker automatic race categorization than when exposed to traditionally delineated racial category terms?

Folk beliefs about genetic variation. Does the belief that races are mostly genetically distinct (rather than highly overlapping) predict automatic race categorization? Research examining these genetic folk beliefs about race showed that people who believe that races are genetically distinct from each other show greater implicit race bias on an evaluative priming paradigm using morphed mixed-race faces than people who believed that races are mostly genetically similar (Plaks, et al., 2012). This work can be extended beyond implicit race bias by examining whether individuals who believe in a greater extent of interracial genetic overlap have a stronger categorical perception of race itself. Whereas research in social cognition has begun to document how beliefs in such concepts as group "essence" (e.g., Haslam & Whelan, 2008) or "entitativity" (e.g., Lickel et al., 2000) promote stereotyping, this work would go an important step further by focusing specifically on people's folk beliefs about the genetic bases of the concept of race.

Liberalism/Conservatism. Research has demonstrated that political ideology influences explicit endorsement of race-related policies and implicit attitudes (Federico & Sidanius, 2002; Jost, Nosek, & Gosling, 2008). But is political ideology associated with differences in how race is actually perceived? Political conservatism is associated with intolerance for ambiguity (Jost, Glaser, Kruglanski, & Sulloway, 2003). Conservatives may be more likely than liberals to try to racially categorize others in an effort to curb ambiguity. Thus, we predict that conservatism (compared with liberalism) would be associated with perceiving ambiguous-race faces in a more categorical ("either/or") fashion, and in fact, there is some preliminary evidence in support of this hypothesis (Malahy, in prep).

Automatic race categorization and racial bias.

Automatic race categorization may be a strong predictor of racial bias. Minimal groups research has shown that explicitly dividing people into groups based on a trivial and irrelevant basis can increase intergroup biases (Tajfel, 1978; Tajfel & Turner, 1986). If some people exhibit weaker automatic race categorization, then they might also be less susceptible to the typically observed racial biases. For example, Devine's (1989) research showed that thinking about a group automatically activates group stereotypes (and vice versa). But weaker race categorization might preempt or weaken this effect. If people perceive race as more continuous, these stereotypic associations may also be less tied to groups. On the other hand, some researchers point out that the effect of categorization on bias has not been adequately shown (Park & Judd, 2005); they argue that researchers need to experimentally manipulate and *measure* increased categorization's effect on bias. With the ASCAT, we are prepared to fully examine this question.

Future studies could examine if the degree to which race is perceived categorically may moderate the effect of race on resource distribution (i.e., Blacks, relative to Whites, are often underpaid, offered fewer job opportunities, and given higher prices on car loans; Ayres &

Siegelman, 1995; Bertrand & Mullainthain, 2004; Black, Kolesnikova, Sanders, & Taylor, 2010) and race bias. Differences in the degree to which race is perceived categorically may also affect people's beliefs that multiracial people are less deserving of minority-based scholarships than their monoracial counterparts (Sanchez & Bonam, 2009) and affect perceptions of a multiracial candidate's eligibility for such resources (i.e., are they are perceived as a minority?). On a more basic level, automatic race categorization could influence who is affected by various social policies. For instance, a stronger or weaker race categorization may determine who is detained and searched according to social policies such as Arizona state's SB 1070 law, which requires non-U.S. Citizens to carry registration documents on their persons at all times.

Other social perception phenomena.

Understanding automatic race categorization may also help illuminate the processes underlying established social perception phenomena. For instance, stronger automatic race categorization may increase susceptibility to *the cross-race effect*, the idea that people are more likely to confuse members of other races than members of their own race (Malpass & Kravitz, 1969). Importantly, the cross-race effect typically does not examine faces from the middle of the racial continuum. One could argue that to obtain a cross-race effect, which presumes ingroup and outgroup members, race perception must be categorical. In theory, people with continuous race perception should not be susceptible to the cross-race effect, since they do not perceive the social world in terms of racial groups. A continuous race perceiver should confuse people equally along the racial continuum, irrespective of society's racial groupings. Thus, individual differences in categorical race perception may have implications for criminal justice. It is widely accepted among expert witnesses for legal trials that the cross-race effect (Malpass & Kravitz) is scientifically supported (Kassin, Tubb, Hosch, & Memon, 2001). In other words, White witnesses may be more likely to confuse one Black man for another Black man, than they are likely to confuse different White individuals. Thus, the present findings have implications for assessing the validity of eye-witness testimony, as well as suggesting that the accuracy of such testimonies may be improved if people's automatic race perception becomes less categorical.

Conclusion.

In sum, the ASCAT is a new methodological tool that enables a nuanced understanding of the ways in which social categories such as race are perceived. Future research with this tool may have important implications for how race is publicly portrayed (e.g. in the media, government, and education). Furthermore, demonstrating that people perceive race categorically even in response to continuously varying racial features provides a novel demonstration of how race is socially constructed. This, in turn, could lead to ways to reduce racial disparities in power, economic and social resources, and interpersonal biases.

REFERENCES

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

Ayres, I. & Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review, 85,* 304-321.

Baldus, D. C., Woodworth G., Zuckerman D., Weiner N. A., & Broffitt B. (1998). Racial discrimination and the death penalty in the post-Furman era: An empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review, 83*, 1638–1770.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217-239.

Berlin, B., & Kay, P. (1969) *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

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

Bivand, R., Leisch, F., & Maechler M (2008). pixmap: Bitmap Images ("Pixel Maps"). R package version 0.4-9, URL http://CRAN.R-project.org/package=pixmap.

Black, D., Kolesnikova, N., Sanders, S. G. and Taylor, L. J. (2010). The role of location in evaluating racial wage disparity. *Federal Reserve Bank of St. Louis Research Division Working Paper 2009-043B*. Available at SSRN: http://ssrn.com/abstract=1468435

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer, Jr. (Ed.), *Stereotype activation and inhibition: Advances in social cognition* (Vol. 11, pp. 1-52). Hilldale, NJ: Erlbaum.

Bonilla-Silva, E. (1999). The essential social fact of race. *American Sociological Review, 64*, 899-906.

Bornstein, M. H. & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 46*, 207-222.

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *A dual-process model of impression formation: Advances in social cognition* (Vol. 1, pp. 1-36). Hilldale, NJ: Erlbaum.

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

Burns, E. M., & Ward, W. D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*, 456-468.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81-117.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J.-M., (2002). Categorical perceptions of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience, 14*, 210-227.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition, 10*, 129-156.

Cantor, N., & Kihlstrom, J. F. (Eds.). (1981). *Personality, Cognition, and Social Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 12, pp. 3-52). New York: Academic Press.

Carroll, J., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, *35*, 283-319.

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,* 1314–1329.

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. In M. P. Zanna, (Ed.) , *Advances in experimental social psychology, Vol 39* (pp. 163-254). San Diego, CA US: Elsevier Academic Press.

Crisp, R. J., Hewstone, M., & Rubin, M. (2001). Does multiple categorization reduce intergroup bias?. *Personality and Social Psychology Bulletin, 27*, 76-89.

Cuddy, A. J. C., & Fiske, S. T. (1998). Doddering but dear: Process, content, and function in stereotyping of older persons. In F. J. Newmeyer, *Language Form and Language Function* (pp. 3-26), Cambridge: The MIT Press.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22-37.

Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin, 29*, 360-370.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P.G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87,* 876-963.

Ehret, G. (1987). Categorical perception of speech signals: Facts and hypotheses from animal studies. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 301-331), Cambridge: Cambridge University Press.

Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*, Cambridge: Cambridge University Press.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227-240.

Fazio, R. H., Jackson, J. R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality & Social Psychology*, *69,* 1013-1027.

Federico, C. M., & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: The antecedents and consequences of "principled objections" to affirmative action. *Journal of Personality and Social Psychology, 82*, 488-502.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357-411). Boston: McGraw-Hill.

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. *Advances in Experimental Social Psychology, 23,* 1–73.

Freeman, J., Pauker, K., Apfelbaum, E., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46,* 179-185.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201-225.

Gaillard, A. W. K., & Verduin, C. J. (1985). Comparisons across paradigms: An ERP study. In M.

I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 375–388). Hillsdale, NJ: Erlbaum.

Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology, 94*, 292-306.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178-200.

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102,* 4-27.

Hall, N. R., & Crisp, R. J. (2005). Considering multiple criteria for social categorization can reduce intergroup bias. *Personality and Social Psychology Bulletin, 31,* 1435-1444.

Harnad, S. (2003). Categorical perception. *Encyclopedia of Cognitive Science.* Nature Publishing Group/Macmillan. Retrieved from http://cogprints.org/3017/1/catperc.html.

Harnad, S. (Ed.). (1987). *Categorical Perception: The Groundwork of Cognition*, Cambridge: Cambridge University Press.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception and Psychophysics, 32*, 409-418.

Haslam, N. & Whelan, J. (2008). Human natures: Psychological essentialism in thinking about differences between people. *Social and Personality Psychology Compass, 2,* 1-14..

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology, 21*, 517-528.

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). Overview of race and Hispanic origin: 2010. *2010 Census Brief, C2010BR-02.*

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*, 616-626.

Ito, T. A. & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, and Behavioral Neuroscience, 5*, 21-36.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129,* 339-375.

Jost, J. T., Nosek, B. A. & Gosling, S. D. (2008). Ideology: Its resurgence in social, personality, and political psychology. *Perspectives on Psychological Science,* 3,126-136.

Kaiser, C. R., Drury, B., Malahy, L. W. & King, K. M. (2011). Nonverbal asymmetry in interracial interactions: Strongly identified Blacks display friendliness, but White respond negatively. *Social Psychology and Personality Science, 2*, 554-559.

Kassin, S., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). The `general acceptance' of psychological research on eyewitness testimony: An issue revisited 10 years on. *American Psychologist*, *56*, 405-416.

Kay, P., & Regier, T. (2006). Language, thought, and color: Recent developments. *Trends in Cognitive Sciences, 10*, 51-54.

Kite, M. E., Deaux, K, & Miele, M. (1991). Stereotypes of young and old: Does age outweigh gender? *Psychology and Aging, 6*, 19-27.

Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults: A meta-analysis. *Psychology and Aging, 3,* 233-244.

Kruglanski, A. W. & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and “freezing”. *Psychological Review, 103,* 263-283.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling: quantitative applications in the social sciences*. London: Sage.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20*, 215-225.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality & Social Psychology, 72,* 275-287.

Levin, D. T. & Angelone, B. L. (2002). Categorical perception of race. *Perception, 31,* 567-578.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception and Psychophysics, 62*, 386-401.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Lingle, J., Altom, M., & Medin, D. (1984). Of cabbages and kings: Assessing the extendibility of natural object concept models to social things. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of Social Cognition* (pp. 71 -118). Hillsdale, NJ: Lawrence Erlbaum Associates.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 732-753.

MacLin, O. H., & MacLin, M. K. (2011). The role of racial markers in race perception and racial categorization. In R. B. Adams, N. Ambady, & K. Nakayama, & S. Shimojo (Eds.), *The Science of Social Vision* (pp. 321-346). New York: Oxford University Press.

MacLin, O. H. & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7,* 98-118.

Maclin, O. H., Peterson, D. J., Hashman, C., & Flach, N. (2009). PsychoPro 2.0: Using multidimensional scaling to examine the perceptual categorization of race. *Behavior Research Methods, 41,* 668-674.

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. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92*, 239-255.

Macrae, C. N., 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*, 37-47.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Maddux, W. W., Galinsky, A. D., Cuddy, A. J., & Polifroni, M. (2008). When being a model minority is good…and bad: Realistic threat explains negativity toward Asian Americans. *Personality and Social Psychology Bulletin, 34,* 74-89.

Malahy, L. W. (in prep). Black and White or shades of gray: Individual differences in automatic race categorization. (Doctoral dissertation).

Malahy, L., Sedlins, M., Plaks, J. E., & Shoda, Y. (2010). Black, White, or shades of gray? Racial labeling of Barack Obama predicts implicit race perception. *Analysis of Social Issues and Public Policy*, *10*, 207-222.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality & Social Psychology, 13,* 330-334.

Martin, D. & Macrae, C.N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology, 37, 806-816.*

Massaro, D. W. (1987). Categorical partitioning: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 254-286), Cambridge: Cambridge University Press.

Massaro, D.W. (1998). Categorical perception: Important phenomenon or lasting myth? In R.H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Congress of Spoken Language Processing* (pp. 2275-2279, Vol. 6). Sydney, Australia. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.3205&rep=rep1&type=pdf

McClelland, J.L., Fiez, J.A., & McCandliss, B.D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology and Behavior, 77*, 657-662.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory and Cognition, 6*, 462-472.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*(12), 1469-1481.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*, 113-138.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1–21.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

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

Pastore, R. E. (1987) Categorical perception: Some psychophysical models . In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 29-52), Cambridge: Cambridge University Press.

Pauker, K. & Ambady, N. (2009). Multiracial faces: How categorization affects memory at the boundaries of race. *Journal of Social Issues, 65*, 69-86.

Peery, D. & Bodenhausen, G. V. (2008). Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science, 19,* 973-977.

Perlman, M. A., & Krumhansl, C. L. (1996). An experimental study of internal interval standards in Javanese and Western musicians. *Music Perception, 14*, 95-116.

Pisoni, D. (1971). *On the nature of categorical perception of speech sounds.* University of Michigan). *ProQuest Dissertations and Theses*, 141 p. Retrieved from http://search.proquest.com/docview/302500955?accountid=14784

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13*, 253-260.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55*, 328-333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15*, 285–290.

Plaks, J., Malahy, L. W., Sedlins, M. & Shoda, Y. (2012). Folk beliefs about human genetic variation predict discrete versus continuous implicit racial categorization and evaluative bias. *Social Psychological and Personality Science, 3*, 31-39.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Quinn, P. C. (2002). Category representation in young infants. *Current Directions in Psychological Science, 11*, 66-70.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Ed.* Thousand Oaks: Sage Publications.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1-20.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 1*, 532-547.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

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

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (Vol. 1). London: Academic Press.

Sanchez, D. T. & Bonam, C. M. (2009). To disclose or not to disclose biracial identity: The effect of biracial disclosure on perceiver evaluations and target responses. *Journal of Social Issues, 65,*129-149.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.

Shepard, R.N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part I. *Psychometrika, 27*, 125-140.

Shepard, R.N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part II. *Psychometrika, 27*, 219-246.

Slawinski, E. B. (1999). Acquisition of /r-l/ phonemic contrast by Japanese children and adults. In M. Pinto, J. Veloso, and B. Maja (Eds.), *Psycholinguistics on the threshold of the year 2000. Proceedings* (pp. 583-590). Porto : Faculdade de Letras da Universidade do Porto

Smith, E. E., Jonides, J., Marshuetz, C. & Koeppe, R. A. (1998). *Proceedings of the National Academy of Sciences, 95*, 876–882.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge: Harvard University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81,* 214-241.

Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 440-463.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62*, 207-218.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

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

Tajfel, H. (1978). *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations.* Oxford: Academic Press.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1,* 149-178.

Tajfel & Turner (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds). *Psychology of intergroup relations*, (pp. 7-24). Chicago: Nelson-Hall Publishers.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*, 778-793.

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science, 24*, 363-396.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Willadsen-Jensen, E. C. & Ito, T. A. (2006). Ambiguity and the timecourse of racial perception. *Social Cognition, 24,* 580-606.

Willadsen-Jensen, E.C. & Ito, T.A. (2008). A foot in both worlds: Asian Americans' perceptions of Asian, White, and Racially Ambiguous Faces. *Group Processes and Interpersonal Relations, 11,* 182-200.

Wilson, M. (1987). Brain mechanisms in categorical perception. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 387-420), Cambridge: Cambridge University Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*, 7780-7785.

Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, trans.). Oxford: Blackwell.

Wright, A. A., & Cumming, W. W. (1971). Color-naming functions for the pigeon. *Journal of the Experimental Analysis of Behavior, 15*, 7-17.