Title
People Construal: Social Categorization and Trait Evaluations of Visually Perceived Groups

Author
Alt, Nicholas

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

People Construal: Social Categorization and Trait Evaluations of Visually Perceived Groups

A dissertation submitted in partial satisfaction of the

Requirements for the degree Doctor of Philosophy

In Psychology

by

Nicholas Peter Alt

2019

ABSTRACT OF THE DISSERTATION

People Construal: Social Categorization and Trait Evaluations of Visually Perceived Groups

by

Nicholas Peter Alt

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2019

Professor Kerri Johnson, Chair

Our social world is populated by groups—students in classrooms, co-workers in teams, soldiers in platoons, and more. Social psychological research on person construal has established the processes and consequences of perceiving an individual, yet, we know less about our perception of groups. Here, I integrate person construal theory with vision science findings on ensemble perception to outline a framework of people construal, or how we make inferences and judgments about visually perceived groups. Overall, I propose that people construal relies on both social categories (e.g., a group's gender ratio) and visual cues to within-category variability (e.g., facial masculinity or femininity). Study Set 1 examined whether perceivers were sensitive to a group's gender ratio and whether this gender ratio impacted evaluative judgments. Studies 1 and 2 showed that within a half second, perceivers were relatively accurate in estimating the number of men within a twelve-person group, and that as a group's gender ratio shifted from men to women participants associated the group with more threat, measured both explicitly and implicitly. Study 3 revealed that group threat evaluations were derived from participants perceived number of men and not inferences about men's perceived cohesiveness within the group. Study 4 underscored the influence of visual cues to men's gender typicality on group evaluations. Results showed that groups of masculine men, compared to feminine men, were rated as having more men and as more threatening. Study Set 2 shifted to questions of social category activation and social categorization. Study 5 revealed that group gender ratio and men's gender typicality interacted to influence subsequent social category activation. Specifically, groups of majority masculine men facilitated activation of the category "man" to a greater degree than groups of majority feminine men. Study 6 revealed that participants categorized majority masculine men groups as "majority men" with higher accuracy and in a faster, more direct manner, compared to majority feminine men groups. Together, these results build a framework of people construal, advancing research beyond our perception of the individual to reveal insights into the role social categories and visual cues play in our perception of groups.

This dissertation of Nicholas Peter Alt is approved.

Hongjing Lu

Carolyn M. Parkinson

Margaret Joan Shih

Kerri L. Johnson, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Study Set 1 include data previously published in Alt, N. P., Goodale, B. M., Lick, D. J., & Johnson, K. L. (2019). Threat in the company of men: Ensemble perception and threat evaluations of groups varying in sex ratio. *Social Psychological and Personality Science, 10*(2), 152-159. The co-authors on this paper helped formulate the study idea and edited the manuscript, but data collection and analysis were completed by the first author.

This dissertation was made possible by funding from the Department of Defense, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.

I would like to express my immense gratitude to the following people who supported my academic journey. First, to my academic advisor, Dr. Kerri Johnson. Your guidance and unwavering support helped me prove to myself that I could be an academic and scholar. I am so grateful for your nurturing, both professional and personal, and will treasure my time at UCLA. You showed me how to be a fantastic advisor, teacher, mentor, and researcher, lessons I hope to pass onto my own students.

To my dissertation committee—Drs. Hongjing Lu, Carolyn Parkinson, and Margaret Shih. I sincerely thank you for your time and devotion. Your questions and feedback produced puzzles to mull over and future directions to investigate for years to come.

I also want to thank my friends for reminding me that work is not central and that community is vital. Special call out to Ryan Tieu for all of the stats advice and nights of WeHo and games. And, thank you Sanford, you bring so much light to my life.

Lastly, thank you Mom, Dad, Jerry, Jen, Rob and Gerry. I am blessed to have your unconditional love and support.

CURRICULUM VITAE

Education

2018	C.Phil., Psychology, University of California, Los Angeles
2014	M.A., Experimental Psychology, The College of William & Mary
2009	B.A., Psychology and East Asian Studies, Wesleyan University

Honors and Awards (Selected)

2018	Division 45 Small Research Grant (American Psychological Association)
2016	Small-Scale Events Grant (Society for the Psychological Study of Social Issues)
2015	National Defense Science & Engineering Graduate Fellowship (Dept. of Defense)
2014	Student Travel Award (Society of Personality and Social Psychology)
2014	Grant-in-Aid Award (Society for the Psychological Study of Social Issues)
2014	East Asia and Pacific Summer Institute (National Science Foundation)
2014	Peter J. Corpus Education Grant (Gay Asian Pacific Support Network)
2014	Distinguished University Fellowship (University of California, Los Angeles)
2014	S. Laurie Sanderson Award for Undergraduate Mentoring (William & Mary)
2013	Clara Mayo Grant (Society for the Psychological Study of Social Issues)
2012	Graduate Research Grant (William & Mary)
2012	William & Mary Graduate Studies Fellowship (William & Mary)
2009	Phil Beta Kappa
2009	Thorndike Award for Excellence in Research (Wesleyan University)

Peer-Reviewed Publications

Alt, N. P., Goodale, B. M., Lick, D. J., & Johnson, K. L. (*in press*). Threat in the company of men: Ensemble perception and threat evaluations of groups varying in sex ratio. *Social Psychological and Personality Science.*

Alt, N. P., Chaney, K. E., & Shih, M. J. (*in press*). “But that was meant to be a compliment!”: Evaluative costs of confronting positive racial stereotypes. *Group Processes & Intergroup Relations.*

Goodale, B. M., Alt, N. P., Lick, D. J., & Johnson, K. L. (2018). Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *Journal of Experimental Psychology: General, 147*(11), 1660-1676.

Alt, N. P., Park, S. H., & Schug, J. (2018). The ironic impact of intergroup apologies on intergroup attitudes: Understanding the role of perceived power for post-apology outgroup attitudes. *Asian Journal of Social Psychology, 21*, 42-53.

Schug, J., Alt, N. P., Lu, P. S., Gosin, M., & Fay, J. (2017). Gendered race in mass media: Invisibility of Asian men and Black women in popular magazines. *Psychology of Popular Media, 6*(3), 222-236.

Schug, J., Alt, N. P., & Klauer K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology, 56*, 121-125.

<u>Presentations and Talks (Selected)</u>

Alt, N. P. (2019). *Social vision of groups: Small ensembles to large social networks.* Chaired symposium presented at the 20th Annual Convention of the Society of Personality and Social Psychology. Portland, OR.

Alt, N. P., Parkinson, C., Kleinbaum, A. M., & Johnson, K. L. (2019, February). *The face of social networks: Naïve observers accurately judge social network characteristics from faces.* Talk presented at the 20th Annual Convention of the Society of Personality and Social Psychology. Portland, OR.

Alt, N. P., Goodale, B. M., Lick, D. J., & Johnson, K. L. (2016, September). *The social vision of groups: Ensemble coding of group gender composition and threat evaluations.* Paper presented at the annual Person Memory Interest Group meeting. Lake Arrowhead, CA.

Alt, N. P., & Dickter, C. L. (2013, November). *Empowered to confront: Embodied power primes and confronting sexist comments*. Talk presented at the annual Southeastern Society of Social Psychologists. Augusta, GA.

Alt, N. P., & Dickter, C. L. (2013, March). *How do we perceive the intersection of race and sexuality? Evidence from social psychology.* Talk presented at Critical Junctures: Exploring the intersections of race, gender, and disability. Emory University, Atlanta, GA.

Morawski, J., & Alt, N. P. (2009, November). *The multiple psychologies of subjectivity: Accountings of experiments in mid-century America.* Paper presented at the annual meeting of History of Science Society. Phoenix, AZ

Chapter 1: Introduction

Take a moment and imagine a crowded sports arena. In your mind's eye, you might conjure a fan cheering on their favorite team. Or a shirtless man with a large letter "L" written on his exposed belly. Or perhaps, the coach yelling about an unfair call. When you think "crowded", however, these specific images are probably not what you first imagined. Instead, you likely conjured the mass of people wearing their team's colors, the slightly skewed gender ratio of more men than women, or the angry emotional reaction of the fans after the unfair call. Much of person perception research has largely centered on the first set of impressions, those made about individual people. This focus, while vital to understanding how we come to perceive and make judgments about others, tells us less about the more common sight, that of individuals in groups. Indeed, our perceptual world is filled with people interacting with people—students in classrooms, protesters in the street, coworkers in work teams, and more. Thus, while groups are a foundational topic in social psychology, underlying a diverse array of research from intergroup relations (Heine, 2010; Richeson & Sommers, 2016; Yzerbyt & Demoulin, 2010) to identity (Swann & Bosson, 2010; Tajfel, 1974), surprisingly little is known about *visual people construal,* or how we perceive groups and the downstream inferences we make about these group percepts.

Social cognition, and in particular person construal research, has established an extensive body of work regarding the processes by which we perceive others—the integration of the visual percept (e.g., face/body) with social category knowledge (e.g., stereotypes) which comes to inform our judgments about others (for reviews see Fiske & Neuberg, 1990; Johnson, Lick, & Carpinella, 2015; Kawakami, Amodio, & Hugenberg, 2017; Macrae & Bodenhausen, 2000; Tajfel, Billig, Bundy, & Flament, 1971). While similar processes likely operate for groups, it is important to acknowledge that groups, due to their complexity, pose a unique set of challenges. For instance, can we perceive social category differences within groups? Are our trait judgments of groups as a whole sensitive to social category variability within, and even more nuanced visual cue variation (e.g., gender typicality)? How do these factors, social category variability and visual cues, affect how we categorize groups?

Social psychologists are starting to turn their eyes towards such questions pertaining to group perception and people construal (Phillips, Weisbuch, & Ambady, 2014). In large part, this interest is spurred by recent surprising findings from vision science on the phenomenon of ensemble perception: our ability to rapidly, and without much conscious attention, extract accurate summary statistics such as the mean and variance from a set of visually perceived objects (Alvarez, 2011; Whitney & Leib, 2018). Overall, by incorporating ensemble perception research with social psychological findings and theory, I am poised both in terms of methodology and extant literature, to advance our knowledge of how we perceive people.

In my dissertation, I outline a series of studies probing the processes and consequences of visual *people construal*, how we perceive groups of people and the downstream evaluative and social categorical judgments we make about such groups. To provide theoretical background, I begin by reviewing social psychology and vision science research, setting the stage to integrate these fields. In Chapter 2, I examine the extensive theory and research on person perception and person construal—that is how we visually perceive and represent single individuals. The empirical and theoretical insights from this area form the foundation for extrapolating beyond the individual to understand the group. In Chapter 3, I discuss findings on ensemble perception, highlighting how vision science approaches can inform our understanding of people construal, while also noting gaps within this work, notably the consequences of group perception as it relates to social inferences and judgments. Finally, in Chapter 4, I outline a set of hypotheses related to people construal and propose a framework for advancing our theoretical and conceptual knowledge about group percepts and how we come to form judgments and evaluations about visually perceived groups.

Chapter 2: Person Construal: Social Categorization and Social Vision

In building towards an integrated theory of people construal it is informative to review the literature on person construal. While many sensory processes underlie person construal, one dominant medium by which we experience others is sight, thus I will focus my review on how our visual perception of others (typically from the perception of faces and bodies) leads to our evaluations and judgments about others. Two key areas of research are critical to developing a *people* construal framework, namely the efficiency and pervasive use of social categories and the recent integration of early perceptual processes (e.g., visual cues) into person construal processes under the field of social vision.

Social Categorization

Our daily lives present us with the complex challenge of evaluating and inferring other people's emotional states, motives, and intentions. One mechanism to alleviate this complexity is the use of social categories, which allow us to integrate previously learned knowledge with newly presented people. To this end, research on person perception has long focused on social categorization. In his seminal work, Allport (1954) argued that social categorization was an inevitable consequence of living in a complex social world and that social categories allowed us to, "…*select* and *interpret* our impressions of the surrounding world" (Allport, 1954, p. 165). In this vein, social categories serve us by reducing the cognitive effort of forming separate and individuated impressions about each newly encountered person, as we apply and rely on learned associations, beliefs, and attitudes about the social categories to which a person belongs.

Early researchers also argued that social categories, often studied with linguistic or semantic descriptions of people and groups (i.e., category labels like “men”, “French”, “White”), were the starting point for attitudinal differences about groups, engendering the potential for prejudice and discrimination (Allport, 1954; Heider, 1958; Warr & Knapper, 1968). With the subsequent integration of more cognitive processes such as attention, memory, and learning, social psychologists revealed the prescience of Allport’s (1954) position about the necessity and ease by which we think about others in terms of social categories (Brewer, 1988; Fiske & Neuberg, 1990; Fiske & Taylor, 1991; Hamilton, 1981). Indeed, an abundance of evidence suggested that we extract and apply social category information efficiently and automatically. For instance, results from semantic priming tasks showed that a “prime” stimulus could activate social category knowledge, subsequently impacting a “target” stimulus (e.g., the category label “Blacks” led to faster response times to stereotypical words such as “musical”; Dovidio, Evans, & Tyler, 1986). It is even the case that social categorical stimuli, presented below the threshold of conscious perception, still affects judgments in a manner consistent with stereotypes about the primed group (Devine, 1989; Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997). This research was further extended by results demonstrating that task relevance had minimal influence on social category activation. For example, in a lexical decision task, where the goal was irrelevant to the social category (i.e., is the target a word or not a word), facilitation for prime-target word pairings (e.g., the word women and the word nurse) still emerged (Macrae, Bodenhausen, & Milne, 1995). Evidence also accumulated suggesting that the use of stereotypes, activated from social categorical thinking, was deemed effective at reducing cognitive demands and acted as an information consolidating tool. Indeed, in one study, researchers found that stereotypes allowed for better performance on a split attention task (i.e., monitoring audio and forming impressions from visually perceived labels, Macrae, Milne, & Bodenhausen, 1994). Across multiple studies and paradigms, results coalesced around the idea that we think in terms of social categories and that the stereotype activation which can arise from viewing others in such terms is a difficult to escape, default way of viewing others (Fiske & Neuberg, 1990; Hamilton, 1981).

Still, the actual object of social categorization—the person—was largely absent from these early studies, a fact noted by Macrae and Bodenhausen (2000, p. 98), "…given Allport's (1954) assertion that category activation follows the registration of a triggering stimulus—specifically, another person—why is it that researchers have typically used verbal labels to investigate this process [social category activation]." Indeed, this challenge posed a serious issue to researchers. If I am studying the efficiency of social category activation regarding "women", however, prime participants with the word "women", did social categorization occur? To assess these questions, researchers needed to shift their investigations away from group labels and instead use visual cues of the person.

Notably, a few studies tried to assess social categories without group labels. For example, the Who-Said-What memory paradigm was developed to tap the spontaneous use of social categories, thus highlighting the truly automatic nature by which social categories emerge. In these studies, participants learned a series of statements by speakers and were later given a surprise memory test in which they recalled which speaker said which statement. Results consistently demonstrated that participants made more within-category errors (e.g., confusing which of two women said the statement) compared to between-category errors (e.g., confusing whether it was a man or a woman who said the statement). This pattern of results was taken as evidence that social categories were being used to guide memory and did not explicitly rely on the use of category labels (Stangor, Lynch, Duan, & Glass, 1992; Taylor, Fiske, Etcoff, & Ruderman, 1978). While certainly evidence that social categories can arise spontaneously, these findings produced a roundabout way of assessing social categorization, first through learning, then memory, and finally recall (Klauer & Wegener, 1998).

Buttressing this evidence for the automatic extraction of social category information from targets, social neuroscience findings suggest that early visual processing is attuned to social category differences extracted from faces. For example, electromyography (EEG) studies revealed that quickly after the visual presentation of a Black face (~170 ms), White participants showed different cortical activation or event-related potentials (ERPs) in visually oriented areas of the brain (Ito & Urland, 2003). Critically, African American participants also showed this same pattern but to White faces, suggesting ingroup/outgroup discrimination via race (Dickter & Bartholow, 2007). Similarly, functional brain imaging (fMRI) findings indicated that racial outgroup faces led to activation in the amygdala, an area of the brain associated with negative and threat-relevant stimuli (Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005). And yet, it was shown that participants attitudes and task relevance moderated amygdala activation, with those who held more positive racial attitudes or who viewed faces under conditions where race was not salient, having lower activation pattern differences between White and Black faces (Phelps et al., 2000; Wheeler & Fiske, 2005). Thus, while these studies demonstrated that social categories differentially affected early face processing outcomes, they also reflected the role of top-down influences such as task instruction or individual attitudes, in moderating these "uncontrolled" responses.

This non-exhaustive review of early social categorization research foregrounds the "cognitive miser" (Fiske & Taylor, 1991) approach to how we think about others—the fact that social categories offer a mental shortcut and in turn represent an efficient and inescapable part of person construal. Further, dual-process theories of person construal delineated how with sufficient attention and cognitive resources we may activate more nuanced processes of individuation, yet our default is to use social categories when little is known about the other or upon first encounter (Brewer & Feinstein, 1999; Fiske, Lin, & Neuberg, 1999). Applied to *people* construal, the lesson from early person construal research is that social categories will likely be a dominant characteristic of how we think about groups, and further that the social categories represented in a group will similarly activate knowledge structures such as stereotypes which will inform our trait and evaluative impressions of groups. These insights provide a foundation for understanding the operation of social categories within groups however a challenge for both person and people construal is how these social categories themselves arise, that is how does the act of social categorization occur given a particular person or group of people?

Social Vision and Person Construal

To address the challenge of how our visual perception of others translates into social categorization and subsequent activation of prior knowledge (i.e., stereotypes, individuating information, learned associations, etc.), a new theoretical framework and area of study emerged—social vision (Adams, Ambady, Nakayma, & Shimojo, 2011; Balcetis & Lassiter, 2010; Johnson et al., 2015). Social vision sought to advance our theoretical and empirical understanding of the antecedent processes associated with social categorization; reframing past models to place social categorization not as the beginning of a long line of person construal processes (Fiske & Neuberg, 1990) but rather as a vital process in need of investigation itself. These principles were cogently argued and summarized in Freeman and Ambady (2011)'s Dynamic Interactive Theory of Person Construal, which explicitly linked bottom-up visual cues (e.g., features, skin tone) with top-down cognitive processes (e.g., stereotypes, attitudes) and placed these two factors in a dynamic and iterative process, with both influencing the ultimate social category judgment. Borrowing theories from cognitive psychology, Freeman and Ambady (2011) developed a dynamic connectionist network model, in which visual cues provide stimulus input which in turn competitively activate different social categories. As these social categories become activated, top-down processes such as motivations come online which enhance or dampen excitatory connections across potential social category options. Importantly, these connections iteratively compete over time, until a stable social category judgment emerges.

Furthermore, Freeman and Ambady (2011), and social vision more broadly, situated social categorization in terms of its more proximal reasons for being, derived from the ecological theory of social perception (McArthur & Baron, 1983). Building on even earlier ideas from Gibson (1986), it was argued that our visual perception is largely driven by functionality, thus social percepts are viewed in terms of affordances which allow us to make inferences about perceived targets. Foregrounding this concept renewed interest in understanding person perception from the standpoint of why our perceptions matter, providing an essential lens for the functional purpose of our everyday judgments and decisions about others.[1] Additionally, in acknowledging the temporal dynamics, the Dynamic Interactive Theory of Person Construal

[1] Indeed, a quick review of google scholar citations shows that McArthur & Baron (1983) has been cited more than 1,016 times, yet most of these citations come between 2000 and 2018 (647) compared to 1983-1999 (362).

theorizes that our judgments can rapidly and efficiently update with changes in our social environment (e.g., momentary facial expressions).

Another key insight from this model is the influence of top-down factors, giving weight to our motivations and the social context in which we perceive others. For instance, we might view a person’s gender, whether someone is categorized as a man or a woman, differently depending on whether it is dark and we are alone or fearful (Johnson, Iida, & Tassinary, 2012). In this case, informed by insights from Error Management Theory (Haselton & Buss, 2000) seemingly irrational biases in social categorization make more “rational” sense, as it is better to err on the side of caution since, on average, men represent greater threat than women (Archer, 2004). Overall, the Dynamic Interactive Theory of Person Construal positioned social categorization not as the starting point but rather as an entire process itself, affected by contextual and motivational factors and in need of exploration.

Two specific themes from social vision and models like the Dynamic Interactive Theory of Person Construal are relevant to my proposal of people construal: 1) the impact of within-category variability (e.g., racial phenotype/gender typicality) on social categorization outcomes and evaluative judgments, and 2) the fact that stimuli features (e.g., hairstyle, gendered cues) alone can compel and even override social categorization processes. The rest of this chapter is devoted to discussing and reviewing literature on these factors.

Visual Cues to Within-Category Variability and Person Construal

Monolithic categories lend themselves to the thinking that there is only one prototypical exemplar. For instance, asking participants to read the label “Black Person” evokes the image of an individual with specific features that match a prototype (e.g., a dark-skinned person, likely a man). We know, however, that racial categorization often others those who only look marginally different (Maddox, 2004). Across multiple studies, and a variety of social categories, researchers have found that this within-category variability in visual cues profoundly influences person construal processes. For instance, Blair, Judd, Sadler, and Jenkins (2002) demonstrated that a target's visually perceived degree of Afrocentricity (e.g., the degree to which one has stereotypical features of a Black person) produced strikingly different outcomes in terms of stereotyping and category accessibility, with high Afrocentric targets evoking greater stereotype activation and category facilitation. Maddox and Gray (2002) also found, using a Who-Said-What paradigm, that perceivers extracted and utilized skin tone in a similar manner as social categories, with more within-"group" (i.e., dark skin tone) than between-"group" (i.e., dark and light skin tones) errors. Together, these findings presented a formidable argument for effects of, and need to account for, within-category differences in visual cues.

Follow-up studies on Afrocentric phenotypic cues continued to reveal the highly consequential real-world impact that these visual cues have on social categorization outcomes. Researchers found that highly phenotypic Black men received harsher criminal sentences and a higher proportion of death sentences compared to less phenotypic Black men (Blair, Judd, & Chapleau, 2004; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Even the visual assessment of physical size was tethered to the degree of Afrocentricity within the face, with people who had more Afrocentric features viewed as more formidable (Wilson, Hugenberg, & Rule, 2017).

Beyond race, the influence of within-category cue variability to gender (i.e., masculinity and femininity) also greatly affects person construal and downstream trait judgments. While not visual in nature, Deaux and Lewis (1984) noted that when sex (i.e., male and female) and gender typicality (i.e., whether someone is masculine or feminine) were crossed, gender typicality was predictive of other social categories (i.e., gay or straight) and gendered traits. This interactive influence of sex and gender typicality afforded novel insights into the construal of a feminine man or a masculine woman. Specific to the face, Friedman and Zebrowitz (1992) demonstrated that manipulating men's faces to appear more babyfaced, (i.e., large eyes and a higher forehead to chin ratio) and women's faces to appear less babyfaced, flipped gender typical trait ratings (i.e., men were rated more warm and caring, and women more dominant). Facial gender typicality was also tethered to ratings of dominance and leadership, with more masculine looking men and women rated higher on these traits than feminine looking men and women (Sczesny, Spreemann, & Stahlberg, 2006).

Other facial sexually dimorphic cues can also vary within gender. For instance, facial width to height ratio (fWHR) is a structural cue to the face measured by dividing face width (e.g., the distance from ear to ear) by face height (e.g., the distance from the upper lip to the upper eyelid). Numerous studies have shown that across both real photos where fWHR is measured, and digital images where fWHR is manipulated, perceivers consistently rate high fWHR individuals as more masculine, as well as more aggressive, threatening, and physically formidable than low fWHR individuals (Carré, McCormick, & Mondloch, 2009; Geniole, Denson, Dixson, Carré, & McCormick, 2015; Geniole, Molnar, Carré, & McCormick, 2014; Třebický, Fialová, Kleisner, & Havlíček, 2016; Zilioli et al., 2015). Overall, a recent meta-analysis found a relatively strong average correlation between threat perceptions and unmanipulated faces (r = .46) and a small-to-medium correlation for manipulated faces (d = .41), supporting the notion that perceivers use fWHR variability to assess threat judgments (Geniole, Denson, Dixson, Carré, & McCormick, 2015).

In summary, across social categories (e.g., race and gender), visual cues to within-category variability (e.g., race and gender typicality) produced different outcomes in relation to person construal processes, both forming their own category subsets (Maddox & Gray, 2002) and shifting evaluative judgments (Friedman & Zebrowitz, 1992; Geniole et al., 2015). A critical insight from social vision models was the explicit goal of examining the impact of these visual cue differences (Johnson et al., 2015). I argue that such within-category variability likely also informs our understanding of people construal, with important downstream consequences for the evaluations of groups.

The Impact of Visual Feature Cues on Social Categorization

The fact that visual cues to within-category variability produced different outcomes associated with social category activation and stereotyping (Blair et al., 2002; Blair et al., 2004; Friedman & Zebrowitz, 1992; Livingston & Brewer, 2002; Sczesny et al., 2006) led researchers to question the role of visual features themselves, both how they engendered social judgments and category activation and whether they alone were sufficient to activate social categories (e.g., priming). Early evidence for the differential effects of visual feature cues on social categorization, emerged from percepts of the body. Prior work on walk-motion, displayed via point-light walkers, found that these stimuli were sufficient to compel accurate sex categorization (Kozlowski & Cutting, 1977). However, percepts of the body provide multiple cues from which sex categorizations can emerge such as walk motion and body morphology (i.e., waist-to-hip ratio). Johnson and Tassinary (2005) investigated these cues in combination and found that perceivers mostly derived their sex categorization judgments from morphology rather than body motion. Indeed, visual attention was focused on the waist-to-hip ratio when making sex categorization judgments, and this focus was dramatically reduced when participants were explicitly told the sex of the walker. Critically, however, walk motion informed judgments of gender typicality, which in turn mediated sex categorizations. Thus, accurate sex categorization from walk motion likely arose from inferred judgments related to gender typicality (i.e., the walker looks masculine, and thus must be a man). These findings show that visual cues play a vital role in social categorization judgments and that categorization does not arise spontaneously in a discrete unambiguous fashion but rather relies on lower-level visual cues (Johnson & Freeman, 2010).

Similar findings that visual features alone can impact social categorization judgments also emerged for percepts of race. Specifically, MacLin and Malpass (2001) demonstrated that race categorization varied as a function of hairstyle cues. In their study, they generated ambiguous faces that overlapped in facial cues to the categories Hispanic and Black. They found that applying different hairstyles to the same face, compelled perceivers to categorize the face as either “Hispanic” or “Black”. More striking, results indicated that participants reported that faces with a “Black” hairstyle had a darker complexion, even though visual cues to luminance were held constant. These findings demonstrated that a single feature—hairstyle—dramatically influenced visual perception and social categorization of a target. Taking these findings a step further, Levin and Banaji (2006) presented participants with faces that were equal in luminance however morphed along a continuum of Black and White facial features. When viewing these stimuli, a compelling visual illusion emerged, whereby White morphed faces were perceived as significantly lighter than Black morphed faces, even though luminance was unchanged. Again, facial features affected how we perceive others. By explicitly modeling these factors—*sensory input* providing the basis for varied social category activation (e.g., Black versus Hispanic) and *top-down influences* (e.g., prior experience of seeing specific hairstyles on Black individuals and

Hispanic individuals) leading to different category judgments—social vision models argued against social categorization being a one and done process and instead saw social categorization as a dynamic, iterative, and motivated process that could lead to varied yet predictable category judgments.

The studies above revealed the role visual features played in both trait and category judgments about others, however stopped short of testing whether visual features alone could activate social category knowledge. Given MacLin and Malpass (2001)'s finding that hair cues could shift category judgments, it is not surprising that researchers sought to test whether this feature, by itself, could elicit category activation. In a series of studies, Macrae and Martin (2007) found that this was the case. In a sequential priming task, they demonstrated that hairstyle cues alone were sufficient to activate gender categories (i.e., short hair facilitated "men" category activation, long hair facilitated "women" category activation). Furthermore, they crossed hairstyle with actual gender of the face, producing congruent and incongruent pairings (i.e., men with long hair and women with short hair) while also manipulating prime duration: short (200 ms) or long (20 ms). Results showed that, at the long prime duration, hairstyle cues overrode the actual gender of the face, with men with long hair and women with short hair priming the "incorrect" category, women and men respectively. Overall, these studies demonstrated that visual features, under certain constraints such as a brief glance, produced faulty category activation.

While Macrae and Martin (2007)'s finding suggested that social category activation was misled by visual features (i.e., hairstyle), we often do not see individuals for a mere 20 ms. Their results also belie the fact that we generally are very accurate with gender categorizations, especially when given sufficient viewing time (Wild et al., 2000). In reconciling these issues, social vision models of person construal (Freeman & Ambady, 2010; Johnson & Freeman, 2010) articulate that social categorization is an iterative process, that continually updates. In this way, even though the final categorization (e.g., that person is man or a woman) is generally accurate, it may be that our decision process unfolds different depending on visual cues such as masculinity and femininity. To investigate this idea, Freeman, Ambady, Rule, and Johnson (2008), had participants make gender categorization judgments from stimuli that again crossed hairstyle cues with gender, as well as stimuli morphed to be more masculine or feminine. Unique to their design, participants made gender categorization decisions by moving a computer mouse to select one of two categories presented on opposite ends of the screen, and their mouse movements were tracked. Results for both hairstyle and morphed faces revealed that while the ultimate categorization was generally correct, participants mouse movements were, on average, attracted to the incorrect gender category label when categorizing incongruent hairstyle cues (i.e., men with long hair, women with short hair) and when faces were gender atypical (i.e., feminine men, masculine women). Similar findings were demonstrated for race categorizations where stereotypic gender associations were mismatched (i.e., Asian men, Black women, Johnson, Freeman, & Pauker, 2012) and racial phenotypicality (Freeman, Pauker, Apfelbaum, & Ambady, 2010). Thus, while isolated visual cues to a particular social category (e.g., gender typicality) were not necessarily enough to prompt incorrect social categorization (e.g., a feminine man is still categorized as a man), recording how the categorization process unfolded over time revealed patterns of incorrect category activation due to mismatched visual cues.

Person Construal and Its Implications for People Construal

Since social psychology's inception researchers have focused on groups as central to the human psyche, particularly as the starting point for stereotyping and more nefarious processes like prejudice and discrimination (Allport, 1954). Indeed, when we encounter another individual, we automatically and efficiently activate social categories to which they belong which affects our downstream explicit and implicit judgments about them (Bargh, Chen, & Burrows, 1996; Bodenhausen, 1988; Brewer, 1988; Fiske & Neuberg, 1990). While this early work largely used semantic labels to denote groups, the person eventually made a return to person perception (Zebrowitz, 2006), however this posed a series of new challenges and questions (Macrae & Bodenhausen, 2000). By investigating social categorization as a process itself, rather than just the beginning of stereotyping and prejudice, it became clear that social categorization was not a discrete all-or-nothing judgment (Johnson & Freeman, 2010) rather visual cues from the face and body served as bottom-up input which in turn activated a multitude of social categories. Further, top-down cognitive structures such as attitudes, stereotypes, and prior experience could moderate categorization (Freeman & Ambady, 2011). In this sense, person construal resulted from numerous processes that could, in turn, account for research findings regarding within-category visual cue variability and feature-level determinants of social categorization. This paradigm shift, from thinking about social categorization as the starting point to viewing it as a product of dynamic and interactive bottom-up and top-down influences, created a more holistic and generalizable model of person construal.

Still, the dominance of person perception processes meant that single individuals were often the focus of attention. My dissertation purposes to shift this focus, examining people and building a theory of how social categories and visual cues impact people perception (Phillips et al., 2014). Person construal research provides the foundation for understanding which processes likely impact group perception, highlighting questions such, as can we visually perceive social categories within groups (i.e., a group's gender ratio)? Does this perception lead to judgments related to gender stereotype associations (e.g., more men in a group should increase the overall threat rating of the group)? And lastly, what role does within-group variability (e.g., gender typicality) play in these judgments and social category activation? Before outlining how I will test these questions, it is first important to review research from a domain outside of social psychology—cognitive vision science findings on a phenomenon known as ensemble perception (Alvarez, 2011; Whitney & Leib, 2018).

Chapter 3: The Visual Perception of Groups

Vision scientists have long been interested in how perceive and represent ensembles, or sets of objects. While we can intuit multiple instances when we see and make inferences about groups, the act of visually perceiving a group offers a provocative challenge to assumptions underlying limitations to vision and cognition. For instance, researchers have noted that we are unable to attend to and represent each individual item we perceive, with findings suggesting that adults can hold in mind information for about three to four objects (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). This four-item limitation exists for a variety of phenomenon including change blindness (Rensink, O'Regan, & Clark, 2000; Simons & Chabris, 1999), object-tracking (Scholl & Pylyshyn, 1999), and feature representation (Luck & Vogel, 1997). Yet, such a limitation is inconsistent with how we perceive and interact with our world and would represent a severe disadvantage to our perceptual and cognitive capabilities. Within the past ten years, there is a growing recognition in vision science that we overcome these limitations by processing and representing visual information in terms of summary statistical properties, leading to research on the phenomenon of ensemble perception.

Ensemble Coding: Low-Level Effects

To answer questions of how we overcome the constraints of our sensory and cognitive systems, early observations by Gestalt theorists acknowledged that our visual experience extends beyond simply the individual features we perceive (Wertheimer, 1938). These insights provided a foundation for understanding how the interaction between sensory input and cognition gives breadth to our visual experience. Much later, psychophysical research demonstrated that we could average across low-level visual features, such as speed (Watamaniuk & Duchon, 1992) and motion (Williams & Sekuler, 1984), which reduced the need to perceive, and represent, each object individually, as the average became the single percept.

Subsequent work on this averaging phenomenon spawned the study of ensemble perception (also know was ensemble coding, or summary representation) defined as: "the idea that the visual system naturally represents sets of similar items using summary statistics…serv[ing] as a computationally inexpensive means of obtaining valuable information about a scene" (Haberman & Whitney, 2012, p. 2) or simply put, the ability to extract visual summary statistical properties (e.g., mean, variance) from sets or groups of objects (Alvarez, 2011). An early demonstration of this effect emerged from the surprising finding that when shown an array depicting a multitude of different sized circles, participants were accurate in reporting the mean size of all the circles; yet when asked whether a particular circle was in the array participants were inaccurate (Ariely, 2001). These findings were evidence against previous propositions that when presented with multiple stimuli, the brain represents each at a low-resolution (Nakayama, 1990). Follow-up studies further highlighted the automaticity and efficiency of extracting ensemble perceptions, as accuracy in size detection remained high even with limited exposure duration (50 ms), memory delays (2000 ms), and different in circle size distributions: uniform, bimodal, and normal compared to homogenous (Chong & Treisman, 2003). Work also demonstrated that ensemble perception was not limited to mean or average percepts as variance could also be extracted and reported (Solomon, Morgan, & Chubb, 2011). Vision scientists demonstrated similar effects across a host of stimuli including size (Chong & Treisman, 2003); line orientation (Dakin & Watt, 1997); spatial location (Alvarez & Oliva, 2008); and number (Halberda, Sires, & Feigenson, 2006). Overall, this research pointed to our visual system's profound flexibility in extracting the gist or summary of a group of stimuli, offering a parsimonious explanation to long established cognitive and visual limitations.

Social Ensemble Coding

Critical to social psychology, vision scientists moved beyond low-level visual stimuli to investigate whether high-level visual characteristics vital to social processes (e.g., emotions, social categories, and identities) could also be extracted and accurately depicted via ensemble perception. In an early demonstration using more socially-relevant stimuli, Haberman and Whitney (2007) showed participants could distinguish the mean facial emotion (happy-sad and neutral-disgust) for groups of four and sixteen faces; even though participants could not discriminate individual faces in the group. They further demonstrated that this same effect emerged for gender, with faces morphed across a single male and female identity. Similar to low-level features, equivalent findings were shown for presentation times as low as 50 ms (Jason Haberman & Whitney, 2009) as well as percepts of variance and not just mean (Haberman, Lee, & Whitney, 2015)

As with low-level ensemble perception, a plethora of research revealed that accurate summary representations could be formed for a range of socially-relevant stimuli including point-light walkers heading direction (Sweeny, Haroz, & Whitney, 2013), eye gaze (Sweeny & Whitney, 2014), race (Thornton, Srismith, Oxner, & Hayward, 2014), humanness (Leib, Kosovicheva, & Whitney, 2016), head rotation (Florey, Clifford, Dakin, & Mareschal, 2016) and identity (J. de Fockert & Wolfenstein, 2009). Furthermore, researchers found that participants could accurately report whether a four-person group had more or less variability in terms of gender, race, and facial dominance compared to another four-person group (Phillips, Slepian, & Hughes, 2018).

Beyond additional demonstrations of ensemble perception's ubiquity, researchers sought to define its underlying mechanism and processes. For instance, while ensemble perception relies on forming a mean of visually perceived stimuli, perceivers do not appear to subsample specific parts of the array to make their judgments. The evidence supporting this assertion rests on the fact that perceivers discounted extreme emotional deviants from their summary representation (Haberman & Whitney, 2010); an effect also demonstrated for size perception (Brady & Alvarez, 2011). If sub-sampling was the mechanism by which ensemble perception operated, the single emotional deviant should sway some participants' judgments, however, accuracy remained high across all trials no matter where the deviant was placed (Haberman & Whitney, 2010).

It was also posited that ensemble perception operates via separate mechanisms for lower and higher level stimuli, as individuals' performance on ensemble tasks highly correlated within lower and higher level domains (e.g., high correlations between color and line orientation; emotion and identity) but not across lower and higher domains (e.g., emotion and line orientation) (Haberman, Brady, & Alvarez, 2015). Additionally, ensemble perception for faces is sensitive to face specific processing disruption, with performance severely hampered for inverted faces (Haberman & Whitney, 2009; Torralba & Oliva, 2003). Lastly, it appears that ensemble perception emerges early in human development, with infants as young as eight months able to store, and subsequently detect changes in up to three different ensembles (Zosh, Halberda, & Feigenson, 2011).

In gaining a better understanding of the mechanisms associated with ensemble perception, three critical areas have garnered intense focus and research: consciousness/awareness, attention, and memory. As with research on person perception, the interplay between cognitive factors and vision helps to establish the precise mechanisms associated with ensemble perception and opens broader theoretical questions (Whitney & Leib, 2018). Since this review focuses more on socially relevant stimuli I will devote primary attention to studies that used such stimuli, however, note that much of this work has also been demonstrated with low-level features.

A tenant of ensemble perception is that this process arises without conscious awareness (Whitney & Leib, 2018). Haberman and Whitney (2007) first demonstrated this effect when they found that perceivers, while accurate in reporting the average emotion of a group (whether the test face was happier or sadder than the mean of the group), were at chance in recognizing which out of two faces was actually in the set. A follow-up investigation found that disgust faces, presented outside of awareness, still influenced the perceived mean emotion of disgust (Haberman & Whitney, 2007). In addition, utilizing a crowding technique, in which a center face was flanked by six others to make it perceptually unrecognizable, researchers showed that the center face impacted the ensemble representation, even though the face itself was undetectable (Fischer & Whitney, 2011). In a further demonstration of awareness and ensemble perceptions, Haberman and Whitney (2011) presented perceivers with two successive 4x4 arrays composed of faces that varied in terms of happy and sad emotions. In the second array, four faces switched emotional extremes. Perceivers, when asked which array was happier (an ensemble percept) were above chance (~70%), however, when asked to locate any of the four faces which changed, perceivers were at chance. This demonstration of ensemble perception even though change blindness occurred, furthered the argument that conscious awareness of individual items was unnecessary for accurate ensemble perception.

A second significant facet of ensemble perception is its seeming independence from visual attention. Initial evidence showed robust ensemble percepts even when attention was divided. For instance, Chong and Treisman (2005) found that perceivers who viewed arrays consisting of circles of different sizes and colors were able to accurately report both cued characteristics (e.g., what is the average circle size of the set) and surprise non-cued characteristics (e.g., what was the average color of the circles). This effect also appeared across sensory modalities, with both visual and auditory ensemble perception emerging at the same time (Albrecht, Scholl, & Chun, 2012). Recent work, however, suggests that divided attention reduces the precision and efficacy of ensemble perception, with worse performance in an attentional blink paradigm (Brand, Oriet, & Sykes Tottenham, 2012; McNair, Goodbourn, Shone, & Harris, 2017). Overall, it has been argued that this question of whether ensemble perception requires or does not require visual attention is too limiting, with calls to investigate how visual attention interacts with ensemble perception to influence accuracy (Whitney & Leib, 2018).

In this vein, intriguing findings emerged from studies where visual attention was cued to particular items within an ensemble. In one of the first studies investigating cuing, Chong and Treisman (2003) showed that mean size perception was more accurate, (i.e., had a smaller threshold for 80% accuracy), when visual attention was global rather than cued, suggesting that approaching ensemble perception with a holistic attentional framework is advantageous. Other researchers showed that signaling attention to a particular circle within an array could bias overall average size estimates to that circle's size (de Fockert & Marchant, 2008), underscoring the effects of individuating attention to particular items within sets.

Attention can also be manipulated via less invasive cueing, with foveal, or center, versus peripheral placement of stimuli within an array biasing ensemble perception. For example, center presented faces in an emotion detection task were weighted to a higher degree when assessing the mean emotion (Ji, Chen, & Fu, 2014). Recent work also suggests exciting possibilities regarding directed visual attention and summary representations of race. Jung, Bülthoff, and Armann (2017) showed arrays of faces, composed of faces morphed to vary between two Asian and White faces, and tested whether faces at the center of a 3x4 array influenced summary representations of average race more than faces at the visual periphery. Results indicated accurate ensemble perception for the whole array, however, faces presented at the center were given more weight in the overall summary representation as demonstrated by performance data matching ideal observer models derived from taking into account only the two center faces. In a methodologically rigorous test of visual attention and ensemble perception, Wolfe, Kosovicheva, Leib, Wood and Whitney (2015) showed that when center viewed faces (measured via eye-tracking) were occluded in real time (i.e., whenever participants eye-gaze focused on them), ensemble perception remained accurate. Overall, these findings suggest intriguing possibilities that visual attention directed at the center of arrays, while not wholly accounting for summary representations, affects ensemble summations more than periphery percepts.

One final cognitive process related to ensemble perception is memory. Researchers have noted that ensemble perception can also occur for stimuli presented in rapid succession rather than all at once (however primacy and recency effects influence the summary perception, Hubert-Wallander & Boynton, 2015). This sequential presentation effect underscores ensemble perception's potential role in visual and working memory (Haberman, Harp, & Whitney, 2009; Brady & Alvarez, 2011). An interesting extension of this is Ying and Xu (2017), who found that visual adaptation aftereffects emerged for emotional faces presented via rapid serial visual presentation. Participants viewed ten different faces all expressing a single emotion (either happy or sad) sixteen times (for a total of 160 face presentations) in the span of approximately 3.7 seconds—developing a summary perception of the faces. After, when asked whether a single face was happy or sad participants showed point of subjective equality shifts dependent on which visual adaptation condition they were in (e.g., if adapted to happy faces, they perceived happy expressions less). This study provides further evidence that we adapt to the multitude of faces we see in any particular environment, with important implications for people perception.

In summary, the explosive growth of research in ensemble perception underscores the importance of our visual system's profound flexibility in extracting the gist or summary of a group of stimuli. Of crucial significance, ensemble perception offers a parsimonious and poignant counter to long-established cognitive and visual limitations, greatly expanding the four-item limit by allowing for items to be defined not as single objects but rather as groups. Furthermore, the early development, automaticity, and limited need for attentional resources, suggests that ensemble perception is a powerful mental capability related to perceiving our social world and the groups of people within it. Questions, however, remain. Specifically, how ensemble perception informs downstream social judgments about groups and what more socially relevant factors influence our group percepts. Overall, the integration of social psychology theory with ensemble perception opens the potential for numerous important and critical insights into people construal.

Chapter 4: People Construal: Integrating Person and Ensemble Perception

In Chapter 2, I reviewed the extensive literature on person perception, and its focus on understanding how social categories underlie the inferences and social evaluations we make about others. This area of research examined the "person" via semantic labels and single faces or bodies, leaving the large visual landscape of "people" untouched. For my dissertation I propose to explore this landscape, focusing on our visual perception of groups and the inferences we draw about these group percepts. Vision science research on ensemble perception provides a strong foundation for developing this perspective. Researchers demonstrated that for both low- and high-level stimuli, we as observers can form accurate summary representations (e.g., mean and variance). Building upon this foundation, I interrogate processes associated with our ability to visually perceive social categories within groups and how these social categories inform the inferences we make about the group as a whole; advancing our theoretical and empirical knowledge of people construal.

Across six studies, I address three distinct but interrelated questions. In Study Set 1 I examined whether perceivers were sensitive, after a brief visual presentation (500 ms), to group social category variability in terms of gender (men or women). In three studies, I manipulated the gender ratio of twelve-person groups and asked participants for numeric estimations of the number of men within the group as well as overall trait judgments about the group. These trait judgments focused on a critical evaluative domain—threat. Indeed, understanding whether a group represents a hostile or hospitable entity is a vital evaluative inference directly impacting our survival and safety (Boyer, Firat, & Leeuwen, 2015). Groups have long been the actors and instigators of some of the worst and deadliest conflicts, and pose higher threat, such as an elevated capacity to inflict damage, compared to a single individual (Schaller & Neuberg, 2008). While past research reveals numerous cues by which we come to evaluate others regarding their imposed threat (e.g., physical formidability, past success, emotional signals, and social categories, Archer, 2004; Holbrook & Fessler, 2013; Neuberg, Kenrick, & Schaller, 2011) little research has investigated our perceptions of threat from visually perceived groups.

Furthermore, theoretical and empirical insights predict that as a group's gender ratio shifts from more women to more men, the group will be rated as more threatening. Consistent evidence across multiple fields of study, indicates that men, compared to women, enact more violence against others. For instance, archeologists found that mass graves of bodies that showed signs of violent deaths (e.g., injuries sustained by weapons or blunt blows), contained exclusively or mostly bodies of men (Boucherie, Jørkov, & Smith, 2017). Additionally, historical overviews and sociological findings often note that men, compared to women, comprise more causalities in warfare (Goldstein, 2003), are more likely to be the victims of homicide (Kellermann & Mercy, 1992) and make up the bulk of arrests for violent crimes (Pastore & Maguire, 2005). Collectively, these patterns strongly suggest that men are overall more associated with violent behavior than women (Archer, 2004). These gender differences in aggression are further reinforced by cultural norms and beliefs, effectively creating a feedback loop whereby men are expected to be more aggressive and violent compared to women, a factor contributing to this difference (Frieze & Li, 2010; Wood & Eagly, 2002). In summary, there is strong support for the assertion that a group's overall threat rating should be affected by the number of men within the group.

Overall, by integrating insights from vision science to examine socially relevant inferences about with groups, I propose three general hypotheses for Study Set 1:

> **Hypothesis 1:** Participants will be sensitive to the number of men within briefly presented (500 ms), visually perceived, groups.

Hypothesis 2: Social evaluations of groups will track group gender ratio, such that as the number of men in a group increases (and the number of women decreases), group threat/hostility judgments will increase. Furthermore, as the number of men in a group increases, ratings of warmth/invitingness will decrease.

Hypothesis 3: This pattern of threat evaluations should hold for more implicit measures, such that majority men groups, when used as primes within the Affect Misattribution Procedure (AMP), will engender subsequently presented unrelated stimuli to be judged as more threatening.

Study Set 1 thus provides a foundation for subsequent research by demonstrating that perceivers are sensitive to a groups' gender ratio and that this sensitivity informs social evaluative judgments about groups as a whole.

In Study 4, I seek to understand whether visual cues to within-category variability (i.e., gender typicality) modulate Study Set 1 results. In particular, I focus on how visual cues to men's facial masculinity and femininity influence group evaluations in terms of social category perception and social evaluations. Related to social vision researchers' findings on the effects of within-category visual cue variability (Freeman & Ambady, 2011; Johnson et al., 2015; Maddox & Perry, 2017), I believe similar aggrandizement of threat inferences should emerge for groups composed of masculine compared to feminine men. These results would demonstrate that perceivers are sensitive not only a group's gender ratio but also visual characteristics (i.e., facial masculinity and femininity) within groups. Overall, for Study 4, I hypothesize that:

Hypothesis 4: The effects found in Study Set 1 will be stronger for groups that contain faces of masculine compared to feminine men.

That is, groups with masculine men, compared to feminine men, should produce higher numeric estimates of the number of men, higher threat ratings, and higher overall estimates of group gender typicality (a new measure added to Study 4).

To examine these questions, I developed images using face morphing software which allows for precise control over the degree of masculinity and femininity of the men's faces. Evidence in support of these hypotheses would further extend our knowledge of people construal, situating this phenomenon in terms of how visual inputs affect group perception akin to person construal processes (Freeman & Ambady, 2011).

Finally, in Study Set 2, I assess how gender typicality moderates social category decisions about groups. A critical question regarding people construal is whether and how social categories become activated following a group percept. While evidence suggests that group percepts do evoke social category activation, and to an even stronger degree than single faces (see Cooley & Payne, 2017), no prior research has directly assessed the interactive effects of social categories (i.e., gender) and within-category variability (i.e., gender typicality) on social category activation from groups. I investigate this question using two social cognitive research paradigms, the sequential priming task and mouse-tracking.

For the sequential priming task, I borrow methods from Macrae and Martin (2007) who found that by using a short prime duration, 20 ms versus 200 ms, visual features (i.e., hairstyle cues) activated incorrect gender category information, overriding the social category itself (i.e., women with short hair facilitated faster reaction times to men's names compared to women's names). To disentangle the effects of visual cues (i.e., gender typicality) and social categories (i.e., majority men or women groups) I use a similar sequential priming task, with participants viewing group arrays composed of majority men (masculine or feminine) and majority women (no gender typicality cue manipulated) as primes, followed by a target face for which they make a gender categorization (man or woman). I hypothesize:

> **<u>Hypothesis 5:</u>** When prime duration is short (200 ms), compared to long (1000 ms), participants should utilize visual cue information. Specifically, majority masculine men arrays should facilitate the category "man" (i.e., categorize target men's faces faster as measured via reaction times) to a greater degree than majority feminine men arrays.

> This relationship should not emerge for long prime durations (1000 ms), as such time is sufficient to override the use of visual cues and instead rely on social category information.
>
> Furthermore, regardless of men's gender typicality and prime duration, majority women groups should facilitate the category "woman" (i.e., participants should be faster to categorize a woman's face, compared to a man's face, following majority women group primes).

Findings that align with these hypotheses would provide further evidence that people construal processes utilize visual cue information like gender typicality. Specifically, category activation aligning with visual cues (i.e., masculine men) and dominant social categories (i.e., majority men) would suggest that gender typicality and gender ratio facilitate social category activation.

While the sequential priming task is sufficient to assess social category activation, it reveals little about the underlying process regarding *social categorization*. In a final study I assess how social categorization decisions unfold over time by using mouse-tracking software that provides online tracking of potential "incorrect" category attraction. By having participants perform a mouse-tracking categorization task (i.e., does the group contain "mostly men" or "mostly women") for group stimuli that are either majority men or majority women and vary in men's gender typicality, I assess the degree to which gender typicality influences social category decisions. I hypothesize that:

> **Hypothesis 6:** When prime duration is short (200 ms) and participants are categorizing majority masculine men groups, participants will show faster reaction times to categorize the group, more direct categorizations, and will make fewer errors, compared to groups composed of majority feminine men. This relationship, however, should be attenuated at the longer prime duration (1000 ms).

The use of mouse-tracking affords a better understanding of participants' underlying processes associated with social category activation, revealing patterns of partial activation beyond simply the final social categorization judgment.

Overall, I aim to provide a foundation for understanding visual people construal—how we perceive groups with varying numbers of men and women, and the downstream implications of this perception in terms of evaluative judgments and social category activation. Critically, by examining visual cues to gender (i.e., gender typicality), I explore not only social categories per se, but also the influence of the visual cues that underlie gender. The integration of ensemble perception with person construal offers numerous insights into the social vision of groups. On a theoretical level, understanding how group perception affects group evaluations, brings our laboratory studies and conceptual models closer to our real-world experience of perceiving people. In turn, these insights have profound implications regarding evaluative judgments such as threat. The novel area of people construal represents a rich and vast unexplored domain, pushing our understanding of the visual perception beyond a single person to that of groups.

Study Set 1:

Threat in the Company of Men—Ensemble Perception and Threat Evaluations of Groups Varying in Gender Ratio

The studies comprising Study Set 1 examined Hypotheses 1-3. In three studies, I (along with my coauthors) examined how a group's gender ratio impacted perceivers' ratings of threat. Because of associations between men and threat, aggression and violence (Archer, 2004; Eagly & Steffen, 1986), I hypothesized that groups would be judged as more threatening as the ratio of men to women increased. I tested this hypothesis using both explicit (Study 1), and more indirect (Study 2) measures. In Study 3, I examined two potential mediators for why groups with more men were judged as more threatening. On the one hand, based solely on numeric probability, more men within a group may directly account for judgments of threat. That is, the social category influence of perceiving more men within a group directly impacts threat evaluations.

On the other hand, more men within a group might increase the perceived coalition of the men in that group, which may explain higher threat evaluations. Indeed, according to the male warrior hypothesis, men are more attuned to intergroup conflicts, and are much more likely aggress when faced with intergroup conflicts (McDonald, Navarrete, & Vugt, 2012; Vugt, 2009) than women. Thus, to the degree that a group is more homogenous regarding the more threatening target, here men, the group is likely to be viewed as higher in entitativity, that is the cohesiveness of the group or how "groupy" the group is viewed (Campbell, 1958; Hamilton, Sherman, & Lickel, 1998). This possibility implies that perceived men's coalition might mediate judgments of group threat. While not listed as a hypothesis, testing these mediators provides insight into what gives rise to social evaluative judgments related to groups and threat. Overall, this work is among the first empirical studies to bridge social vision and ensemble perception research (also see Phillips et al., 2014), providing novel insights into the accuracy and social evaluative consequences of visual group perception.

Study 1

Method

Participants. Power analyses, utilizing parameters derived from our models, were run with MPlus simulations following guidelines by Muthén and Muthén (2002). Given the high power afforded by our within-subject designs, these simulations determined that at least 20 participants were needed to detect an effect of the size estimated in our study at 90% power. Based on our lab's prior rate of data collection, I collected data for two weeks to reach the desired number of participants. In fact, I exceeded the target number, recruiting eighty-two undergraduate participants (50 women, $M_{age} = 19.19$, $SD_{age} = 1.55$).

Stimuli. Group or ensemble stimuli consisted of 3x4 arrays depicting faces of White men and women from the Chicago Face Database (CFD) Version 1.0 (36 men and 38 women; Ma, Correll, & Wittenbrink, 2015). I used a custom Python script to generate arrays with faces randomly selected without replacement within trial, but with replacement between trials, from available stimuli. The location of each face was randomized within each ensemble. Ensembles varied in the ratio of men to women from 0:12 to 12:0 in increments of one. For each of the 13 possible gender ratios, I generated a set of 300 unique ensembles, resulting in a total of 3,900 ensembles from which to sample (for sample stimuli see Figure 1, please note faces depicted in figure are not the real faces used in stimuli, stimuli faces were drawn from the CFD).

Figure 1. Sample group stimuli, depicted faces generated from thispersondoesnotexist.com, Chicago Face Database faces used in real stimuli

Procedure. Stimuli were presented and responses were recorded using Inquisit software. The protocol consisted of two blocks presented in counterbalanced order. Each trial began with a fixation cross (500 ms), followed by a randomly selected ensemble (500 ms) about which participants provided judgments. In one block, participants estimated the number of men in each ensemble (*0 men – 12 men*). In the other block, participants rated each ensemble along four evaluative dimensions: invitingness (1 = *not at all inviting,* 7 = *extremely inviting*), warmth (1 = *extremely cold,* 7 = *extremely warm*), hostility (1 = *not at all hostile, 7 = extremely hostile*), and threat (1 = *not at all threatening,* 7 = *extremely threatening*). Each block included 260 arrays randomly selected from the total population of 3,900 ensembles, with 20 ensembles for each level of gender ratio. After completing their judgments, participants provided demographic information and were debriefed.

Results and Discussion

I analyzed data both including and excluding participants who had incomplete data (n = 13), and results were unchanged. I therefore included data from all participants.

I predicted that numeric estimates would reflect the ratio of men to women in each ensemble, and that overall judgments of threat would increase as the ratio of men to women increased. I used the R packages “lme4” and “lmerTest” to compute hierarchical linear models which account for within-subject variation and ratings nested within participant (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2015). All models included random intercept, random slopes with residual maximum likelihood estimation (REML). All coefficients were unstandardized. Also, I tested potential interactions with participant gender (effect coded as *-1 = men, 1 = women*); in all studies participant gender did not qualify any effect unless otherwise noted.

First, I examined whether participants were sensitive to the gender ratio of ensembles by regressing participant's numeric estimates of men onto Gender Ratio (mean centered). Consistent with hypotheses, as the ratio of men to women increased, so did participants' numeric estimates of men in an ensemble, $B = 0.52$, $SE = 0.02$, $t(77.63) = 28.10$, $p < .001$, 95% CIs [0.48, 0.55].[2] Next, I tested whether the actual number of men in each ensemble affected threat judgments of groups. To achieve this, I calculated a mean threat score by averaging the four evaluative items (with warmth and invitingness reverse-coded), such that higher scores represented higher threat. The composite of these four items showed high within-subject reliability (0.94) based on the method outlined by Cranford et al. (2006). Regressing the overall threat judgment score onto actual Gender Ratio, I found that as the ratio of men to women increased, threat judgments also increased, $B = 0.06$, $SE = 0.01$, $t\,(78.65) = 6.55$, $p < .001$, 95% CIs = [0.04, 0.07]. In summary, consistent with our hypotheses, participant's numeric estimates of men in an ensemble corresponded to the actual number of men depicted in each ensemble. Importantly, perceivers also judged the group to be more threatening as the ratio of men to women increased.

Study 2

In Study 2, I extended these findings beyond self-report measures by using a more indirect measure: the Affect Misattribution Procedure (AMP; Payne, Burkley, & Stokes, 2008). The AMP has been used as an indirect measure of both emotion (e.g., fear) and semantic (e.g.,

---

[2] There was a significant Gender Ratio by Participant Gender interaction, $B = 0.06$, $SE = 0.02$, $t(66) = 2.85$, $p = .006$, 95% CIs [0.02, 0.10]. We decomposed the interaction by examining simple effects of Gender Ratio broken down by Participant Gender. Results showed that women's numeric estimates of men were more strongly tethered to gender ratio, $B = 0.54$, $SE = 0.02$, $t(49) = 23.62$, $p < .001$, 95% CIs [0.50, 0.59], than were men's numeric estimates, $B = 0.42$, $SE = 0.03$, $t(17) = 13.61$, $p < .001$, 95% CIs [0.36, 0.48]. However, both simple effects were significant and in the same direction, indicating that participants' numeric estimates of men increased as the ratio of men to women increased.

personality traits; animacy) concepts and a previous meta-analysis indicated it is a valid and reliable means of measuring attitudes on par with other similarly implicit tasks (Deutsch & Gawronski, 2009; Gawronski & Ye, 2014). Aligning with results from Study 1, I hypothesized that the arrays with higher ratios of men to women would prime the concept of threat, leading to a misattribution of threat for subsequently presented targets.

Method

Participants. As in Study 1, power analyses using MPlus simulations were conducted with parameters derived from our models. Based on these simulations, I determined that a sample size of at least 20 was needed to detect an effect of the size estimated in our study at 90% power. Again, based on our lab's prior rate of data collection, I collected data for two weeks to reach the desired number of participants. In this time, I exceed the desired number, recruiting 102 undergraduate participants (74 women, M_{age} = 18.92, SD_{age} = 1.23).

Stimuli. The AMP utilizes two types of stimuli: primes and targets. Prime stimuli consisted of ensembles from Study 1 depicting five different gender ratios of men to women: 0:12, 3:9, 6:6, 9:3, and 12:0 (300 unique arrays per gender ratio). Target stimuli for the AMP traditionally consist of a set of Chinese characters (Payne et al., 2008). Many participants in our subject pool are familiar with Chinese, so I generated a new set of 360 target stimuli that depicted unique combinations of Thai letters.

Procedure. Utilizing the AMP script developed for Inquisit software (Borchert, 2016), participants viewed pairs of images flashed sequentially on a computer screen. Participants' task was to judge whether the second image—the target stimulus—appeared to be more threatening or less threatening than the average stimulus by pressing "E" or "I" on the keyboard (counterbalanced across participants). Each trial consisted of the following: a prime stimulus (i.e., an ensemble) presented for 75 ms, a blank screen for 125 ms, a target stimulus (i.e., Thai letters) for 100 ms, and a visual mask that remained on the screen until the participant made a judgment. Participants completed ten practice trials followed by 360 critical trials (randomized), consisting of 60 trials for each sex ratio and 60 additional trials that contained no prime. Upon completion, participants provided demographic information and indicated whether or not they knew Thai.

Results and Discussion

Eight participants were excluded for providing invariant responses (e.g., rated all stimuli as more threatening) and one participant was excluded for indicating they knew Thai, leaving a total sample of 93 participants.

I analyzed data using the R packages "lme4" and "lmerTest", ensuring models fit a binomial distribution. I coded AMP responses dichotomously (*0 = less threatening, 1 = more threatening*) and the gender ratio of prime stimuli continuously (from *0 (0 men:12 women)* to *4* (*12 men:0 women*), mean centered). I then regressed AMP responses onto Gender Ratio. As predicted, for each unit increase in Gender Ratio, the odds ratio for categorizing the target as more threatening, compared to less threatening, increased by 1.07, $B = 0.07$, $SE = .02$, $z = 2.89$, $p = .004$, 95% CIs = [0.02, 0.12], OR = 1.07. This effect was qualified by a higher-order quadratic trend (see Figure 2),[3] whereby the likelihood of a target stimulus being judged as more threatening was relatively equal when women outnumbered men (i.e., 0 men:12 women, 3 men:9

[3] For this quadratic trend, I found a significant interaction with Participant Gender, $B = 0.02$, $SE = 0.01$, $z = 2.46$, $p = .014$, 95% CIs = [0.01, 0.04]. I decomposed the interaction by exploring the simple quadratic effect separately for men and women. For women, the quadratic trend was significant, $B = 0.03$, $SE = 0.009$, $z = 3.25$, $p = .001$, 95% CIs [0.01, 0.05]; for men, the quadratic trend was not significant, $B = -0.01$, $SE = 0.02$, $z = -0.81$, $p = .418$, 95% CIs [-0.05, 0.02]. I should note, however, that there were few men ($n = 18$ after exclusions) thus strong interpretations should be cautiously drawn.

women) and when men and women were equal (i.e., 6 men:6 women), but increased exponentially as men became the majority in the arrays, $B = 0.02$, $SE = 0.01$, $z = 2.53$, $p = .011$, 95% CIs [0.01, 0.04], OR = 1.02. Thus, when exposed to majority men groups, participants judged a subsequent target to be more threatening. These results corroborate the explicit evaluations from Study 1.

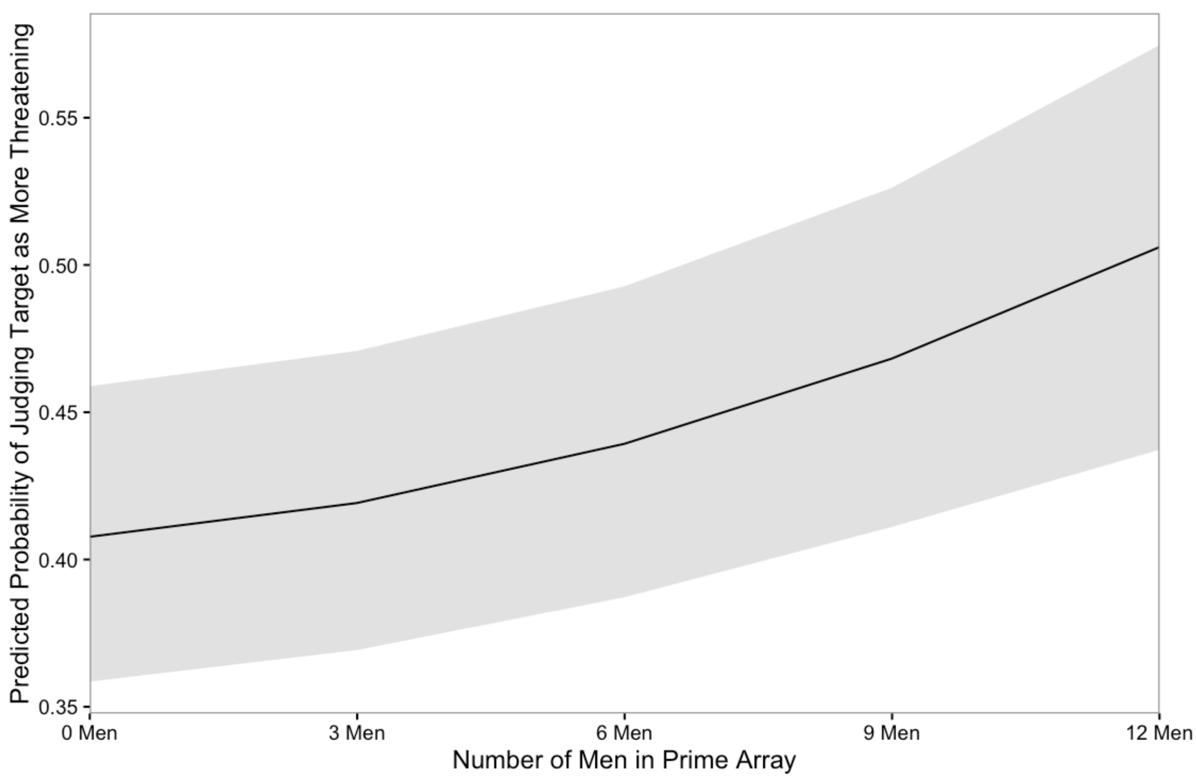

Figure 2. Predicted probabilities from Study 2 of categorizing the target as "more threatening" as a function of number of men in the prime array.

Study 3

In Study 3 I examined two potential underlying factors accounting for why a group of men may be viewed as more threatening. As argued above, one straightforward explanation may be that because men are associated more with violence and aggression (Archer, 2004; Eagly & Steffen, 1986), perceived threat might increase as the ratio of men to women increases. Alternately, because men, compare to women, more often enact coalitional aggression and violence (McDonald et al., 2012; Vugt, 2009), the degree to which the men in an array are perceived to be a collective entity may account for higher threat evaluations. I examined these ideas in Study 3, as well as replicated results from Study 1, by testing whether the perceived number of men or the perceived coalition of the men in the group mediated the relationship between the gender ratio and threat evaluations.

Method

Participants. I replicated the recruitment strategy for Study 1, collecting data for a total of two weeks. In this time, I recruited 99 undergraduate participants (68 women, M_{age} = 19.42, SD_{age} = 1.28).

Stimuli. I utilized the stimuli from Study 1, however I excluded ensembles with men to women ratios of: 0:12, 1:11, 2:10 and 10:2, 11:1, 12:0 because groups of zero, one, or two individuals would be unlikely to constitute a coalition, thereby rendering our measure meaningless for these groups. Thus, our pool of potential ensembles was 2,100, with 300 unique ensembles per gender ratio.

Procedure. Overall the protocol replicated Study 1, with two notable differences. First, the blocks were presented in sequential order, with numeric estimations first, followed by evaluative judgments, and a new block assessing perceived coalition of the group. For this coalition judgment, participants were instructed to respond to questions I adapted from the entitativity literature (Ip, Chiu, & Wan, 2006): “To what extent do the men seem more like a group rather than just a bunch of individuals?” (1 = *Much more like a collection of individuals* *than a group,* 7 = *Much more like a group than a collection of individuals*).[4] To allow for tests of mediation, participants judged the same set of randomly selected ensembles from the total pool of potential ensembles from the first block in each subsequent block, presented in a random order within each block. After completing the task, participants provided demographic information and were debriefed.

Results and Discussion

As in Study 1, I hypothesized that numeric estimates of the number of men and threat judgments for ensembles would increase as the gender ratio of men to women increased. I also hypothesized that as the gender ratio of men to women increased, judgments of men's coalition would increase. Finally, I tested whether perceived number of men *or* perceived men's coalition mediated threat judgments.

I used the same R packages and analytic strategy as Study 1, fitting hierarchical linear models to the data. To test our hypotheses regarding mediation I used the "mediation" package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2013) and examined two different mediation models, one with the perceived number of men as the mediator, and the other with judgments of men's coalition as the mediator. For each model run, I set the simulation parameter to 1,000. No participants had incomplete data.

Results for numeric estimations and threat judgments followed the same pattern found in Study 1. In terms of numeric estimations, as the ratio of men to women increased, so did participants' numeric estimates of men in the ensemble, $B = 0.33$, $SE = 0.02$, $t(89.12) = 20.21$, p

[4] Immediately following the men's coalition question, I also asked the same question but replaced the word men with women. Overall results with women were in the predicted direction (e.g., as the Gender Ratio of men to women in the ensemble increased, perceived women's coalition decreased). Multilevel mediation analyses did reveal that women's coalition mediated threat judgments, unlike the finding for men.

$< .001$, 95% CIs [0.30, 0.37]. Unlike Study 1, the Participant Gender by Gender Ratio interaction did not reach significance, $B = 0.004$, $SE = 0.04$, $t(88.55) = 0.11$, $p = .911$. Threat judgments (which had high within-subject reliability, 0.97) also showed the same pattern as Study 1, as the ratio of men to women increased, so did threat judgments, $B = 0.04$, $SE = 0.01$, $t(83.51) = 6.04$, $p < .001$, 95% CIs = [0.02, 0.05]. Consistent with hypotheses, I found that as the ratio of men to women increased, participants judged the men in the group to represent a coalition, rather than a collection of individuals, $B = 0.09$, $SE = 0.01$, $t(87.46) = 7.09$, $p < .001$, 95% CIs = [0.07, 0.12]. Lastly, I tested whether Perceived Number of Men and judgments of Men's Coalition predicted overall threat judgments. As numeric estimates of men in the ensemble increased so did threat judgments, $B = 0.03$, $SE = 0.004$, $t(66.55) = 5.66$, $p < .001$, 95% CIs = [0.02, 0.04]; however Men's Coalition did not predict higher threat judgments, $B = 0.01$, $SE = 0.01$, $t(79.36) = 1.24$, $p = .218$, 95% CIs = [-0.001, 0.003].

Finally, I examined two different mediation models to determine whether the relationship between Gender Ratio and overall threat judgments was mediated by the Perceived Number of Men or judgments of Men's Coalition. I conducted mediation tests using multilevel mediation analyses (see Zhang, et al., 2016 for review), examining the average causal mediation effect (ACME) and the average direct effect (ADE). First, I tested whether Perceived Number of Men was a mediator of the relationship between Gender Ratio and threat judgments. Results indicated that the ACME was significant, ACME = 0.01, $p < .001$, 95% CIs = [0.002, 0.01]. I also found evidence for a significant ADE = 0.03, $p < .001$, 95% CIs = [0.02, 0.04]. Overall, the significant ACME, indicates that Perceived Number of Men mediated the relationship between Gender Ratio (i.e., the actual number of men) and threat judgments.

Next, I tested Men's Coalition judgments mediated the relationship between Gender Ratio and threat judgments. In this model, the ACME did not reach significance, ACME = 0.0003, $p = .610$, 95% CIs = [-0.0001, 0.001]; there was however a significant ADE = 0.04, $p < .001$, 95% CIs = [0.02, 0.05]. Thus, I did not find evidence that Men's Coalition judgments mediated the relationship between Gender Ratio and threat judgments.

General Discussion

Results from Study Set 1 confirmed hypotheses, indicating that perceivers were sensitive to the gender ratio of twelve person groups, even after a very brief visual presentation (500 ms). Furthermore, trait ratings of groups, assessed both explicitly and more implicitly, tracked the groups' gender ratio—as the number of men increased (and the number of women decreased), the group was rated as more threatening and hostile, and less warm and inviting. Interestingly, participants' numeric estimates of the number of men and not perceived men's coalition, mediated group threat ratings, suggesting that *perceived* social categories serve as the critical cue for evaluative inferences about briefly presented groups.

Study Set 1 offers a first glimpse into how we perceive and evaluate groups of people. Results confirmed long standing tenants of person construal, specifically the capacity to make social category judgments and their importance for how we evaluate others (Macrae & Bodenhausen, 2000). It is particularly telling that facilitated priming effects emerged even after a group was presented for as little as 75 ms, with majority men groups associated with more threat. This finding is particularly interesting in the context of ensemble perception. Given that our perception of groups results in extracting the average or mean gender typicality of the group, a group of six men and six women may result in an average that is ambiguously gendered (neither a man nor a woman). Thus, it is only when the men become the majority, resulting in a distinctly masculine or "man" ensemble summation, that priming effects emerge. This may be especially the case when viewing arrays for such a short duration (75 ms), which mimics speeds in vision research (Haberman & Whitney, 2007).

This research is some of the first to explicitly focus on the visual perception of groups, merging ensemble perception from vision science to address social psychological questions such as group evaluations. Indeed, aligning with shifts to focus on our perception of actual persons (e.g., the face of a man or woman), as opposed to category labels (i.e., the word man or woman, Zebrowitz, 2006); this work situates our perception of actual groups rather than group descriptors (e.g., faces of men and women as opposed to a bar graph showing the percentage of men and women). Thus, this work extends knowledge about our sensitivity and capacity to perceive these more complex group stimuli and highlights the critical role that social category knowledge plays in our trait evaluations of groups.

Study 4: Impact of Men's Gender Typicality Cues on Group Perception and Group Evaluation

Results from Study Set 1 revealed insights into people construal drawn from category-based group perception (i.e., men and women). It remains unknown, however, whether perceivers are sensitive to phenotypic visual cues to gender (i.e., masculinity and femininity) in the context of groups. That is, how might categorical perception and the inferences drawn about groups change as a function of the visual cues which support such categorical judgments? By investigating the role of gender typicality via visual gender cues, I begin to situate people construal within person construal frameworks—specifically, the utilization and impact of bottom-up visual cues to gender typicality on both categorical judgments (i.e., the number of men within a group) as well as group trait evaluations.

I again turn to ensemble perception to understand whether phenotypic variation, such as gender typicality, affects people construal. Ensemble perception of facial identity provides indirect evidence that perceivers are sensitive to gender typicality when forming summary representations of groups. For instance, when presented with groups composed of faces morphed across three distinct identities, which varied in gender typicality, perceivers could accurately represent the mean identity of the group (Bai, Leib, Puri, Whitney, & Peng, 2015). While identity and gender typicality are undoubtedly unique, this evidence suggests that perceivers are sensitive to fine-grained visual cues (e.g., denoting facial identity) and not just broader social categories (e.g., gender). In addition, Phillips and colleagues (2018), demonstrated that perceivers were above chance in determining whether a four-person group had more or less gender typicality variability than a previously presented four-person group, and this effect was robust to backward masking and reduced presentation times (200 ms). One limitation to prior work is the few unique facial identities used (three and seven respectively). Additionally, the scope of the questions focused on perceptual capability (i.e., summary representation and the ability to distinguished variability) rather than potential downstream consequences associated with visual gender typicality cues. Still, these findings suggest that a target's gender typicality within groups should be perceptual apparent and in turn, should influence social evaluative impressions of the group as a whole.

Broadly, Study 4 investigated within-category, bottom-up, visual cues to gender typicality and their impact on people construal phenomenon—category-based numeric estimations and social evaluative group judgments. To investigate these questions, I generated a new set of face stimuli in which each man's face was morphed to composite men's or women's faces, resulting in distinctly masculine or feminine gender typicality. These morphed faces comprised the pool of available stimuli for group array generation. The use of morphed faces was advantageous as it allowed for the same man's face to be used as the base for a masculinized or feminized version, thus facial characteristics indicative of identity were preserved and a direct effect of gender typicality was assessed. As in Study Set 1, I focused on men's faces, as questions related to threat judgments were more relevant for differences related to masculine and feminine men's faces. I note, however, that future extensions will investigate the influence of gender typicality on women's faces.

Study 4 largely follows methods from Studies 1 and 3, with perceivers making category-based (i.e., the number of men) and trait judgments (i.e., threat/hostility; warmth/invitingness) about visually perceived groups. In addition, I included a new measure assessing the perceived average gender typicality of the group. This new measure provided an assessment of gender typicality that does not rely on category labels (i.e., the number of men). Overall, I predicted that perceivers' judgments would be influenced by the gender typicality of men within the arrays. That is, as the number of men in the group increases, masculine men groups should be judged to have a higher number of men in them, higher average masculinity, and higher stereotype-consistent threat trait judgments, compared to groups composed of feminine men.

Method

Participants

I recruited participants for two weeks through the UCLA Communication Department SONA pool. During this time 96 participants were recruited (22 men, 74 women, $M_{\text{age}} = 19.17$, $SD_{\text{age}} = 0.99$, 30.2% Asian/Asian American, 21.9% White/European American).

Stimuli

As in previous studies, the final stimuli were 3x4 arrays composed of men's and women's faces. I used a custom python script to generate ensembles, in real-time, with random placement and random selection of faces within a 3x4 grid. Faces were pulled from the pool of available faces (see below), with the limitation that no face appeared twice with the same array.

Different from Studies 1-3, the pool of available men's faces was composed of masculine morphed and feminine morphed men's faces. These faces were generated using FantaMorph which allows for the two-dimensional averaging of facial images based on 114 landmarks placed upon faces. To generate these facial morphs, I first created 10 composite morphs of men's faces by randomly selecting White men's faces from the Chicago Face Database (Ma et al., 2015) and morphing them together. A similar process was used to create 10 composites of women's faces. In total, I created 10 men's faces and 10 women's faces which served as the anchors for subsequent morphs. I then morphed 66 White men's faces from the Chicago Face Database (excluding faces which were categorized below 80% White based on pre-test data (see Ma et al., 2015), faces with hair below the shoulders, and faces which produced artifacts in the morphing) 35% towards one of the ten composite men's and women's faces (randomly selected). In total I generated a masculine and feminine version of all 66 faces (in total 132 faces). Since morphing can produce subtle yet impactful facial changes (e.g., smoother skin, see Sutherland, Rhodes, &

Young, 2017) I also morphed each woman's face, selected from the Chicago Face Database (n = 72), to a randomly selected women's base face. Thus, I created morphs for each woman's face which still preserved gender typicality. In sum, the pool of faces from which ensembles were created contained 66 White masculine men's faces, 66 White feminine men's faces, and 72 White women's faces (see Figure 3 for sample stimuli).

Figure 3. Sample stimuli, feminine men's faces on the left, masculine men's faces on the right

Procedure

The procedure closely mirrored Study 1. Participants made three different sets of judgments across three randomly presented blocks. In one block, participants estimated the number of men within the arrays (0 *men – 12 men*). In another block, participants rated each ensemble in terms of four evaluative dimensions: invitingness (1 = *not at all inviting,* 7 = *extremely inviting*), warmth (1 = *extremely cold,* 7 = *extremely warm*), hostility (1 = *not at all hostile, 7 = extremely hostile*), and threat (1 = *not at all threatening,* 7 = *extremely threatening*). These dimensions were presented in random order following each array. New to Study 4, participants also made judgments assessing the overall gender typicality of the array (i.e., "please select the face which best represents the mean or average facial masculinity/femininity from the group"), using a visual scale composed of FaceGen faces. This scale depicted faces morphed along a gender continuum and participants were asked to indicate via numbers below each face, which face best represented the average face from the array in terms of masculinity/femininity (see Figure 4).

Figure 4. Gender Typicality Scale

The stimuli presentation program randomly selected from the pool of available faces when forming each array with the constraint that no face would appear twice in one array. In total 200 arrays were created per participant, with 40 arrays for each gender ratio level (0 men: 12 women; 3 men:9 women, 6 men:6 women; 9 men:3 women; and 12 men:0 women), 20 composed of masculine men's faces, and 20 composed of feminine men's faces. Each face in the array was randomly placed into a 3x4 grid. For each trial, participants saw the array for 500 ms and then were presented with the question. For each participant, they rated the same set of 200 arrays, that is, the array creation process occurred once at the start of the experiment and participants then saw the same 200 arrays across the three blocks, presented in a random order.

After completing the three blocks of judgments, participants provided demographic information and were debriefed.

Results

I first present numeric estimations, followed by gender typicality and lastly threat judgments. For all analyses, I used hierarchical linear models. I included as fixed effects the actual Number of Men in the array (mean-centered) and Gender Typicality, that is whether the men's faces were masculine or feminine (effect coded, -1 = feminine, 1 = masculine), as well as the interaction between Number of Men and Gender Typicality. Since stimuli were randomly generated only participant was included as a random factor. Models included random intercepts and random slopes however random slopes were removed if convergence issues emerged, if models resulted in singular fit, or if reduced models did not result in lower model fit as assessed by deviance values from maximum likelihood estimation ANOVA tests being significant. Exact model specification is reported within results. Analyses were conducted in R using lmer and lmerTests (Bates et al., 2015; Kuznetsova et al., 2015). For all analyses, results indicated no significant interaction with Participant Gender and thus this factor was removed.

Numeric Estimations

A model with random slopes for Gender Typicality and Number of Men was fit to the data. Results revealed that as the Number of Men in the array increased, so did participants' numeric estimates of men in the ensemble, $B = 2.56$, $SE = 0.08$, $t(97.69) = 33.14$, $p < .001$, 95% CIs [2.41, 2.71] (see Figure 5). There was also a significant effect for Gender Typicality with arrays that contained feminine faces resulting in lower estimates of men in the ensemble, $B = -0.36$, $SE = 0.03$, $t(99.79) = -13.79$, $p < .001$, 95% CIs [-0.42, -0.31]. Critically, there was a significant interaction between the Gender Typicality and Number of Men, $B = -0.31$, $SE = 0.02$,

$t(13957.11) = -13.38$, $p < .001$, 95% CIs [-0.40, -0.27]. Simple slopes indicated that for masculine men's faces, there was a steeper slope for Number of Men, $B = 2.87$, $SE = 0.08$, $t(135.13) = 34.30$, $p < .001$, 95% CIs [2.71, 3.04], compared to feminine men's faces, $B = 2.25$, $SE = 0.08$, $t(99.33) = 28.95$, $p < .001$, 95% CIs [2.09, 2.40].

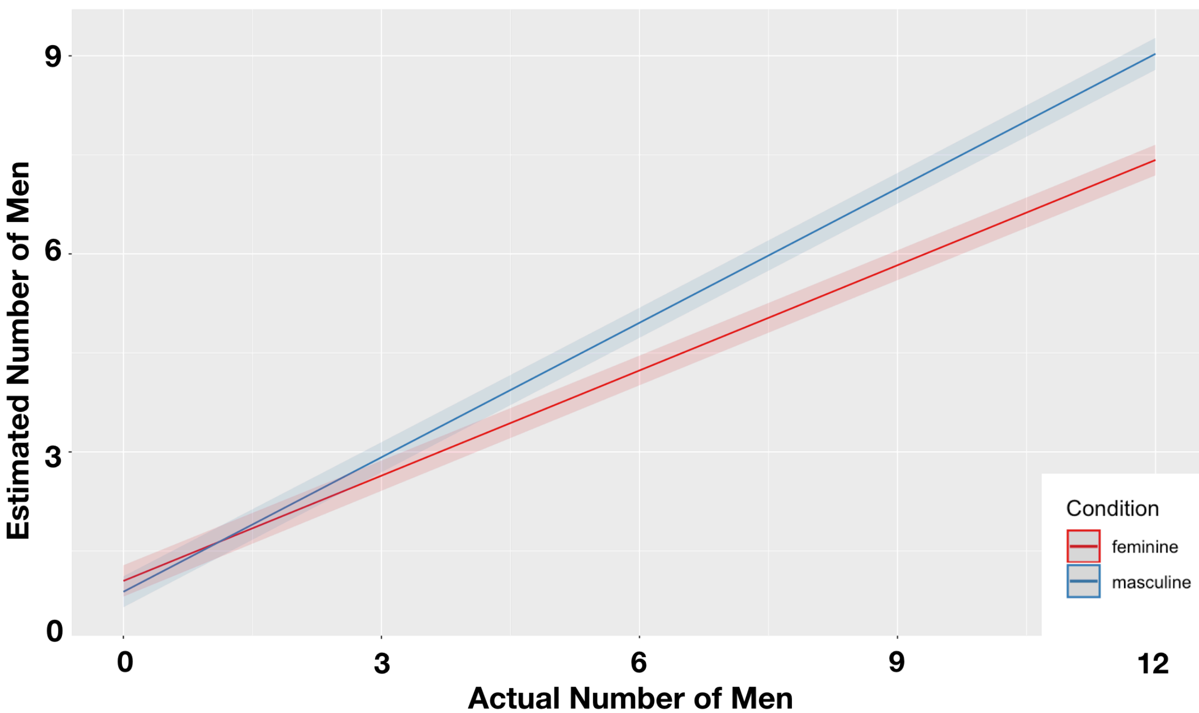

Figure 5. Estimated Number of Men by Actual Number of Men and whether arrays contained Feminine or Masculine men. Bands represent 95% confidence intervals.

Average Gender Typicality

Similar results were obtained for average gender typicality, again with a model which included random slopes for Number of Men and Gender Typicality. As the Number of Men in the array increased, so did participants' ratings of average gender, $B = 2.91$, $SE = 0.14$, $t(96.73) = 21.17$ $p < .001$, 95% CIs [2.64, 3.18] (see Figure 6). There was also an effect of Gender Typicality, such that arrays which contained feminine faces resulted in lower scores on average gender, *B* = -0.43, *SE* = 0.04, $t(98.18) = -9.78$, $p < .001$, 95% CIs [-0.51, -0.34]. These results were qualified by a significant interaction between the Number of Men and Gender Typicality, *B* = -0.36, *SE* = 0.04, $t(13860.61) = -9.96$, $p < .001$, 95% CIs [-0.43, -0.29]. Simple slopes indicated that when the men's faces were masculine, there was a steeper slope for Number of Men, *B* = 3.27, *SE* = 0.15, $t(123.76) = 22.36$, $p < .001$, 95% CIs [2.98, 3.56] compared to when the men's faces were feminine, *B* = 2.55, *SE* = 0.14, $t(97.94) = 18.50$, $p < .001$, 95% CIs [2.28, 2.82].

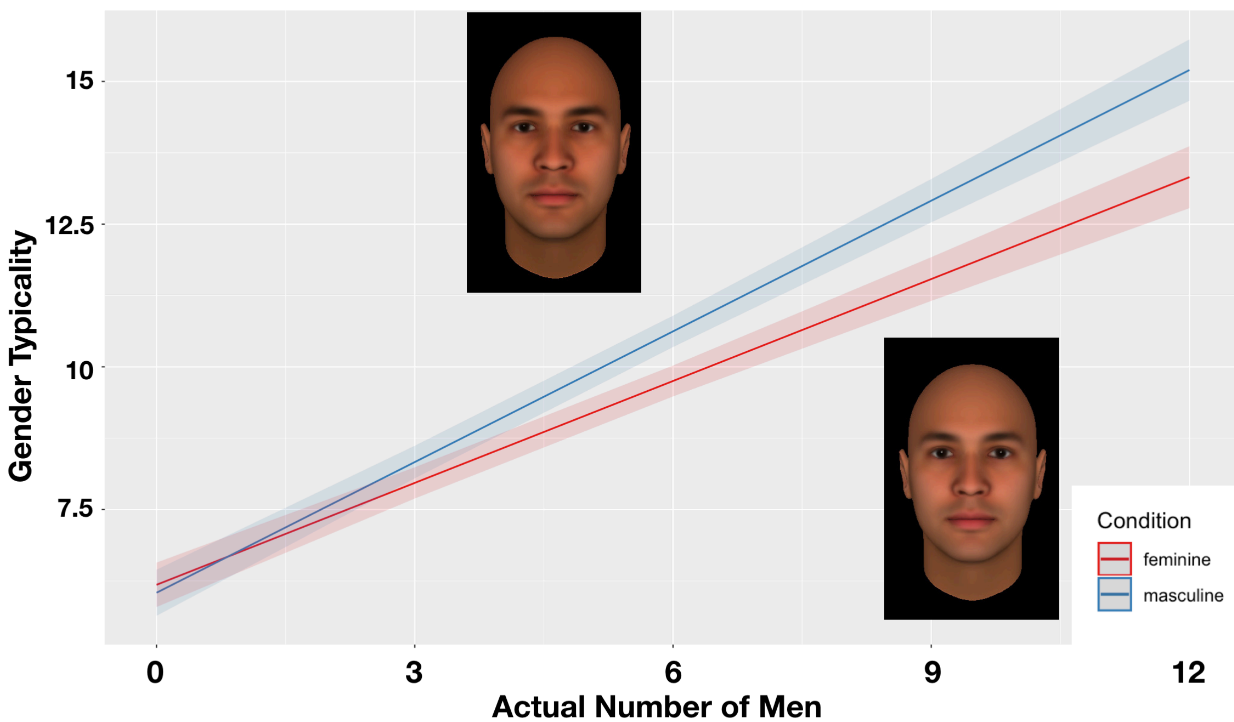

Figure 6. Average Gender Typicality Rating by Actual Number of Men and whether arrays contained Feminine or Masculine men. Depicted faces represent faces from gender typicality scale for value shown from 12 men:0 women groups. Bands represent 95% confidence intervals.

Threat Ratings

A mean score was formed from the four evaluative items with warmth and invitingness reverse coded. A model which included random slopes for Number of Men and Gender Typicality revealed that as the Number of Men in the array increased, participants judged the array as more threatening, $B = 0.19$, $SE = 0.02$, $t(96.42) = 9.65$, $p < .001$, 95% CIs [0.15, 0.23] (see Figure 7). There was also an effect of Gender Typicality, such that arrays which contained feminine faces resulted in the array being viewed as less threatening, $B = -0.04$, $SE = 0.01$, $t(97.99) = -6.39$, $p < .001$, 95% CIs [-0.06, -0.03]. These results were qualified by a significant interaction between the Number of Men and Gender Typicality, $B = -0.02$, $SE = 0.01$, $t(13769.93) = -3.64$, $p < .001$, 95% CIs [-0.03, -0.01]. Simple slopes indicated that when the men's faces were masculine, there was a steeper slope for Number of Men, $B = 0.21$, $SE = 0.02$, $t(130.69) = 9.93$, $p < .001$, 95% CIs [0.17, 0.25] compared to when the men's faces were feminine, $B = 0.17$, $SE = 0.02$, $t(97.93) = 8.55$, $p < .001$, 95% CIs [0.13, 0.21].

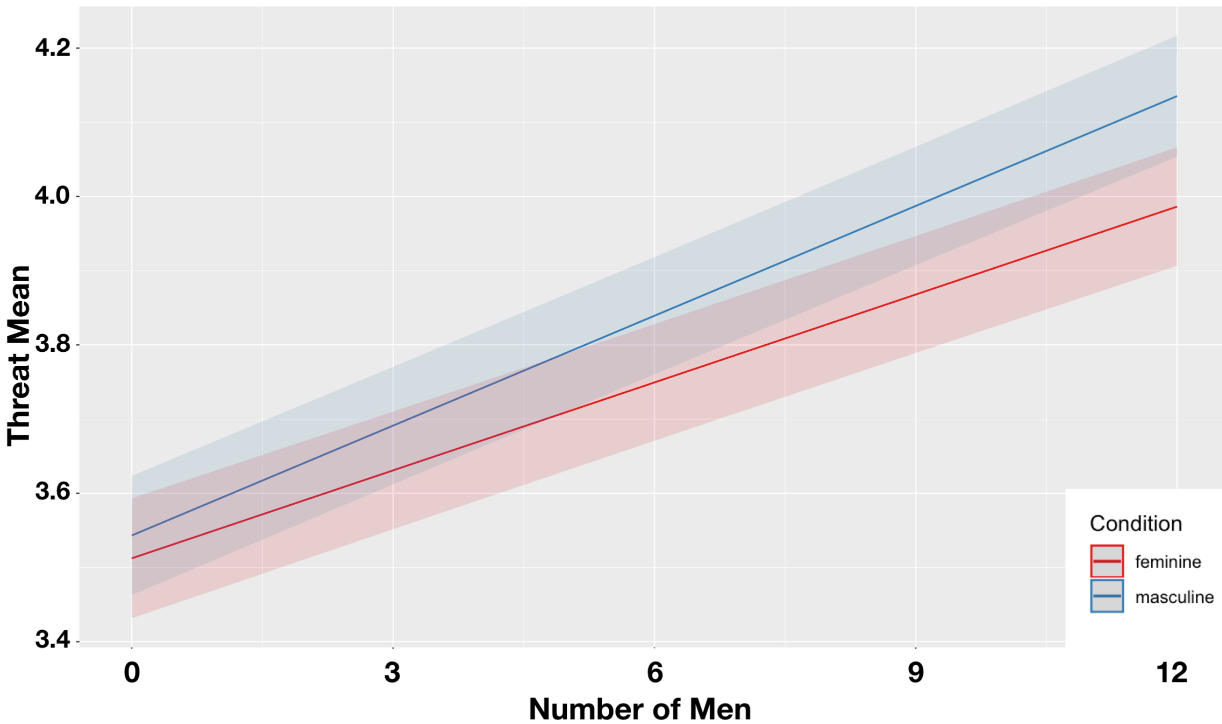

Figure 7. Average Threat Rating by Actual Number of Men and whether arrays contained Feminine or Masculine men. Bands represent 95% confidence intervals.

Discussion

Study 4 revealed that perceivers were sensitive to bottom-up visual cues to gender and critically that these cues impacted perceptual and trait judgments about groups. Specifically, results indicated that as the number of men in the array increased, steeper slopes were found for arrays composed of masculine men compared to feminine men across a number of measures—participants' estimated number of men, average gender typicality, and threat ratings. Together with Study Set 1, this work supports the assertion that a people construal framework must incorporate both the perception of social categories (e.g., this is a group mostly of men) and visual cues to within-category difference (e.g., this is a group of mostly masculine men).

Results from Study 4 aligned with tenants from models of person perception, namely the influence of bottom-up visual cues, here demonstrated with facial morphing changes in gender typicality (Freeman & Ambady, 2011). Again, I found that following surprisingly brief visual presentations (500 ms) perceivers were sensitive to a group's gender ratio and visual cues to gender typicality. As social vision researchers have argued, these phenotypic visual cues are vital to how we perceive the person (Johnson & Adams, 2013), and here I show they are also vital to how we perceive a group, affecting social category-based judgments as well as group evaluations. These results also inform ensemble perception findings as they highlight perceivers flexibility in integrating visual information across a much larger range of identities (66 unique faces) as previously examined (4 and 7 unique faces, Bai et al., 2015; Phillips et al., 2018). While our measures were not as precise with regards to the exact assessment of the average percept (Bai et al., 2015), the addition of a gender typicality measure results in a close correlate to ensemble averaging. The other advantage of the gender typicality measure was that it reduced the potential influence of using a label when asking for numeric estimations (e.g., "men" as opposed to "women"). The fact that findings were similar across both a visual scale and numeric estimations suggests perceivers do note differences without having to use an explicit category label (Macrae & Bodenhausen, 2000). In sum, Study 4 extends people construal by focusing on the capacity and impact of perceiving bottom-up visual cues to gender typicality when making perceptual and social evaluations from groups.

Study Set 2: Visual People Construal and Social Category Activation

Studies 1-4 investigated perceptually derived numeric estimates and trait judgments of groups, however, did not address processes associated with social category activation. A tenant of person construal research is that social category information is extracted from targets with relative ease, subsequently activating category-based knowledge which reduces the complexity of our social world (Macrae & Bodenhausen, 2000). This notion, applied to visually perceived groups, has immense implications for downstream consequences such as stereotyping and prejudice (Lick & Johnson, 2016). For instance, Cooley and Payne (2017), using the AMP, demonstrated that groups of three Black men were implicitly associated with more negative judgments compared to a single Black man. Still, prior research has yet to investigate whether visually perceived groups varying in gender ratio facilitate social category activation, the topic addressed in Studies 5 and 6.

Within this context, I also build upon Study 4, examining the role of bottom-up visual cues to social categories (i.e., gender typicality). As shown in Study 4, differences in visual cue utilization emerged even after a short visual presentation (500 ms); reflecting the importance of differentiating visual cues and social categories. To this end, I used social cognition methods employed by Macrae and Martin (2007) who found that by varying prime duration, different patterns of category activation emerged. In their study, participants, after seeing a prime stimulus for a very brief presentation (20 ms) showed mismatched category activation due to features (i.e., hairstyle) rather than the prime's gender category (i.e., man or woman). These findings showcased the potency of gendered feature cues, such as hairstyle, in leading to erroneous social category activation. Additionally, recent studies using mouse-tracking software, demonstrated that social categorization decisions about gender atypical men and women resulted in partial opposite-category activation (e.g., a feminine man's face resulted in attraction to the category label "woman" to a greater extent compared to a masculine man's face, Freeman et al., 2008). Thus, while overall categorization accuracy was high, visual cues to gender swayed the decision process. In summary, in Study Set 2, I investigate people construal in terms of social category activation, answering the question—do groups activate social category knowledge in a manner similar to past person construal research.

Study 5

In Study 5 I borrowed methods from Macrae and Martin's (2007) work examining the effects of social categories versus visual cues on category activation. In their study, they found that by varying presentation duration, short (20 ms) versus long (200 ms), participants showed different social category activation patterns reflecting either social category use or visual cue use. Specifically, their results showed that presenting a prime for a short duration led to priming via visual cues (e.g., women with short hair led to faster categorization of men's names) compared to long primes which led to priming via social category (e.g., women, regardless of hair length, led to faster categorization of women's names). By manipulating prime duration (short versus long), men's gender typicality (masculine versus feminine), prime (majority men versus majority women groups) and assessing gender categorization of a subsequent target (man or woman's face), I tested a similar process with groups. Specifically, I hypothesized that majority masculine men groups, compared to majority feminine men groups, will facilitate categorization of a target man's face, when these groups are presented at a short prime duration, that is, when visual cues should be most impactful (Macrae & Martin, 2007). Furthermore, I anticipate that majority women groups, regardless of men's gender typicality and prime duration, will facilitate categorization of a target woman's face (i.e., a traditional priming effect). In sum, I explored the independent and interactive effects of a groups' gender ratio and visual cues to gender typicality, demonstrating how both operate to activate social categories.

Method

Participants. Following standard lab procedures, I recruited UCLA undergraduates using the Communications SONA subject pool. Data collection began at the start of the Spring 2019 quarter and continued until halfway through the quarter.[5] In this time, I collected data from 55 participants. Due to computer errors, 14 participants were excluded as incomplete data was recorded, leaving a sample size of 41 participants (33 women, 8 men, $M_{age} = 19.10$, $SD_{age} = 1.07$, 20 White/European American, 13 Asian/Asian American, 3 Black/African American, 5 did not respond to the race/ethnicity question).

Stimuli. For this study, there were two types of stimuli: primes and targets. The primes consisted of 3x4 arrays with a gender ratio of 9 men:3 women (Majority Men) and 3 men:9 women (Majority Women). Within each of these arrays, I used faces from Study 4, which varied in men's gender typicality (Masculine men or Feminine men). Thus, there were four types of primes: majority men/masculine men, majority men/feminine men, majority women/masculine men, and majority women/feminine men. These primes were generated in a manner similar to Studies 1-3 and the groups were created from the pools of available faces from Study 4. In total, there were 80 primes per type, totaling 320 primes.

For the targets, I used a set of men's and women's faces from the Center for Vital Longevity Face Database (Minear & Park, 2004). Target faces were only of White men and women and were a standard image size with neutral facial expressions.

Procedure. Overall, the study was a 2 (Prime: majority men, majority women) by 2 (Target: match, mismatch) by 2 (Gender Typicality: masculine men, feminine men) by 2 (Prime Duration: 200 ms, 1000 ms) repeated measures design. Each trial in this task proceed as follows:

[5] Due to dissertation deadlines data analysis occurred before complete data collection was finished. Data collection, however, continued in order to reach participant numbers comparable to other studies presented in this dissertation.

participants saw a fixation cross (500 ms), followed by a prime array (majority men/masculine men, majority men/feminine men, majority women/masculine men, majority women/feminine men) for a set duration (200 ms or 1000 ms). After viewing the prime array, participants saw a target face (man or woman) and were asked to categorize the target face as either a man or a woman by pressing the "e" or "i" key, counterbalanced across participants. The target face remained on the screen until a response was made. The inter-trial interval was 1500 ms. There were two counter-balanced blocks in which primes were presented for either the short (200 ms) or long (1,000 ms) duration. All prime-target pairings were random with an even split across the two blocks, such that there were 40 trials per prime array type, resulting in 160 trials per block and 320 trials total.

Results and Discussion

The dependent measure was reaction time to categorize the target (i.e. the subsequently presented face) as either a man or a woman. Following Macrae and Martin (2007), trials were excluded if an incorrect categorization was made (3.1%, n = 824 out of 26240 trials). Additionally, reaction times below 50 ms (n = 5) and three standard deviations above the mean reaction time (RT_{mean} = 704.035, RT_{sd} = 505.914, n = 355) were excluded. This left a total of 24,640 trials.

A crossed hierarchical linear model including all predictors (effect-coded) and all higher order interactions was fit to the data. Predictors included Prime (majority men, majority women), Target (match, mismatch, e.g., a match for majority men primes would be a man's face), Gender Typicality (masculine men, feminine men) and Prime Duration (200 ms, 1000 ms). The model included random intercepts for both participants and for target. Results indicated a significant main effect of Duration, B = -28.02, SE = 1.69, t(24514.52) = -16.61, p < .001, 95% CIs [-31.33,

-24.72] and a significant Target by Match interaction, $B = -7.51$, $SE = 2.27$, $t(77.71) = -3.31$, $p = .001$, 95% CIs [-11.98, -3.03]. Critically these results were qualified by the predicted four-way interaction across Prime, Target, Gender Typicality, and Prime Duration, $B = 3.79$, $SE = 1.68$, $t(24511.20) = 2.25$, $p = .024$, 95% CIs [0.50, 7.09].

Decomposing this four-way interaction, I first examined the three-way interaction, Target by Gender Typicality by Prime Duration, separately for Prime—Majority Men and Majority Women. Since my hypothesis centered on the effect of visual cues (i.e., Gender Typicality) on category activation, it is reasonable that Majority Men trials (9 men in the group) would be more likely to show the effect of (Men's) Gender Typicality compared to Majority Women trials (only 3 men in the group).

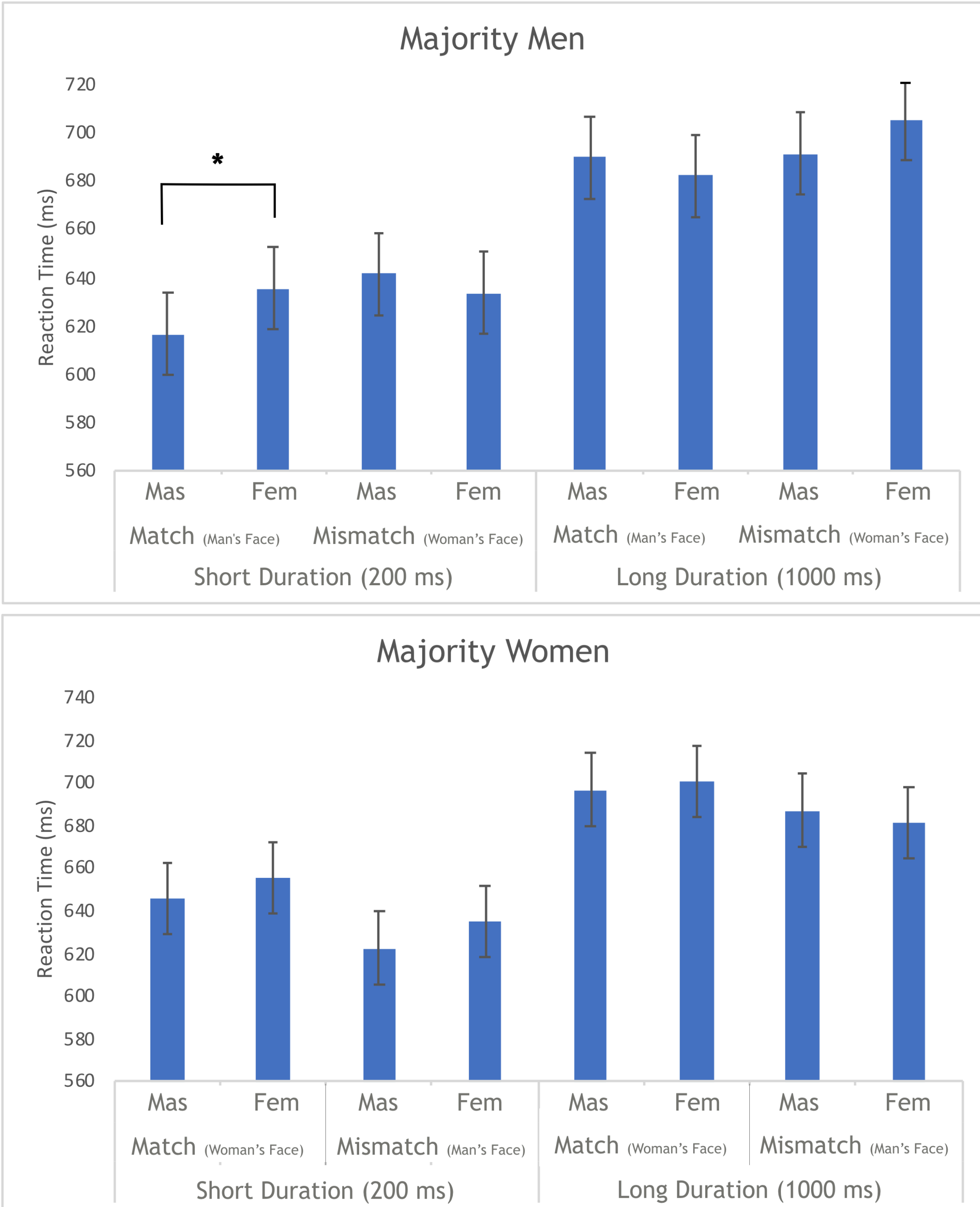

Figure 8. Reaction times by Trial Type (Majority Men, *top*, Majority Women, *bottom*), Prime Duration (Short, Long), Target (Matched Prime, Mismatched Prime), and Men's Gender Typicality (Masculine, Feminine). *$p < .05$.

Results for Majority Men primes revealed a significant effect of Duration, B = -30.21, SE = 2.37, t(12227.43) = -12.77, p < .001, 95% CIs [-34.85, -25.57] and a significant effect of Target, B = -5.98, SE = 2.78, t(77.93) = -2.15, p = .035, 95% CIs [-11.46, -0.50] however these were qualified by a significant three-way interaction across Target, Gender Typicality, and Duration, B = 6.06, SE = 2.36, t(12223.76) = 2.57, p = .010, 95% CIs [1.44, 10.69] (see top of Figure 8). Further decomposing of this three-way interaction by Duration, showed that for the 1,000 ms duration there were no significant effects, ps > .118, suggesting no differences in reaction times for gender categorizations targets depending on prime at this duration.

However, for the 200 ms duration, wherein based on Macrae & Martin (2007) visual cues should have greater impact, results revealed a significant two-way interaction between Target and Gender Typicality, B = 6.88, SE = 3.12, t(5966.64) = 2.21, p = .027, 95% CIs [0.77, 13.00] such that for matching trials, participants were faster to categorize the target as a man when the prime was a majority masculine men group compared to a majority feminine men group, B = -18.72, SE = 8.78, Z = 2.13, p = .033, 95% CIs [-35.93, -1.50] and for mismatch trials no significant effect of Gender Typicality emerged, B = 8.81, SE = 8.86, Z = 0.99, p = .320, 95% CIs [-8.56, 26.18].

Results for Majority Women primes revealed a significant effect of Duration, B = -25.84, SE = 2.41, t(12180.60) = -10.74, p < .001, 95% CIs [-30.55, -21.12] such that participants were faster to categorize targets after seeing a 200 ms prime compared to a 1000 ms prime, and a significant effect of Target, B = 9.02, SE = 2.88, t(77.73) = 3.13, p = .002, 95% CIs [3.34, 14.69]. Contrary to expectations, this effect indicates participants were faster to categorize non-matching trials (i.e., men's faces) compared to matching trials (i.e., women's faces) following majority women primes. No other effects were significant, ps > .050 (see bottom of Figure 8).

To summarize, results indicated that participants showed facilitated category activation for men's faces following majority masculine men group primes compared to majority feminine men group primes, specifically when the primes were presented for a short duration (200 ms), that is, when visual cues should be most impactful (Macrae & Martin, 2007). While our pattern of results did not show a reversal in category activation due to visual cues, such as found by Macrae and Martin (2007), our results do align with the broader idea that participants are sensitive to and utilize visual cue information when it comes to social category activation following group primes.

Interestingly, the other significant effect of note was faster categorization of men's faces (i.e., mismatch trials) following majority women primes, regardless of men's gender typicality and prime duration. This effect is at odds with predictions, however, speculatively, it could be due to participants consistently seeing changes in men's facial cues while women's facial cues remained static (e.g., all feminine). This constant perceptual change in men's faces may lead participants to focus greater attention on men, and thus be primed to make faster men categorizations of targets, regardless of which category is dominant in the array. This notion, however, is rather speculative.

Study 6

Study 5 revealed evidence that visual gender typicality cues within groups facilitated social category activation. Specifically, that men's faces were categorized faster following majority masculine men groups, compared to majority feminine men groups, at a short prime duration. Although informative, these results do not take into account how group *categorization* decisions unfold over time and the potential category competition that could arise from visual gender typicality cues versus social category based people construal. To assess the underlying decisional processes that occur when social category decisions are made about visually perceived groups, I conducted a mouse-tracking study (Freeman & Ambady, 2010). This unique methodology allows for an examination into the underlying cognitive decision-making process associated with people construal and social category activation of groups.

Method

Participants. I recruited participants through the UCLA Communication Department SONA. In total, 89 participants were recruited (68 women, 20 men, 1 did not answer the gender demographic question, $M_{\text{age}} = 19.07$, $SD_{\text{age}} = 0.97$, 36.0% Asian/Asian American, 36.0% White/European American, 13.5% Latinx American, 9.0% Multiracial, 4.4% Black/African American).

Stimuli. Stimuli were twelve person groups that varied in gender ratio—majority men (9 men:3 women) and majority women (3 men:9 women). These groups also varied in men's gender typicality (masculine or feminine), with faces drawn from the pool generated in Study 4. Overall, 480 stimuli were created, 120 per type of stimulus (i.e., majority men/masculine men; majority men/feminine men; majority women/masculine men and majority women/feminine men).

Procedure. To record participants' mouse movements, I used the software package MouseTracker (Freeman & Ambady, 2010). This software tracks, in real-time, the computer mouse as participants move the mouse to select one of two options. Participants were instructed that they would see a group of people appear on screen, and that their task was to indicate, as quickly and as accurately as possible, whether the group was composed of "mostly men" or "mostly women". To make this judgment, they selected one of two labels presented with these options in the upper left and upper right of the screen (counter-balanced across participants). To ensure that participants moved the mouse while making their decisions, as opposed to making the decision then initiating mouse movement, a message appeared telling participants to initiate movement faster if participants' mouse movements did not begin within 400 ms post-stimuli presentation (see Hehman, Stolier, & Freeman, 2015).

Each trial began with a "Start" button presented at the bottom center of the screen. Participants clicked that button to begin the trial which also served to center the computer mouse at a standard *x*- and *y*-coordinate (0, 0). Target groups were randomly selected from the four possible target types and participants categorized the target using the "mostly men" or "mostly women" labels (counter-balanced at the participant level on the left or right). Target groups were displayed for either a short (200 ms) or long (1000 ms) duration. This manipulation mirrored Study 5 to compare the influence of visual phenotypic (i.e., gender typicality) versus social category (i.e., gender) information at different stimulus presentation times. All participants completed six practice trials before the experiment began to familiarize themselves with the procedure. In total, participants completed 160 trials, presented in a random order (40 per trial type), with a self-timed break halfway through the trials.

Results

Data Preparation and Trial Exclusions. MouseTracker software allows for multiple measurements to be collected, including participants' judgment errors, categorization reaction time, and spatial attraction to the alternative category selection. For reaction times and spatial attraction, I only used trials for which correct categorizations were made, this resulted in excluding 2854 trials, however 11,386 trials remained (or approximately 80% of the data). Additional to the exclusions described above, I also excluded reaction time and spatial attraction trials that were three standard deviations (SD) above or below their respective means.

For all analyses, I used hierarchical linear models to account for the nested structure of judgments within participants. Participant was the random factor while Target (Majority Men, Majority Women), Gender Typicality (masculine, feminine) and Duration (200 ms, 1,000 ms) were included as fixed factors (all effect coded, -1 and 1) as well as all possible interactions. All models included random intercept and random slopes unless model convergence issues or overfitting emerged in which case random slopes were removed until models converged and/or appropriate fit was reached. All analyses indicate which model was ultimately fit to the data.

Judgments. A hierarchical linear model with random intercept and random slope for Target was fit to a binomial distribution (correct judgments = 0, incorrect judgments = 1). Results revealed that participants were less likely to make a categorization error after the 1,000 ms compared to 200 ms presentation time, $B = -0.40$, $SE = 0.02$, $Z = -17.68$, $p < .001$. Additionally, a significant effect for Target revealed that participants were more likely to error for Majority Men compared to Majority Women trials, $B = 0.19$, $SE = 0.04$, $Z = 4.96$, $p < .001$ and a significant effect for Gender Typicality indicated that participants were more likely to error for feminine compared to masculine trials, $B = 0.11$, $SE = 0.02$, $Z = 4.76$, $p < .001$. Results revealed two significant two-way interactions. The first significant two-way interaction was between Duration and Target, $B = 0.07$, $SE = 0.02$, $Z = 2.92$, $p = .004$, the second two-way interaction was between Target and Gender Typicality, $B = 0.20$, $SE = 0.02$, $Z = 8.73$, $p < .001$. These two-way interactions were qualified by a significant three-way interaction across Duration, Target, and Gender Typicality, $B = 0.05$, $SE = 0.02$, $Z = 2.03$, $p = .043$ (see Figure 9). Decomposing this significant three-way interaction, results indicated that for both 1,000 ms and 200 ms there were significant Target by Gender Typicality interactions, although this effect was larger for the 1,000 ms duration, $B = 0.25$, $SE = 0.04$, $Z = -6.88$, $p < .001$ compared to the 200 ms duration, $B = 0.15$, $SE = 0.03$, $Z = 5.39$, $p < .001$. Further breaking down these two-way interactions, revealed that for Majority Men groups at the 1,000 ms duration, there was a significant effect of Gender Typicality, $B = 0.79$, $SE = 0.09$, $Z = 8.48$, $p < .001$, with more errors when categorizing majority feminine men groups compared to majority masculine men groups. There was no significant effect of Gender Typicality for Majority Women groups at 1,000 ms, $B = -0.20$, $SE = 0.11$, $Z = -1.80$, $p = .071$.

At 200 ms, for Majority Men groups, there was also a significant effect of Gender Typicality, $B = 0.43$, $SE = 0.08$, $Z = 5.70$, $p < .001$, demonstrating a similar pattern as above, with more errors when categorizing majority feminine men groups compared to majority masculine men groups. There was also a significant effect of Gender Typicality for Majority Women groups at 200 ms, $B = -0.16$, $SE = 0.08$, $Z = -2.03$, $p = .042$, reflecting a pattern whereby participants made more errors when categorizing majority women groups if that group contained three masculine men compared to three feminine men. In summary, participants' pattern of errors when making majority men or majority women categorizations reflected the impedance of gender typicality cues. Specifically, majority feminine men groups produced significantly more errors than majority masculine men groups (i.e., when gender typicality cues misaligned), with this difference greater for the 1,000 ms compared to 200 ms viewing duration. Additionally, for majority women groups, participants errored more when the men in the group were masculine (misaligning with the category 'mostly women') compared to feminine.

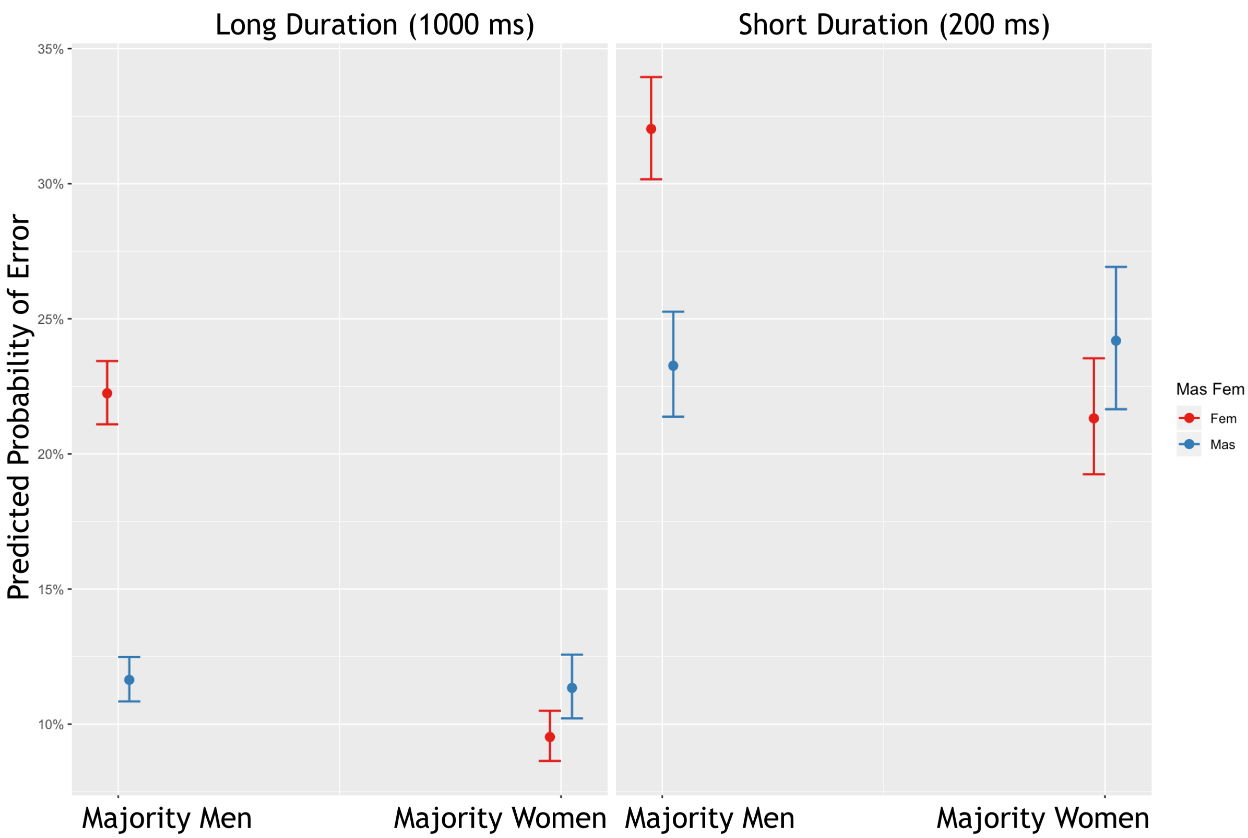

Figure 9. Predicted Probability of Error by Prime Duration (Long, Short), Target (Majority Men, Majority Women) and Men's Gender Typicality (Masculine, Feminine)

Reaction Time. I hypothesized that reaction times for categorizing groups would be faster when visual gender typicality cues matched social category information (e.g., masculine men arrays compared to feminine men arrays). In addition, I anticipated that this effect would be stronger for majority men arrays, as there would be a greater number of men in such arrays compared to majority women arrays. I also examined whether prime duration would affect results, aligning with Study 5 results.

There were 151 trials that exceeded three SD above the mean for reaction time ($M_{\text{reaction time}}$ = 1048.62, $SD_{\text{reaction time}}$ = 461.96) and thus these trials were excluded. Results from a full model (random intercept and random slope for Target and Duration) revealed no significant three-way interaction for Target, Gender Typicality, and Duration, $B = -2.29$, $SE = 2.58$, $t(10988.34) = -0.89$, $p = .375$, 95% CIs [-7.35, 2.77], as well as no significant interaction between Duration and Target ($p = .155$) and Duration and Gender Typicality ($p = .722$). Thus, for model parsimony, I reported the model results without the interaction terms for Duration.

Results from this reduced model (random intercepts and random slopes for Target and Duration), revealed only a significant two-way interaction between Target and Gender Typicality, $B = 6.27$, $SE = 2.58$, $t(10996.91) = 2.43$, $p = .015$, 95% CIs [1.22, 11.32] (see Figure 10). This two-way interaction was decomposed by examining the effect of Gender Typicality in the Majority Men and Majority Women arrays. Results, using custom contrasts (i.e., matrix(c(0, 0, 0, 2, [2 or -2]), 1), revealed that in the Majority Women arrays, there was no significant effect of Gender Typicality, $B = -5.57$, $SE = 7.14$, $z = -0.78$, $p = .435$, 95% CIs [-19.56, 8.43] however for the Majority Men arrays, there was a significant effect of Gender Typicality, $B = 19.51$, $SE = 7.43$, $z = 2.63$, $p = .009$, 95% CIs [4.94, 34.07], such that participants were faster to categorize masculine men arrays ($M = 999.55$, $SE = 13.68$) compared to feminine men arrays ($M = 1020.91$, $SE = 13.70$). These findings align with hypotheses such that when visual gender typicality information was most present (majority men groups), this visual cue information resulted in faster categorizations when cues aligned (i.e., masculine men) compared to when they misaligned (i.e., feminine men).

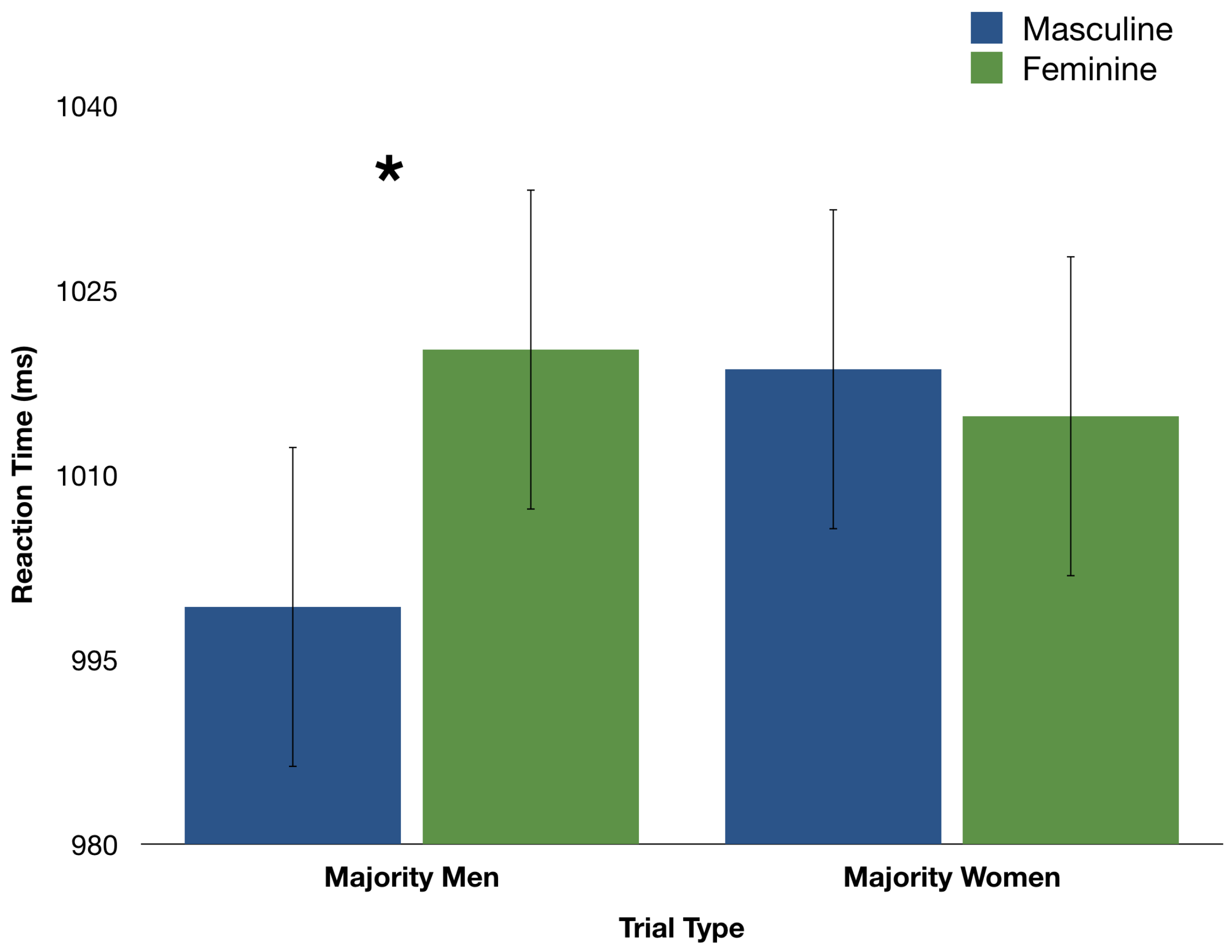

Figure 10. Reaction times to categorize target by Trial Type (Majority Men, Majority Women) and Men's Gender Typicality (Masculine, Feminine). *$p < .05$

Spatial Attraction. To assess spatial attraction, I followed standard procedures (Hehman et al., 2015), scaling mouse trajectories onto a standard coordinate space (top-left: "-1, 1.5"; bottom-right: "1, 0") with the bottom center at the "0, 0" trial starting position. Additionally, the trajectories were fit to 101-time steps series with linear interpolation. To assess the degree to which "incorrect" category activation emerged (e.g., a majority feminine men group showing greater partial activation to the “mostly women” label compared to a majority masculine men group) I calculated, by trial, the area-under-the-curve (AUC). This AUC consists of finding the maximum deviation, and then taking the area between the idealized response (i.e., a straight direct line to the correct category label) and the actual mouse trajectory.

I hypothesized that when categorizing groups with manipulated visual gender typicality cues most present (e.g., majority men arrays), misaligned gender typicality would create greater opposite category attraction, that is majority feminine men groups should have higher AUC than majority masculine men groups.

There were 406 trials which exceeded three SD above the mean ($M_{\text{AUC}} = 0.43$, $SD_{\text{AUC}} = 1.11$) and thus these trials were excluded. Results from a full model (random intercept and random slope for Target and Duration) revealed no significant three-way interaction for Target, Gender Typicality, and Duration, $B = 0.01$, $SE = 0.01$, $t(10745.14) = 1.24$, $p = .217$, 95% CIs [-0.01, 0.02] as well as no significant interaction between Duration and Target ($p = .066$) and Duration and Gender Typicality ($p = .646$). Thus, as with reaction times, I reported the model results without the interaction terms for Duration.

Results from this reduced model (random intercepts and random slopes for Target and Duration) revealed a significant effect of Duration, $B = -0.08$, $SE = 0.01$, t(84.15) = -8.61, $p < .001$, 95% CIs [-0.10, -0.06] such that participants who viewed arrays for longer durations had more direct categorizations compared to participants who viewed the arrays for shorter durations. Results also revealed a significant two-way interaction between Target and Gender Typicality, $B = 0.02$, $SE = 0.01$, $t(10752.07) = 2.57$, $p = .010$, 95% CIs [0.004, 0.03]. This two-way interaction (see Figure 11) was broken down by separately examining the effect of Gender Typicality in the Majority Men and Majority Women arrays. Results, using custom contrasts (i.e., matrix(c(0, 0, 0, 2, [2 or -2]), 1), revealed that in the Majority Women arrays, there was no significant effect of Gender Typicality, $B = -0.03$, $SE = 0.02$, $z = -1.52$, $p = .129$, 95% CIs [-0.07, 0.01] however for the Majority Men arrays, there was a significant effect of Gender Typicality, $B = 0.04$, $SE = 0.02$, $z = 2.10$, $p = .035$, 95% CIs [0.003, 0.09] such that participants had higher AUCs when categorizing feminine men arrays ($M = 0.24$, $SE = 0.02$) compared to masculine men arrays ($M =0.19$, $SE = 0.02$). These findings align with hypotheses such that when visual cue information was most present (majority men groups), this visual cue information resulted in greater attraction to the incorrect category label when cues misaligned (i.e., feminine men) compared to when cues aligned (i.e., masculine men).

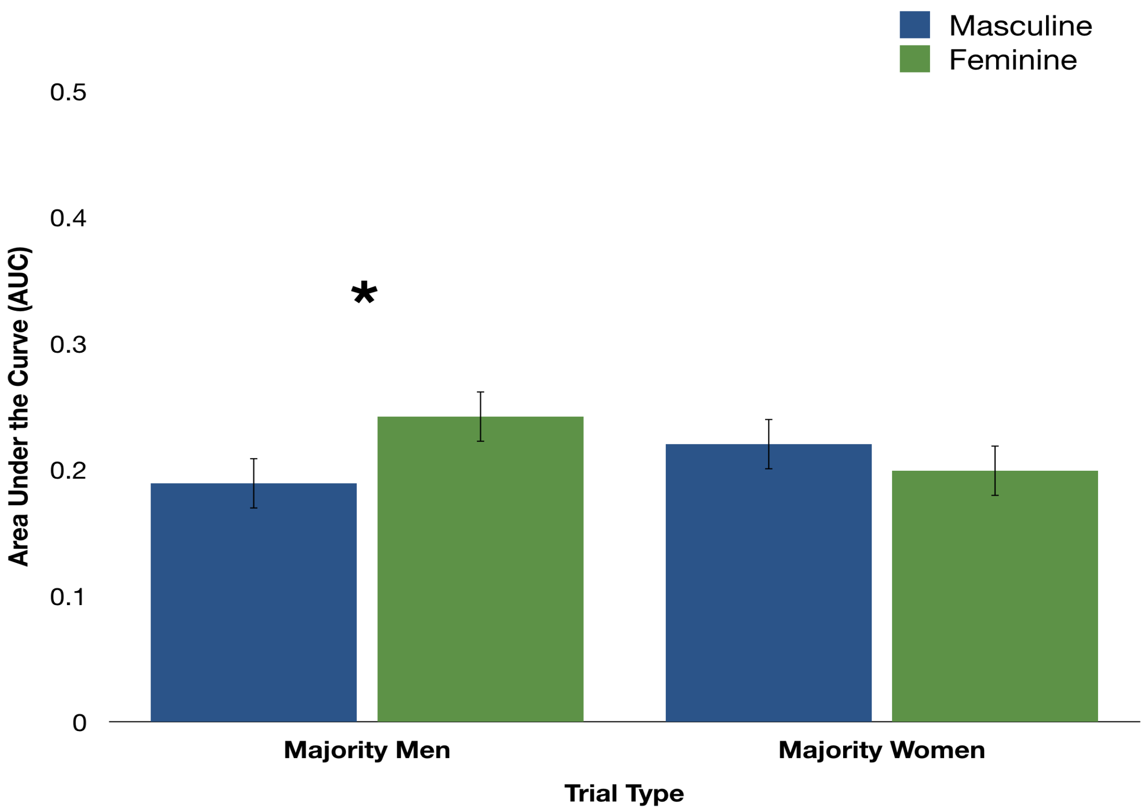

Figure 11. Area Under the Curve (AUC) by Trial Type (Majority Men, Majority Women) and Men's Gender Typicality (Masculine, Feminine). *$p < .05$.

Discussion

Study Set 2 examined people construal processes related to social category activation and social category decision making. In Study 5, by utilizing a sequential priming paradigm, I assessed the degree to which group primes varying in visual cues to men's gender typicality (i.e., masculine/feminine) and majority group gender (i.e., majority men/majority women) facilitated category priming. Results revealed that group gender typicality impacted subsequent category activation when shown for a short duration and for groups with majority men (i.e., when cues should be most utilized, Macrae & Martin, 2007). Specifically, majority masculine men group primes facilitated category activation as indicated by faster reaction times to categorize a subsequently presented man's face, compared to majority feminine men group primes. Study 6 further demonstrated the role of men's gender typicality on participants' social category decisions about groups. Participants, when categorizing a majority masculine men group, compared to a majority feminine men group, had fewer errors, faster reaction times, and more direct social categorizations. Together, these converging results provide evidence that visual gender cues affect social category activation processes associated with groups.

Study Set 2 further situates people construal within established person construal frameworks, advancing our understanding of the visual perception of groups as it relates to social category-based processes. While Studies 1-4 focused on numeric estimations and trait evaluations of briefly presented groups, Studies 5 and 6 examined how the social categories depicted within a group activate and facilitate social categorization. Through both a traditional sequential priming task and a mouse-tracking paradigm results converge on the notion that perceivers integrate visual cues to gender typicality when making assessments regarding groups.

Specifically, when social category information (e.g., men) aligns with visual cue information (e.g., masculine men), perceivers social category activation was facilitated across numerous dependent measures (e.g., reaction time, errors, directness), yet misalignment of these cues hindered category activation and categorization.

Findings did reveal some noteworthy unpredicted effects. From Study 5, results showed that perceivers, regardless of men's gender typicality, showed facilitated categorization of a man's face following majority women group primes. Since these groups have fewer men within them, 3 out of 12 faces, it makes sense men's gender typicality should not impact results. However, since the remaining 9 faces are all of women, majority women groups should facilitate categorization of women's, not men's, faces. I hypothesized that this divergence could be attributed to shifting men's gender typicality, since this change could draw attention to the men's faces throughout the perceptual process, especially compared to the relatively static gender typicality of women's faces (e.g., all were feminine). Still, follow-up tests should explore this notion, possibly by varying women's gender typicality and keeping men's gender typicality constant. In this case, majority men groups should facilitate categorization of women's faces because attention is drawn to the women's shifting gender typicality.

Also, results from Study 6, while indicative of gender typicality impacting participants categorical decisions, did not show interactive effects with stimulus presentation timing, at least for reaction times and area-under-the curve. It may be the case that timing effects specifically impact a prime/target relationship (e.g., see a prime then make a decision about a subsequent target) rather than a pure categorization process (e.g., categorize this target). The change in the nature of the task could have affected how participants utilized and integrated these cues, however the fact that the pattern of results across both studies was indicative of gender typicality affecting performance is still noteworthy as it suggests integration of these cues when visually processing groups.

CONCLUSION

Our social environments are full of people—co-workers mingling in an office, strangers filing into a movie theatre, students strolling on a college campus. Person perception research has long noted that we effortlessly and efficiently sort these people into social categories, which allows for an efficient interpretation of their relevance, motives, emotions, thoughts, and actions (Bruner, 1957; Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000). As the nomenclature suggests, person perception largely treats the person as the object of focus, examining how we come to perceive each individual person within the sea of people. Dating back to Gesalt psychology (Asch, 1946) we know, however, that the whole equals more than the sum of its parts, and my dissertation focused on the larger percept of groups. In proposing and developing a theory of people construal, or how we perceive, evaluate, and make sense of groups within our social world, I integrated social cognition and person construal theory with novel findings from vision science on ensemble perception, to propose and evaluate unique hypotheses about group perception, judgment, and social categorization.

First, I found that perceivers were sensitive to social category differences (i.e., gender) within twelve person groups and that these differences in turn affected explicit and more implicit group evaluations. Specifically, as the gender ratio within a group shifted from predominately women to predominately men, participants reported seeing more men within the group and rated the group as more threatening/hostile and less warm/inviting, aligning with gender stereotype associations. Critically, these results emerged following a very brief visual presentation of the group—500 ms for explicit ratings and 75 ms for the AMP—highlighting the efficiency of our categorical and evaluative group judgments. Additionally, I showed that participants estimated number of men mediated group threat judgments, further highlighting the impact of perceived social categories on group social evaluations.

Second, I tested whether perceivers were sensitive to within-category variability (i.e., gender typicality) for visually perceived groups, and if these visual cues affected group evaluative judgments. To test these questions, I generated a novel stimulus set of masculine and feminine men using face morphing, and then created groups with these faces. Thus, I varied both the representation of social categories (i.e., how many men and women were in a group) and the visual cues that denote gender (i.e., whether the men in the group were masculine or feminine). Results from Study 4 demonstrated that perceivers were sensitive to cues of men's facial gender typicality and that these cues impacted categorical, perceptual, and evaluative judgments. That is, as the number of men to women in a group increased, if the men within the group were masculine, compared to feminine, participants rated the group as having more men, as being more masculine, and as more threatening. These findings provided a more nuanced perspective into how we perceive and integrate not only social categories within groups but also the visual cues underlying social category perception of groups.

Finally, in Study Set 2, I directedly assessed social category activation and social categorization processes for groups varying in gender ratio and men's gender typicality. In Study 5, I conceptually replicated Macrae and Martin (2007)'s work which showed that by varying prime duration (i.e., a short versus a long prime), one can tease apart the role of social category (i.e., gender) and visual feature cues (i.e., long/short hair), such that visual features, at a short prime duration, can override social category information (e.g., a woman with short hair facilitated activation of men, not women). While results did not show a reversal in category activation for groups, the pattern of findings did suggest that visual cues within groups were utilized. Specifically, for the short prime duration, majority masculine men group primes, compared to majority feminine men group primes, facilitated faster reaction times to categorize a subsequently presented face as a man. Thus, when visual cue information should be most impactful (i.e., short prime, majority men), gender typicality cues impacted category accessibility as assessed through reaction times to categorize a target as a man. Finally, Study 6 examined the role of visual cues, again men's gender typicality, on social categorization judgments as these decisions unfolded over time. Using a mouse-tracking paradigm, I assessed participants categorization of groups (i.e., whether a group was majority men or women) that varied in men's gender typicality. Results again suggested that perceivers were sensitivity to visual cue information, with majority feminine men groups producing more errors in categorization, slower reaction times to categorize, and less direct categorization (i.e., greater area-under-the-curve), compared to majority masculine men groups. Together, these results demonstrated perceivers' sensitivity to visual cue information (i.e., men's gender typicality), over and above social category differences (i.e., dominant gender within the group), revealing the critical role for both categorical and visual phenotypic information in people construal.

The Social Vision of Groups

By situating this work at the intersection of social cognition research on person construal and vision science research on ensemble perception I aimed to build an integrated understanding of how we come to perceive and make judgments about groups. Ensemble perception (Whitney & Leib, 2018) provided the foundational evidence that perceivers were capable of group perception. Early work in this area approached these questions by testing low-level stimuli, addressing the accuracy of ensemble perception and establishing ensemble perception as a means to overcome known limitations in terms of object perception (typically set at 3-4 objects, see (Luck & Vogel, 1997; Rensink, et al., 2000; Scholl & Pylyshyn, 1999; Simons & Chabris, 1999). By expanding an "object" to include not only one thing but a group of things these researchers outlined a parsimonious way to explain how we adeptly perceive and move through our complex sensory world (Whitney & Leib, 2018). While vision scientists increasingly addressed more socially relevant stimuli including faces, bodies, and eye-gaze (Haberman & Whitney, 2007; Sweeny, et al., 2013; Sweeny & Whitney, 2014; Thornton et al., 2014), methodological differences such as few participants (offset by a large number of trials) and tight control over stimuli (frequently utilizing a small set of faces) meant the critical focus of this work largely centered on perceptual processes and accuracy. While vital to establishing the capability and operation of group perception, my dissertation studies resituated ensemble perception by focusing on the socio-functional, ecological aspects of perceiving people (McArthur & Baron, 1983), addressing questions not just of how we perceive groups but to what end and why.

In addressing questions regarding the functional purpose of people construal, it was critical to examine tenants from the rich history of person construal. Throughout the presented studies, a number of these tenants guided and were upheld by my research. Of critical importance were processes associated with social categorization, both the use of social categories and how these social categories activated stereotypic knowledge which, in turn, affected downstream evaluations of groups. It is these processes that informed why a group of majority men would be associated with more threat than a group of majority women (Fiske & Neuberg, 1990; Freeman & Ambady, 2011; Hamilton, 1981; Macrae & Bodenhausen, 2000). Additionally, the Dynamic Interactive Theory of Person Construal (Freeman & Ambady, 2011) highlighted the need to examine not just social categories but also the bottom-up visual cues to those social categories. Again, findings paralleled past research, with visual cues to men's gender typicality impacting estimates of the number of men, overall average gender typicality, and trait evaluations (Study 4). Perceivers' social categorization processes were also impacted by such bottom-up cues, as shown by the pattern of results for priming facilitated category activation and categorization decisions of groups (Studies 5, 6).

Overall, this work on people construal represents a novel domain for understanding our daily perception not only of the individual but also individuals with individuals. While it is likely that tenants of person construal research will continue to be upheld when applied to groups, group perception research will also lead to a modification and extension of person construal theory. Simply the fact that we are sensitive to group social category variability (both across categories and within categories) after seeing a group for only half a second is astonishing yet resonates with our lived experience of seeing groups, crowds, and people. It is also telling that these lower level percepts appear to be the factor behind our group trait evaluations, more so at least than higher-order thinking about a group (i.e., entitativity) at least for quick and repeated group judgments (Study 3). Findings from Study Set 2 also both replicate and diverge from prior research. While it is clear visual cues to gender typicality were utilized both in the priming paradigm (Study 5) and for categorization decisions (Study 6), it is also clear that they operate differently compared to person perception. While hairstyle cues and visual gender typicality are clearly different, I did not find a complete reversal of priming effects in Study 5, as suggested by Macrae and Martin (2007). Perhaps more interesting was the divergence when it came to majority women groups, which appeared to show category facilitation to men compared to women's faces, regardless of duration and men's gender typicality within the array. As noted, the fact that men within the group were changing on a constant basis between masculine and feminine might have cued participants to focus on the men, leading to facilitation of men as the category, regardless of gender typicality and the dominant social category within the group (i.e., women). Still, these findings suggest unique effects of seeing such change within a group, a fact illuminated only by focusing on groups as category primes.

Future Directions & Limitations

The domain of people construal is rich in opportunities to test novel questions about the perception of groups. As future research continues, I envision three areas of critical importance. First, and perhaps most obvious, it is important to understand how group perception operates for a variety of social categories as well as the interaction of these social categories. By testing both single social category variability (e.g., race, age), within-social category variability (e.g., racial phenotypically) and intersectional social category variability (e.g., race and gender) novel insights may be gained regarding our perception of people.

Second, there is still much to be known about the exact process by which visual information about groups is integrated and consolidated into a group percept. Research from vision science on ensemble perception is still developing and continued integration between vision science and social psychology is critical to advancing a cogent theory of people construal. Two relevant and novel findings from ensemble perception are how sequential percepts (e.g., showing images one after another, rather than as a group) can also be accurately summarized and how changes in group size (e.g., adding or subtracting elements from a group) may affect group perceptions.

Lastly, it is important to explore more applied aspects of people construal, taking an ecological perspective (McArthur & Baron, 1983) to ask why group perception matters in the first place. I believe, group perception offers critical insights into a variety of applied social psychological applications such as understanding group representativeness and the development of interventions. Here, I spend the rest of this section developing these future directions, highlighting specific areas to direct future work.

The Who's Who within the Group. As this was an initial foray into people construal, I purposefully limited stimuli to White men and women. In doing so, however, this research fell into the dominant paradigm of researching the default category—White individuals when varying gender (Purdie-Vaughns & Eibach, 2008). Clearly other perceptually obvious social categories such as race and age, and even some "perceptually ambiguous" categories such as sexual orientation, are perceivable from the face and of great importance (Milord, 1978; Macrae & Bodenhausen, 2000; Tskhay & Rule, 2013). Simple extensions to this research call for varying the social categories depicted within groups and testing whether similar stereotyped evaluations emerge. Examining differences in stereotype activation and application for groups compared to individuals also suggest some intriguing possibilities. For instance, Cooley and Payne (2018) show that Black and Asian individuals were rated more stereotypically when in groups of other Black and Asian faces then when presented by themselves. These effects suggest that groups play a critical role in modulating our evaluations of individuals, and how stereotypes may become activated and applied to affect evaluations.

Furthermore, it is important to examine the intersection of social categories. As prior work demonstrates race and gender are tethered with both bottom-up visual cues such as facial gender typicality (Carpinella, et al., 2015; Johnson, Freeman, et al., 2012) and top-down stereotypic associations (Galinsky, Hall, & Cuddy, 2013) showing significant overlap. Social vision research has indicated that such overlap produces unique effects on social categorization, such that Asian men and Black women are categorized less efficiently because of contrasting racial and gender stereotype roles (e.g., the category Black is associated with masculinity, which is at odds with the category women, Carpinella, et al., 2015; Johnson, Freeman, et al., 2012). Taking this a step further, intersectional invisibility argues that the default for Asian is a woman and the default for Black is a man, with findings from paradigms like Who-Said-What supporting these categorical defaults through the pattern of errors that emerge (Sesko & Biernat, 2010; Schug, Alt, & Klauer, 2015). While the intersection of social categories leads to an exponential increase in the degree of complexity, one potential future study could examine whether Black women and Asian men may be less influential on questions related to gender numeric estimations compared to groups composed of Black men and Asian women. If true, this finding would give greater credence to intersectional invisibility when it comes to group percepts. A further extension may be examining the mean or average percept of a group's gender typicality. To a purely rational observer, groups composed of equal numbers of Black men and women (6 Black women:6 Black men) should have equal gender mean summations as groups composed of Asian men and Asian women (6 Asian men: 6 Asian women), however the tethering of race and gender suggest that this would not be the case, with the Asian group skewed more feminine and the Black group skewed more masculine. Overall, the expansion of social categories within group percepts results in a multitude of novel extensions to develop both ensemble perception as well as social vision theory.

A theme of my dissertation is the importance of visual cues to social categories in addition to the social categories themselves. As shown in Studies 4-6 visual cues to gender typicality were discernable within groups and impacted evaluations of the group as a whole. One omission from the present work is the role of gender typicality for faces of women. While I hypothesize that gender typicality should operate in a similar fashion (e.g., groups composed of masculine women will be perceived as having fewer women and be rated as more threatening, compared to groups composed of feminine women), it is also likely the case that these effects will not be as strong as when gender and gender typicality match (e.g., a group of masculine women should not lead to even higher ratings of threat compared to a group of masculine men). Even more complex are the possible iterations of groups composed of men and women who vary in gender typicality. For instance, what would be the effect of a group of masculine men and feminine women compared to a group of feminine men and feminine women, or feminine men and masculine women on social evaluations and social categorical outcomes. These combinations lead to interesting divergent and interactive hypotheses between ensemble perception research and social cognition. On one hand, ensemble perception would hypothesize that on a continuum of summary gender representations (such as the gender typicality face scale used in Study 4), masculine men/masculine women and feminine men/feminine women groups would anchor extreme group masculinity and extreme group femininity, respectively. It is less clear however where groups of masculine men/feminine women and feminine men/masculine women would fall along such a continuum. While a purist would suggest they should fall at roughly the same spot, social psychological theory would suggest that gender atypical men and women are not weighted equally (Glick, Gangl, Gibb, Klumpner, & Weinberg, 2007) and as such, differences in outcomes may emerge. Similar examination of visual cues to racial phenotypically, as well as the intersection of gender and race open a host of questions regarding the impact of visual cues on group perception. Overall, it is no longer the case that group differences can be thought of in terms of monolithic category differences, and instead within-group variability is a critical element to examine in people construal.

Continued Integration with the Field of Ensemble Perception. The field of ensemble perception is still relatively nascent with novel findings emerging about both low-level and more socially relevant group stimuli. Of particular interest is recent work showing that perceivers who viewed serially presented photos, that is in sequence one after the other rather than as a single group, reported accurate summary representations about the "group" as a whole (Hubert-Wallander & Boynton, 2015). This slight change in how ensemble perception was tested has ramifications that parallel technological advances of today, such as "new feeds" viewed on a cellphone or profile pictures viewed on common dating apps. This sequential ensemble perception suggests that even when people are presented one after another, a group percept may still be available to perceivers. From a social psychological perspective this has strong implications for how a prototype about a certain group of people might develop across time and experience (see Tong, Dubé, & Sekuler, 2019).

The integration of ensemble perception and person construal theories also suggests possibilities with how group percepts are continually updated. Indeed, a critical, yet understudied, element of social vision models such as the Dynamic Interactive Theory of Person Construal (Freeman & Ambady, 2010) is that the social perception of others occurs in an iterative fashion and updates throughout the temporal sequence (e.g., the processing of fleeting facial emotions). While socially perceived characteristics about groups certainly change over time (e.g., the age of a group of friends), groups themselves can also change, with new members added and old members dropping out. We likely do not hold a static, un-updatable representation of groups; however little research has examined how such updating might occur for people construal and to what effect. This offers a unique window into how both ensemble perception updates (Whitney & Leib, 2018) and what the resultant effects will be for how we perceive and evaluate a group.

Implications of People Construal. The act of consciously thinking about visually perceived groups entails interpreting or making some meaning out of that visual perception (Bruner, 1957; McArthur & Baron, 1983). In my dissertation studies I presented evidence that our perception of a groups gender composition serves as a means of evaluating that group's level of threat or their capacity to do harm. In other work, my collaborators and I showed that perceiving a group's gender ratio interacted with one's own gender identity to impact evaluations of fit and belonging within the group (Goodale, Alt, Lick, & Johnson, 2018). A similar finding would likely emerge for differences in other social categories such as race, with implications for underrepresented minorities in classrooms and college campuses. The integration of people construal with other established research domains such as stereotype threat (Steele & Aronson, 1995) would be fruitful for developing a holistic theory of how minorities perceptual experience of others and their environment translates into negative outcomes (see Murphy, Steele, & Gross, 2007). As research into people construal continues it will be important to remain focused on the social implications and potential for interventions surrounding group perception.

The second half of my dissertation largely focused on the impact of bottom-up visual cues (i.e., gender typicality) and their independent and interactive effects with social categories to affect group evaluations and group categorization processes. It is important to recognize that the Dynamic Interactive Theory of Person Construal (Freeman & Ambady, 2011) posits that top-down cognitive and motivational factors also affect how we perceive individuals, and it is likely the case that these factors also play a role in our perception and social judgments about people. One extension building towards this theoretical point would be to examine how individual differences in attitudes about and motivations towards women and racial minorities would impact perceptual estimations and group evaluations. In fact, aligning with recent work by Cooley and Payne (2017, 2018), it may be the case that individual differences in attitudes held about a group (e.g., global attitudes about Black Americans) would be even more predictive of attitudes about a perceived group of Black men compared to an individual Black man as the group percept may serve as a means of removing the "individuating" processes potentially applicable to the social perception of an individual (Fiske & Neuberg, 1990). This could have strong implications for understanding what the effects would be of depicting an individual versus a group, and to what end these different portrayals might serve.

Other top-down factors of interest include the role of knowledge surrounding the context in which a group is perceived. As Freeman, Ma, Han and Ambady (2013) demonstrated, the background image (e.g., stereotypically American or Chinese) affected social categorization decisions of Asian and White faces as assessed via mouse-tracking. They found that an American background led to less direct categorizations of Asian faces and a Chinese background led to less direct categorizations of a White face. Applied to groups, I hypothesize that contextual factors such as knowing whether a classroom is associated with STEM versus non-STEM may bias the estimated number of women or racial minorities within the class, ultimately affecting beliefs about representation and even performance within such settings (Murphy et al., 2007). On-going work is also examining how identity and attitudinal differences may impact numeric estimations within a group, with results showing a tendency by White Americans to over-estimate the number of Black men, compared to White men, in briefly presented, large 100 person groups and for White Americans to be relatively poor at gauging whether their in-group (White individuals) are in the racial majority or minority when viewing groups composed of 70%

racial minorities, 30% Whites (Alt, Chaney, & Johnson, *in prep;* Alt, Does,& Shih, *in prep*). In sum, this preliminary evidence suggests that people construal processes likely mirror person construal processes regarding the impact of top-down factors on group perception.

Conclusion

In summary, the studies presented in this dissertation integrate novel findings from vision research on ensemble perception with social psychological theory on person construal to develop and explore our understanding of how we visually perceive and draw inferences about groups of people. In developing a theory of people construal, I investigated and found evidence for three critical aspects; first, that perceivers were sensitive to visually depicted group gender ratio and that judgments of groups tracked gender stereotype associations (i.e., the more a group shifted from women to men, the higher the threat evaluations of the group). Second, I showed that perceivers were sensitive to both category differences (i.e., men and women) and within-category variability (i.e., gender typicality) as assessed by numeric estimations, average gender typicality, and trait evaluations. And third, I found that social category facilitation occurred following group primes when gender and gender typicality aligned, leading to faster reaction times to categorize a subsequent face as a man. Furthermore, tracking perceivers categorical decisions as they unfolded over time revealed that visual cue and social category alignment produced fewer errors, and faster, more direct, group categorizations. These studies all build new knowledge around a critical yet under-investigated visual percept—groups—showcasing points of convergence and divergence with past models of person perception (Fiske & Neuberg, 1990; Freeman & Ambady, 2011; Hamilton, 1981; Macrae & Bodenhausen, 2000) and ensemble perception (Whitney & Leib, 2018) to advance theory and research on *people* construal.

References

Adams, R. B., Ambady, N., Nakayma, K., & Shimojo, S. (Ed.). (2011). *The science of social vision*. New York: Oxford University Press.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810–815. https://doi.org/10.3758/s13414-012-0293-0

Allport, G. W. (1954). *The Nature of Prejudice* (Unabridged, 25th anniversary ed). Reading, Mass: Addison-Wesley Pub. Co.

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

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

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

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology, 4*, 291–322. doi:10.1037/1089-2680.8.4.291

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

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258-290.

Bai, Y., Leib, A. Y., Puri, A. M., Whitney, D., & Peng, K. (2015). Gender differences in crowd perception. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01300

Balcetis, E., & Lassiter, D. G. (Ed.). (2010). *Social psychology of visual perception*. New York, NY: Psychology Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674–679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

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

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: testing process models of stereotype use. *Journal of Personality and Social Psychology*, *55*(5), 726–737.

Boucherie, A., Jørkov, M. L., & Smith, M. (2017). Wounded to the bone: Digital microscopic analysis of traumas in a medieval mass grave assemblage (Sandbjerget, Denmark, AD 1300–1350). *International Journal of Paleopathology*, *19*, 66–79. https://doi.org/10.1016/j.ijpp.2017.10.005

Borchert, K. (2016). User manual for Inquisit's AMP procedure. Retrieved from http://www.millisecond.com/download/library/v5/AMP/amp.manual

Boyer, P., Firat, R., & Leeuwen, F. V. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science*, *10*(4), 434–450. doi:10.1177/1745691615583133

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brand, J., Oriet, C., & Sykes Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(1), 63–69. https://doi.org/10.1037/a0026950

Brewer, M. B. (1988). A dual process model of impression formation. In *A dual process model of impression formation.* (pp. 1–36). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Brewer, M. B., & Feinstein, A. S. H. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 255-270). New York: Guilford Press.

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

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3,* 14-25.

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

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, *41*(3), 405–419. http://doi.org/10.1177/0146167214567153

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

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

Cooley, E., & Payne, B. K. (2017). Using groups to measure intergroup prejudice. *Personality and Social Psychology Bulletin*, *43*(1), 46–59. doi:10.1177/0146167216675331

Cooley, E., & Payne, B. K. (2018). A group is more than the average of its parts: Why existing stereotypes are applied more to the same individual when viewed in groups than when viewed alone. *Group Processes and Intergroup Relations,* 1-15.

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, P., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32,* 917–929. doi:10.1177/0146167206287721

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*(5), 991–1004. http://doi.org/10.1037/0022-3514.46.5.991

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology (2006)*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. doi:10.1037/0022-3514.56.1.5

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on “automatic evaluations” as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, *45*(1), 101–114. doi:10.1016/j.jesp.2008.09.001

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*(3), 189–198. https://doi.org/10.1093/scan/nsm012

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*(1), 22–37. https://doi.org/10.1016/0022-1031(86)90039-9

Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 309–330. https://doi.org/10.1037/0033-2909.100.3.309

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

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

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Elsevier. https://doi.org/10.1016/S0065-2601(08)60317-2

Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition*. McGraw-Hill.

Fiske, S. T., Lin, M. H., & Neuberg, S. L. (1999). The Continuum Model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231-254). New York, NY: Guilford.

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*, 32210. https://doi.org/10.1038/srep32210

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. doi:10.1037/a0022327

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology. General*, *137*(4), 673–690. https://doi.org/10.1037/a0013875

Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology, 49,* 206-210.

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46,* 179-185. doi:10.1016/j.jesp.2009.10.002

Friedman, H., & Zebrowitz, L. A. (1992). The contribution of typical sex differences in facial maturity to sex role stereotypes. *Personality and Social Psychology Bulletin*, *18*(4), 430–438. https://doi.org/10.1177/0146167292184006

Frieze, I. H., & Li, M. Y. (2010). Gender, aggression, and prosocial behavior. In *Handbook of Gender Research in Psychology* (pp. 311–335). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1467-5_13

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. C. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, *24*(4), 498–506. doi:10.1177/0956797612457783

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin*, *40*(1), 3–15. doi:10.1177/0146167213502548

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLOS ONE*, *10*(7), e0132726. https://doi.org/10.1371/journal.pone.0132726

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1526–1541. https://doi.org/10.1037/a0036732

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Psychology Press.

Glick, P., Gangl, C., Gibb, S., Klumpner, S., & Weinberg, E. (2007). Defensive reactions to masculinity threat: More negative affect toward effem- inate (but not masculine) gay men. *Sex Roles, 57,* 55–59. doi:10.1007/s11199-007–9195-3

Goldstein, J. S. (2003). *War and Gender: How Gender Shapes the War System and Vice Versa*. Cambridge University Press.

Goodale, B. M., Alt, N. P., Lick, D. J., & Johnson, K. L. (2018). Groups at a glance: Perceivers infer social belonging in a group based on perceptual summaries of sex ratio. *Journal of Experimental Psychology: General, 147*(11), 1660-1676.

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

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

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

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

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

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

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

Haberman, J., & Whitney, D. (2012). Ensemble perception. In J. Wolfe & L. Robertson (Eds.), *From Perception to Consciousness* (pp. 339–349). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576. https://doi.org/10.1111/j.1467-9280.2006.01746.x

Hamilton, D. L. (Ed.). (1981). Cognitive processes in stereotyping and intergroup behavior. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamilton, D. L., Sherman, S. J., & Lickel, B. (1998). Perceptions of groups: The importance of the entitativity continuum. In C. Sedikides, J. Schopler, & C. A. Insko (Eds.), Intergroup cognition and intergroup behavior (pp. 47–74). Mahwah, NJ: Lawrence Erlbaum Associates.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78,* 81-91

Heider, F. (1958). *The Psychology of Interpersonal Relations* (1st edition). John Wiley & Sons.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations*, *18*(3), 384–401. http://doi.org/10.1177/1368430214538325

Heine, S. J. (2010). Cultural psychology. In *Handbook of Social Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002037

Holbrook, C., & Fessler, D. M. T. (2013). Sizing up the threat: The envisioned physical formidability of terrorists tracks their leaders' failures and successes. *Cognition, 127*(1), 46–56. http://dx.doi.org/10.1016/j.cognition.2012.12.002.

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, *15*(4), 5. https://doi.org/10.1167/15.4.5

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

Ip, G. W., Chiu, C. Y., & Wan, C. (2006). Birds of a feather and birds flocking together: Physical versus behavioral cues may lead to trait- versus goal-based group perception. *Journal of Personality and Social Psychology*, *90*(3), 368–381. https://doi.org/10.1037/0022-3514.90.3.368

Ji, L., Chen, W., & Fu, X. (2014). Different roles of foveal and extrafoveal vision in Ensemble representation for facial expressions. In *Engineering Psychology and Cognitive Ergonomics* (pp. 164–173). Springer, Cham. https://doi.org/10.1007/978-3-319-07515-0_17

Johnson, K. L., & Adams, R. B. (2013). Social Vision: An Introduction. *Social Cognition*, *31*(6), 633–635. https://doi.org/10.1521/soco.2013.31.6.633a

Johnson, K. L., & Freeman, J. B. (2010). A “new look” at person construal: Seeing beyond dominance and discreteness. In E. Balcetis & D. Lassiter (Eds.) *The Social Psychology of Visual Perception* (pp. 253–272). New York: Psychology Press.

Johnson, K. L., Iida, M., & Tassinary, L. G. (2012). Person (mis)perception: Functional biased sex categorization of bodies. *Proceeding of the Royal Society B, 279,* 4982-4989.

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science*, *16*(11), 890–897. https://doi.org/10.1111/j.1467-9280.2005.01633.x

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent research in social vision: an integrative approach to the determinants and consequences of social categorization. *Social and Personality Psychology Compass,* 9(1), 15-30. http://doi.org/10.1111/spc3.12147

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*(1), 116–131. doi:10.1037/a0025335

Jung, W., Bülthoff, I., & Armann, R. G. M. (2017). The contribution of foveal and peripheral visual information to ensemble representation of face race. *Journal of Vision*, *17*(13), 11. https://doi.org/10.1167/17.13.11

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup Perception and Cognition: An Integrative Framework for Understanding the Causes and Consequences of Social Categorization. Advances in Experimental Social Psychology (1st ed., Vol. 55). Elsevier Inc. http://doi.org/10.1016/bs.aesp.2016.10.001

Kellermann, A. L., & Mercy, J. A. (1992). Men, women, and murder: Gender-specific differences in rates of fatal violence and victimization. *Journal of Trauma and Acute Care Surgery*, *33*(1), 1.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality and Social Psychology*, *75*(5), 1155–1178. https://doi.org/10.1037/0022-3514.75.5.1155

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, *21*(6), 575–580. https://doi.org/10.3758/BF03198740

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models, R package version 2.0-30. https://CRAN.R-project.org/package=lmerTest

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186. https://doi.org/10.1038/ncomms13186

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *Journal of Experimental Psychology. General*, *135*(4), 501–512. https://doi.org/10.1037/0096-3445.135.4.501

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, *8*(6), 720–722. https://doi.org/10.1038/nn1465

Lick, D. J., & Johnson, K. L. (2016). Perceptually Mediated Preferences and Prejudices. *Psychological Inquiry*, *27*(4), 335–340. https://doi.org/10.1080/1047840X.2016.1215211

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*(1), 5–18.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

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

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

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

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*(3), 397–407.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology*, *33*(5), 471–489. https://doi.org/10.1006/jesp.1997.1328

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: feature-based processing and person construal. *European Journal of Social Psychology*, *37*(5), 793–805. https://doi.org/10.1002/ejsp.406

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

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*(4), 383–401. https://doi.org/10.1207/s15327957pspr0804_4

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*(2), 250–259. https://doi.org/10.1177/0146167202282010

Maddox, K. B., & Perry, J. M. (2017). Racial appearance bias: Improving evidence-based policies to address racial disparities. *Policy Insights from the Behavioral and Brain Sciences, 5*(1), 57-65.

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

McDonald, M. M., Navarrete, C. D., & Vugt, M. V. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosphical transactions of the Roysal Society B, 367*, 670-679. doi:10.1098/rstb.2011.0301

McNair, N. A., Goodbourn, P. T., Shone, L. T., & Harris, I. M. (2017). Summary statistics in the attentional blink. *Attention, Perception, & Psychophysics*, *79*(1), 100–116. https://doi.org/10.3758/s13414-016-1216-2

Milord, J. T. (1978). Aesthetic aspects of faces: A (somewhat) phenomenological analysis using multidimensional scaling methods. *Journal of Personality and Social Psychology, 36,* 205-216.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633. https://doi.org/10.3758/BF03206543

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599-620. doi:10.1207/S15328007SEM0904_8

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879-885.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blackmore (Ed.), *Vision: Coding and efficiency* (pp. 411-422). Cambridge: Cambridge University Press

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human Threat Management Systems: Self-Protection and Disease Avoidance. *Neuroscience and Biobehavioral Reviews*, *35*(4), 1042–1051. https://doi.org/10.1016/j.neubiorev.2010.08.011

Pastore, A. L., & Maguire, K. (Eds.). (2005). *Sourcebook of Criminal Justice Statistics, 2003* (31st ed. edition). Washington, DC: Office of Justice Programs, Bureau of Justice.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*(1), 16-31. doi:10.1037/0022-3514.94.1.16

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*(5), 729–738.

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

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

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles, 59*(5–6), 377–391. http://dx.doi.org/10.1007/s11199-008-9424-4.

Rensink, R. A., O'Regan, K. J., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, *7*(1–3), 127–145. https://doi.org/10.1080/135062800394720

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, *67*(1), 439–463. http://doi.org/10.1146/annurev-psych-010213-115115

Schaller, M., & Neuberg, S. L. (2008). Intergroup prejudices and intergroup conflicts. In C. Crawford & D. L. Krebs (Eds.), Foundations of evolutionary psychology: Ideas, issues, and applications (pp. 399 – 412). Mahwah, NJ: Erlbaum.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259–290. https://doi.org/10.1006/cogp.1998.0698

Schug, J., Alt, N. P., & Klauer, K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology*, *56*, 121–125. doi:10.1016/j.jesp.2014.09.012

Sczesny, S., Spreemann, S., & Stahlberg, D. (2006). Masculine = competent? Physical appearance and sex as sources of gender-stereotypic attributions. *Swiss Journal of Psychology*, *65*(1), 15–23. https://doi.org/10.1024/1421-0185.65.1.15

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*(2), 356–360. http://doi.org/10.1016/j.jesp.2009.10.016

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 13–13. https://doi.org/10.1167/11.12.13

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*(2), 207–218. https://doi.org/10.1037/0022-3514.62.2.207

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

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool for investigating social perception. *Social Psychological and Personality Science, 8*(5), 538-551.

Swann, W. B., & Bosson, J. K. (2010). Self and identity. In *Handbook of Social Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy001016

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

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Tajfel, H. (1974). Social identity and intergroup behaviour. *Information (International Social Science Council)*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

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*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Thornton, I. M., Srismith, D., Oxner, M., & Hayward, W. G. (2014). Establishing a majority: Observer race influences estimates of crowd ethnicity. *I-Perception*, *5*(5), 490–490. https://doi.org/10.1068/ii59

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2013). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*, 1–38. doi: 10.18637/jss.v059.i05

Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, & Psychophysics.*

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network (Bristol, England)*, *14*(3), 391–412.

Třebický, V., Fialová, J., Kleisner, K., & Havlíček, J. (2016). Focal Length Affects Depicted Shape and Perception of Facial Images. *PLOS ONE*, *11*(2), e0149313. https://doi.org/10.1371/journal.pone.0149313

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

Vugt, M. V. (2009). Sex differences in intergroup competition, aggression, and warfare: The male warrior hypothesis. *Annual New York Academy of Science, 1167,* 124-134. doi: 10.1111/j.1749-6632.2009.04539.x

Warr, P., & Knapper, C. K. (1968). *Perception of People and Events*. London, New York etc.: John Wiley & Sons Ltd.

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

Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A source book of Gestalt psychology.* (pp. 71–88). London, England: Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56–63. https://doi.org/10.1111/j.0956-7976.2005.00780.x

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

Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology, 77,* 269-291. https://doi.org/10.1006/jescp.1999.2554.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*(1), 55–62.

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. http://dx.doi:10.1037/pspi0000092

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4), 11. https://doi.org/10.1167/15.4.11

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. *Psychological Bulletin*, *128*(5), 699–727.

Yzerbyt, V., & Demoulin, S. (2010). Intergroup relations. In *Handbook of Social Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002028

Ying, H., & Xu, H. (2017). Adaptation reveals that facial expression averaging occurs during rapid serial presentation. *Journal of Vision*, *17*(1), 15. https://doi.org/10.1167/17.1.15

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*(5), 657–701. doi:10.1521/soco.2006.24.5.657

Zhang, Z., Zheng, C., Kim, C., Van Poucke, S., Lin, S., & Lan, P. (2016). Causal mediation analysis in the context of clinical research. *Annals of Translational Medicine*, *4*(21). doi:10.21037/atm.2016.11.11

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, *41*(4), 322–330. https://doi.org/10.1002/ab.21544

Zosh, J. M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, *140*(2), 141–158. https://doi.org/10.1037/a0022925