Categorization of Gender: Race, Stereotypes and Androgyny

Zoe Vogeli

Department of Psychology, McGill University

PSYC 380D1/D2: Honours Research Project Seminar

Dr. Jordan Axt

Dr. David Ostry

April 22, 2021

Abstract

Through a gender categorization task in which participants select a target face as male or female, Johnson et al (2012) revealed that race is gendered – certain combinations of race and gender lead to facilitated categorization. These results indicate that race and gender are psychologically confounded –the concept "Black" was stereotype congruent with male, and the concept of "Asian" was stereotype congruent with female – resulting in facilitated categorization of Black male faces and Asian female faces (Johnson et al., 2012). There is yet to be research that assesses how this stereotype congruency moderate's categorization of gender *at varying levels of androgyny* – that is, among faces varying in the degree to which they are clearly male or female. If "race is gendered", what effect does race have when gender categorization is difficult (i.e., when using more androgynous targets)? This project examined how categorization of target gender at varying levels of androgyny is moderated by whether the face is "stereotype congruent" (i.e., Black faces that are predominantly male or Asian faces that are predominantly female) or "stereotype incongruent" (i.e., Black faces that are predominantly female or Asian faces that are predominantly male). Using a mouse-tracking gender categorization task, we measured participants decision conflict (via latency, x-flips and area under the curve) and found a significant interaction between stereotype congruency and androgyny, such that stereotype congruency decreased decision conflict in gender categorization for more androgynous faces. We discuss how cognitive load perspectives may shed light on the results found in the present study.
Keywords: androgynous, stereotyping, race, gender, decision conflict, mouse tracking

Categorization of Gender: Race, Stereotypes and Androgyny

Every day as humans we observe the people in our environment without a second thought. This social perception occurs with minimal effort and visual information. When we see a face, we immediately work to categorize that face along social dimensions, such as by gender or race. For instance, research using event-related brain potentials (ERPs; Ito & Uland, 2003) investigated the extent to which race and gender are automatically encoded. Participants were shown an image of a face for 1000ms, and then had to categorize the face as either male or female in the gender condition, or Black or White in the race condition. Race information was processed as early as 122ms after stimuli was presented, and gender information was processed only slightly after (Ito & Urland, 2003). This provides clear support that social dimensions such as gender and race are perceived automatically, which aligns with previous research using more behavioral measures (e.g., Brewer, 1988; Fiske & Neuberg, 1990; Stroessner, 1996). However, recent studies have indicated that instead of being treated as two distinct categories in perception, race and gender are correlated in perception, meaning that the process of identifying someone's gender is influenced by their race (Johnson et al., 2012). That is, the existence of shared phenotypes and stereotypes between race and gender promote or inhibit one's ability to categorize a face as either male or female (Johnson et al., 2012).

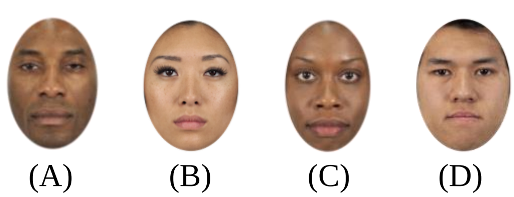
Race is Gendered

In Johnson and colleagues' (2012) study, participants categorized stimuli that depicted gender-ambiguous faces that varied in a level of apparent race (Black, White, Asian), with results finding that as the race of the stimuli changed from Black to White to Asian, the probability of categorizing the face as female increased significantly (Johnson et al., 2012), such that the most phenotypically Asian faces were more than twice as likely to be categorized as women relative to the most phenotypically Black faces. Johnson and colleagues (2012) then used a mouse tracking task to record participants mouse trajectories in order to measure decision conflict throughout the categorization process (Freeman & Ambady, 2010). Again, they found that race biased gender categorization; for female stimuli the categorization of Black faces was impaired, and the categorization of Asian faces was facilitated, and the opposite trend appeared for male stimuli (Johnson et al., 2012). They proposed that the racially biased gender categorization can be attributed to common stereotypes between the race and gender (Johnson et al., 2012). More specifically, that there are shared stereotypes between Asian race and female gender, and Black race and male gender, which led to facilitated categorization of female Asian faces and Black male faces (Johnson et al., 2012).

In our study, we describe the shared stereotypes between race and gender in terms of stereotype congruency. Based on Johnson and colleagues' (2012) study, we describe the concept "Black" as "stereotype congruent" with male, and the concept of "Asian" as "stereotype congruent" with female. Likewise, the concept of "Black" can be described as stereotype incongruent with female, and the concept of "Asian" can be described as stereotype incongruent with male. Therefore, the stereotype congruent faces (i.e., Black male faces or Asian female faces) should elicit less decision conflict compared to "stereotype incongruent" faces (i.e., Black female faces of Asian male faces; Johnson et al., 2012). Examples of stereotype congruent and stereotype incongruent stimuli are presented in Figure 1.

Figure 1

Stereotype Congruent and Stereotype Incongruent Face Stimuli



Note. (A) Black male face, (B) Asian female face, (C) Black female face, (D) Asian male face. (A) and (B) are stereotype congruent. (C) and (D) are stereotype incongruent.

Androgyny

We introduce the role of androgyny given how previous work (e.g., Johnson et al., 2012) in this research topic relied on faces that were relatively unambiguous in terms of gender when using the mouse-tracker paradigm (i.e., all of the faces used in the study were selected to be clearly male or female). Thus, it is less clear how race influences gender categorization when gender is less easily identifiable.

Prior work has argued that masculinity and femininity are empirically and logically distinct and, therefore, they can be treated as two independent dimensions (Bem, 1974). This allows for characterization of a person as masculine, feminine, or androgynous based on where their traits lie on the masculine and feminine dimensions (Bem, 1974). Bem (1974) also found results which indicated that androgyny is a reliable psychological construct. Here, androgyny can be defined as the degree to which a face is clearly male or female. Androgyny is characterized by a combination of feminine and masculine traits which creates gender ambiguity (Nowak & Denis, 2016). Androgyny can arise when one has both masculine and feminine features, an absence of gendered features, or when one's features lie somewhere between masculine and feminine (Nowak & Denis, 2016). When a person's femininity or masculinity does not overpower the other, it leads to the inability to identify the person as either male or female. Although feminine and masculine traits are on a spectrum, research has shown - through face categorization tasks - that people perceive gender categorically even when evaluating androgynous faces (Campanella et al., 2001).

Cognitive Load and Androgyny

Gender categorization of androgynous faces may elicit higher cognitive load compared to non-androgynous faces. Cognitive load can be defined as a multidimensional construct that represents the load imposed on the cognitive system by performing a particular task (Paas & van Merriënboer, 1994). Cognitive Load Theory was first developed by Sweller (1988) in the context of problem-solving and learning. Since then, cognitive load theory has been applied to a variety of areas (Chandler & Sweller, 1992; Kirschner et al., 2009; Moreno & Mayer, 1999; Vrij et al., 2008). Recent research has studied the effects of cognitive load on decision making (e.g., Burgess, 2009; Deck & Jahedi, 2015; Tinghög), and stereotypes (e.g., Biernat et al., 2003; Burgess, 2009; Dijksterhuis & Van Knippenberg, 1995; Gilbert & Hixon, 1991; Macrae et al., 1993; Stangor & Duan, 1991; Wigboldus et al., 2004).

Evidence suggests that gender is categorized automatically and effortlessly (Brewer, 1988; Fiske & Neuberg, 1990; Ito & Urland, 2003; Stroessner, 1996), meaning the process should typically bring low cognitive load. However, having to categorize an androgynous face as male or female is uncommon, novel and presumably requires greater mental effort. Novel tasks are typically associated with high cognitive load (Paas & van Merriënboer, 1994). Thus, when a face is gender ambiguous (due to androgyny), the task of gender categorization may impose greater cognitive load on the participant. To summarize, gender categorization of non-androgynous faces may elicit low cognitive load, whereas gender categorization of androgynous faces may elicit high cognitive load.

Mouse-Tracking and Decision Conflict

To measure decision conflict in gender categorization, we used a mouse-tracking paradigm. Johnson et al. (2012) were some of the first to use the mouse-tracker paradigm, which has since become one of the most popular methods in research on social categorization. A mouse-tracking task is completed on a computer, where the participant is presented with a stimulus in the centre of the screen and then must click the appropriate button at the top left or right of the screen in order to select which category the stimuli belongs to. Mouse-tracking is most often used to measure real-time decision conflict of participants when choosing between two categories (Stillman et al., 2018). The area under the curve (AUC), number of x-flips, and latency (outcomes discussed in more detail below) are all mouse-tracking measures that cumulatively represent the amount of decision conflict present for participants (Stillman et al., 2018). Mouse tracking methods have been used to assess how participants make decisions and how the decision conflict is related to real world behavior such as risk preferences (Stillman et al., 2020), or discriminatory behavior (Melnikoff et al., 2020)

More specifically, area under the curve (AUC) is the geometric area between the participants mouse-trajectory and an idealized straight-line trajectory (Hehman et al., 2015). It is assumed that if there were no decision conflict - and the categorization choice were clear - a participant would use their cursor to move in a straight line and click the correct category button. And, conversely, if there were high decision conflict – and categorizing the stimuli was difficult – the participant may not move their cursor directly to one category but instead the experience of categorization uncertainty would result in a curved mouse trajectory that leans more towards the category in one corner, deviating from the ideal straight-line trajectory.

X-flips is the number of times the participant's cursor reverses direction along the horizontal x-axis. This outcome also represents the participant's uncertainty, as participants are presumably changing their decision between one category and the other when they change the direction of the cursor heading towards the category on the left versus the one on the right. A stimulus which elicits no decision conflict should result in no x-flips, as the participant would move the cursor in only one direction, towards one category, but a greater number of x-flips is associated with more decision conflict (Freeman, 2018).

Finally, latency is the amount of time participants take between being presented the stimulus and making their categorization decision. The longer it takes the participant to make the decision, it is assumed that there is more decision conflict, as an easy decision is made quickly with more efficiency (Johnson et al., 2012).

In our study, we used a mouse tracking paradigm in order to measure decision conflict when categorizing a face as either male or female. Face stimuli were either Black or Asian faces of varying levels of androgyny. This enabled us to analyse the effect stereotype congruency has on categorizing androgynous faces where stereotype congruent is a male Black face or an Asian female face (Johnson et al., 2012).

Overview

This study examined how categorization of target gender at varying levels of androgyny is moderated by whether the face is "stereotype congruent" (i.e., Black faces that are predominantly male or Asian faces that are predominantly female) or "stereotype incongruent" (i.e., Black faces that are predominantly female or Asian faces that are predominantly male; Johnson et al., 2012). Using a mouse-tracking paradigm, this project demonstrated that racial stereotypes can facilitate or impair gender categorization at varying levels of androgynous faces. In this study we investigated how shared stereotypes between race and gender (stereotype congruency) affects the categorization of a face as male or female at varying levels of androgyny. This work then explores whether participants rely on racial cues more when gender categorization is difficult and may impose high cognitive load (i.e., when using more androgynous targets), or whether the use of race in gender categorization is consistent among targets that are either more or less difficult to categorize.

Methods

Participants

We collected 1885 eligible participants for greater than 95% power to detect a correlation as small as r = .20. Participants were volunteers who came from the Project Implicit research pool (implicit.harvard.edu; Nosek, 2005). Project Implicit is an online research laboratory where volunteer participants can complete studies related to implicit social cognition. Our study was approved by the McGill REB and participants provided informed consent.

The average age of participants was 35 years old and ranged from 17 to 80. 66.5% of participants were female, 31.7% were male, and the remaining 1.8% chose not to disclose. 66.6% of participants were citizens of the United States of America, 8.9% were citizens of the United Kingdom, 5.7% were citizens of Canada, 1.8% were citizens of Australia, 1.3% were citizens of India. The race of the participants, as self-reported, was 68.8% White, 8.9% Black or African American, 6.6% was more than one race, 4.1% East Asian, 3.7% South Asian, 0.6% Native Hawaiin or other Pacific Islander, 0.5% American Indian/Alaska Native, and 6.8% was other/unknown.

For exploratory research, participants completed an Implicit Association Test (IAT; Greenwald et al., 1998). The IAT is a categorization task that is commonly used to assess implicit attitudes. Here, participants completed a novel IAT seeking to assess implicit attitudes towards androgynous vs. gender-conforming people. Though IAT data was not included in the present analysis, participants were excluded from mouse-tracking analyses if more than 10% of trials in the IAT were faster than 300ms, an indication of careless responding (Greenwald et al.,

2003) . In addition, trials in which participants took excessively long to respond on the mouse-tracking task (>2000 ms; +3SD from the average response time) were excluded (Freeman et al., 2016). Finally, we also removed trials with no mouse tracking data after the 30ms mark in order to complete analyses.

Procedure

All participants completed a gender categorization mouse-tracking task. Participants were given written instruction, a four-trial training block of the task for practice, before going on to complete 120 trials where they had to categorize a face as male or female.

Mouse-Tracking Task

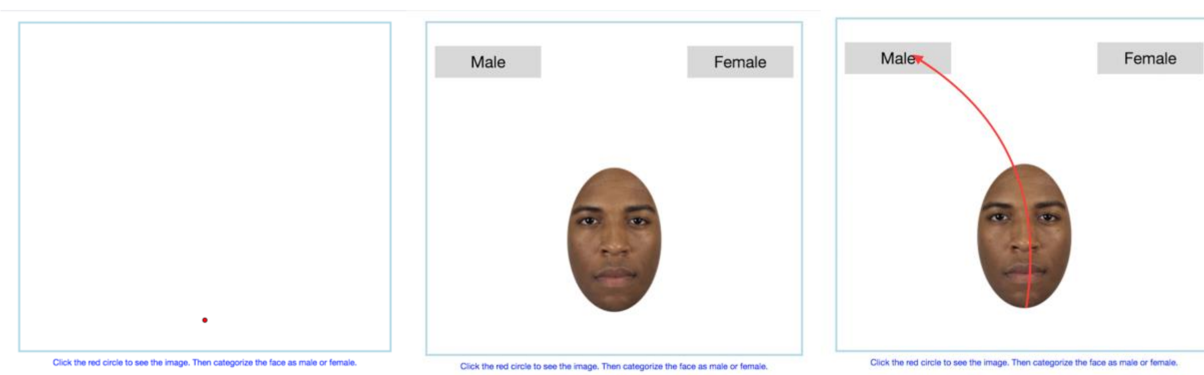
Participants completed a 120-trial mouse-tracking gender categorization task on the computer. Participants received instruction prior to beginning the task. They were informed that they would see an image of a person's face at the bottom of their screen, and for each image they were to categorize whether they think the face is male or female by dragging their mouse to the appropriate button at the top of the screen. They were also told to categorize the face as quickly as possible, and that it was okay if they made an occasional mistake.

The task runs as follows: The task begins with a small red circle at the bottom-centre of the screen, and two rectangular buttons in the top left and right corners of the screen labeled as "male" or "female" (Figure 2A). At the bottom of the screen, below the red circle, there are instructions that state "Click the red circle to see the image. Then categorize the face as male or female." When the participant clicks the red circle with their cursor, an image of a face immediately appears – just above the red circle - bottom centre of the screen (Figure 2B). The image of the face is one of 120 face stimuli (as described below) presented in a random order. The participant then must choose to categorize the face as either male or female by moving the mouse cursor from the automatically pre-set location at bottom-center of the screen (the red circle) to the ‘‘male’’ or ‘‘female’’ response boxes in either top corners of the screen (Figure 2C). Once the participant selects either “male” or “female” with their cursor, the face stimuli disappears, and the participant must click the red circle to continue onto the next trial where they will categorize a new face. Participants were shown a warning to “move faster” if their mouse fails to move for more than 400ms during any portion of the trial. Participants complete a four-trial training task (with two face stimuli), using the method described above.

After the training task, participants complete 120-trials. This is to ensure the participants are familiar with the task before measurement begins. After 60 trials, the “male” and “female” labeled buttons swap sides so that they are now in opposite top corners for the next 60 trials.

Figure 2

Gender Categorization Mouse-tracking Task



(A) (B) (C)

Note. (A) At the start of each trial, participants must click the red dot to begin, (B) Once clicked, a target face will immediately appear, (C) The participant will move their mouse cursor to select either male or female in order to categorize the gender of the target face. The red arrow represents the participants mouse trajectory, which is measured, but not visible to the participant.

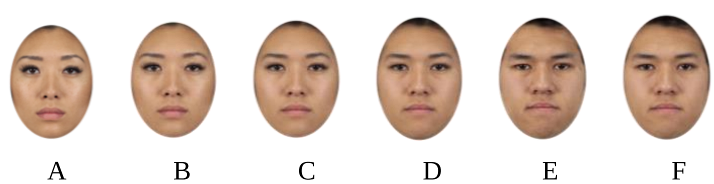
To familiarize the participants with the mouse-tracking measure, participants first complete a four-trial training block (using the same method as described above) in which they sorted two faces (not present in the following trials), prior to beginning the critical 120 trials. In the critical trials, the image of the face presented after clicking the red circle is one of 120 face stimuli (as described below) presented in a random order.

The face stimuli were either Black or East Asian and varied in their level of androgyny. Face stimuli for this task were created by morphing a subset of 10 of the most masculine and most feminine faces across two races (East Asian, and Black) currently represented in the Chicago Face Database (Ma et al., 2015). Pairs of male and female faces were generated randomly such that each face was used exactly once. Each pair was morphed twice at a 20% gradient to create 4 morphs at the levels of 20% Male / 80% Female, 40% Male / 60% Female, 60% Male / 40% Female, and 80% Male / 20% Female (Figure 3 below). In total, there were 6 photos (2 originals and 4 morphs) for each of 10 base pairs across the 2 races, which yielded 120 faces total.

After the mouse-tracking task, participants reported whether they used a physical mouse, a trackpad, or a touchscreen.

Figure 3

Face Stimuli, Original and Morphs



Note. (A) Original Female face, (B) 20% Male / 80% Female morph, (C) 40% Male / 60% Female morph, (D) 60% Male / 40% Female, (E) 80% Male / 20% Female, (F) Original Male face

The goal of the mouse-tracking task was to quantify decisional conflict in real time by continuously monitoring mouse movement while the participant categorizes each face stimuli as male or female. For each trial, participants' cursor position on the x-y plane was recorded every

~15ms to track the mouse movement during decision making. Based on existing recommendations on how to effectively use mouse tracking in the context of social cognition (e.g., Stillman et al., 2018), we used the measured mouse movement data to calculate three specific characteristics; latency, trajectory (AUC), and uncertainty (x-flips). These characteristics are assumed to be a good indication for the level of decision conflict experienced in a categorization task (Stillman et al., 2018).

We measured latency by calculating the amount of time (in milliseconds) it took the participant to categorize each stimulus item (e.g., from the time the mouse is released to the time the mouse tracking software has registered the stimulus as sorted) with the assumption that greater latency reflects more decisional conflict and less efficiency of judgements (Johnson et al., 2012). Additionally, we measured trajectory by calculating the area under the curve (or AUC) for each stimulus. AUC is measured by comparing the actual path of the mouse to an idealized straight trajectory from the starting location to response termination, and it is assumed that the more the mouse path deviates from the straight trajectory, the greater decisional conflict between the two options (Stillman et al., 2018). Finally, we measured uncertainty using x-flips, which is the number of times the mouse reverses direction in the x-plane while categorizing a stimulus. Here it is assumed that a greater number of reversals represent greater levels of uncertainty and thus more decisional conflict (Freeman, 2018).

Results

We analyzed the mouse-tracking data to see the effect that androgyny, stereotype congruency, and their interaction had on the three indices of decision conflict (Stillman et al., 2018): area under the curve (AUC), x-flips, and overall latency (Stillman et al., 2018).

Hierarchical linear modeling (HLM) was used to analyze the data from the mouse tracking task where trials (Level 1) were nested within participant (Level 2). Each trial-level outcome – latency, AUC, x-flips – was predicted by target androgyny level, target stereotype congruency and their interaction

Androgyny level was operationalized by giving a score to each face that represented the amount of androgyny present. The score was determined with the following formula: 100 – |(% male - % female)|, with higher values indicating the face had greater androgyny. This formula meant that 100% male or 100% female faces would have an androgyny value of 0, 20% male or 20% female faces would have a value of 40, and 60% male or 60% female faces would have a value of 80.

Stereotype congruency was coded for each face/image such that a value of one (1) indicated the target face was stereotype congruent and a value of zero (0) indicated that the face was stereotype incongruent. Specifically, targets were classified as stereotype congruent when they were either Black and predominantly male (i.e., having ≥60% of the face come from a male image) or Asian and predominantly female (i.e., having ≥60% of the face come from a female image). The remaining faces were classified as stereotype incongruent, being either Black and predominantly female or Asian and predominantly male.

For the latency analysis, the coefficient for androgyny level was positive and statistically significant (b = 1.11, SE = 0.025, t = 44.32, p < .001) indicating that the more androgynous the face, the higher the response latency (more decision conflict). The stereotype congruency coefficient was negative and statistically significant (b = -15.00, SE = 1.68, t = -8.89, p < .001), indicating that when a face was stereotype congruent the response latency was lower (less decision conflict). As shown in Figure 4 (below), the interaction of androgyny level and stereotype congruency was also negative and significant (b = -0.18, SE = 0.036, t = -5.04, p < .001), indicating that the more androgynous the face, the more effect stereotype incongruency had on latency; that is, a face high in androgyny elicited more decision-conflict when the target was stereotype incongruent (Asian male or Black female) compared to stereotype congruent (Asian female or Black male).

The analyses for AUC and X-flips revealed a similar pattern. For AUC, the coefficient for androgyny level was positive and statistically significant (b = 0.002, SE = 0.0001, t = 16.84, p < .001) indicating that the more androgynous the face, the more area under the curve (more decision conflict). The stereotype congruency coefficient was negative and statistically significant (b = -0.069, SE = 0.007, t = -9.35, p < .001), indicating that when a face was stereotype congruent there was less area under the curve (less decision conflict). As shown in Figure 5 (below), the interaction of androgyny level and stereotype congruency was negative and significant (b = -0.001, SE = 0.0002, t = -7.96, p < .001), indicating that the more androgynous the face, the more effect stereotype incongruency had on area under the curve.

Finally, analyses for X-flips showed that the coefficient for androgyny level was positive and statistically significant (b = 0.003, SE = 0.0001, t = 21.09, p < .001) indicating that the more androgynous the face, the more x-flips (more decision conflict). The stereotype congruency coefficient was negative and statistically significant (b = -0.065, SE = 0.008, t = -7.75, p < .001), indicating that when a face was stereotype congruent the fewer number of x-flips (less decision conflict). As shown in Figure 6 (below), the interaction of androgyny level and stereotype congruency was also negative and significant (b = -0.0006, SE = 0.0002, t = -3.28, p < .001), indicating that the more androgynous the face, the more effect stereotype incongruency had on number of x-flips.

Figure 4

Interaction of Stereotype Congruency and Androgyny Level on Latency

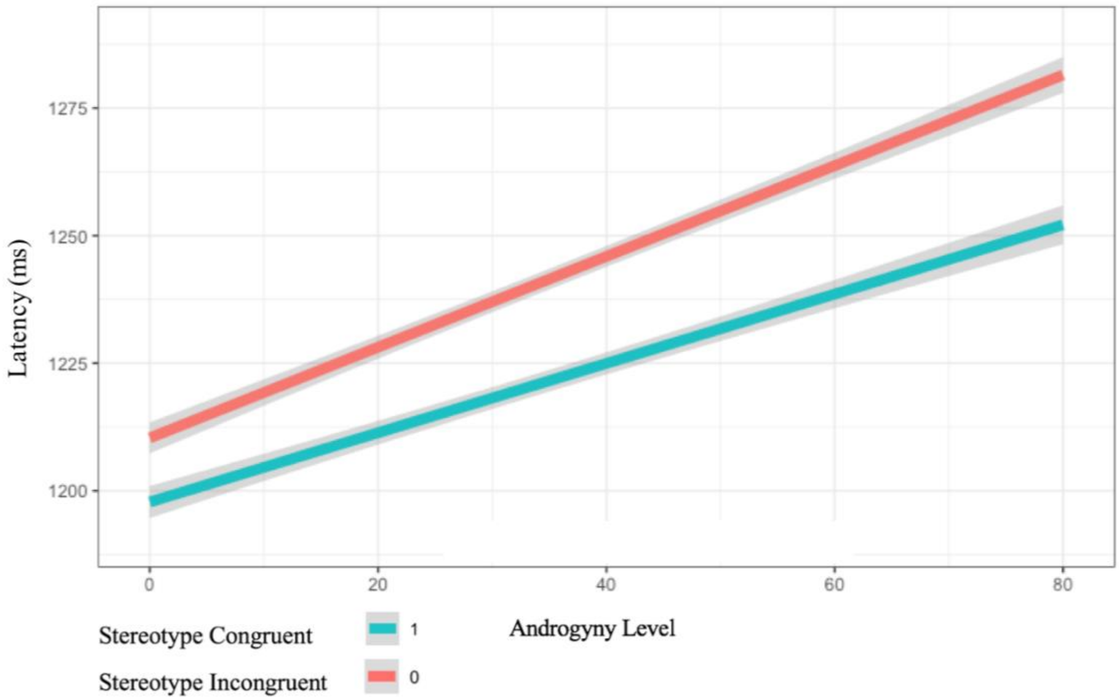


Figure 5

Interaction of Stereotype Congruency and Androgyny Level on AUC

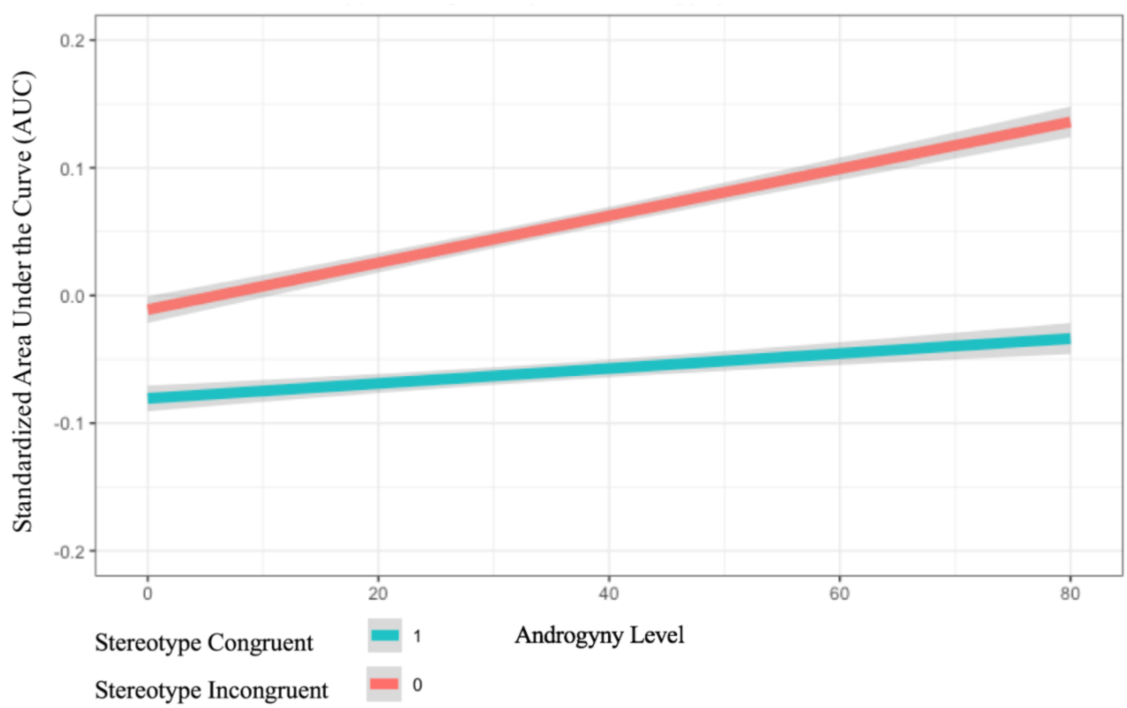
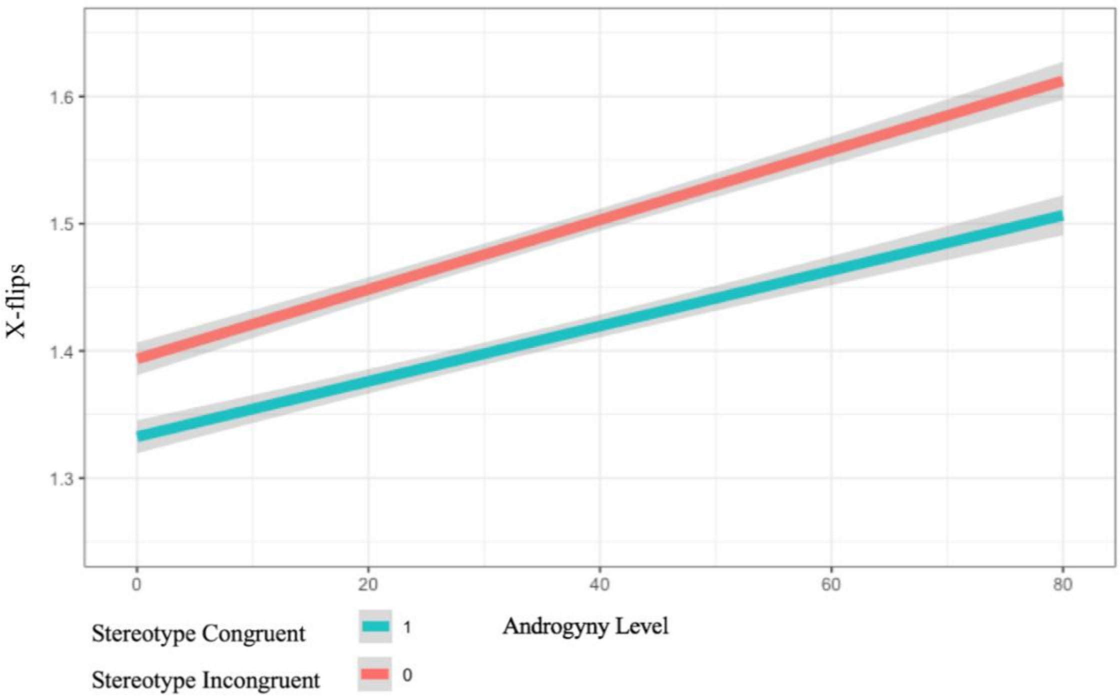


Figure 6

Interaction of Stereotype Congruency and Androgyny Level on X-flips



Discussion

Our study aimed to determine the effect stereotype congruency, level of androgyny, and their interaction had on gender categorization of Asian and Black faces. We found that increased androgyny of a face impairs gender categorization by increasing decision conflict – measured through latency, AUC, and x-flips on a mouse-tracking task. When a face had social dimensions (race and gender) that were stereotype congruent (Black and male, or Asian and female), gender categorization was facilitated and decreased the level of decision conflict. Finally, we found that level of androgyny and stereotype congruency had an interaction effect, such that stereotype congruency had a greater impact on gender categorization for more androgynous faces.

These findings are in line with previous research on gender and race put forward by Johnson et al. (2012). Our study has replicated Johnson et al.'s (2012) results that found a "race is gendered" effect; the concept "Black" was stereotype congruent with male gender, and the concept of "Asian" was stereotype congruent with female gender – resulting in facilitated categorization of Black male faces and Asian female faces. The "race is gendered" effect was marked by a main effect of stereotype congruency creating lower levels of decisional conflict on all three mouse-tracking outcomes; latency, AUC, and x-flips.

However, Johnson et. al.'s (2012) original research did not explore androgyny. Our study extended the work put forward by Johnson et al. (2012) and found that the "race is gendered" effect is not of the same magnitude for all faces – it is even more true for faces that are harder to categorize (i.e., those that display androgyny). Specifically, our analysis produced an interaction effect of stereotype congruency and level of androgyny for all three measures of decision conflict. Together, these findings indicate that the more androgynous the face, the more of a role stereotype congruency plays on categorization.

Cognitive Load and Stereotype Effects

It is possible that cognitive load perspectives may also shed light on the results found in the present study; when androgyny introduces cognitive load to the gender categorization task by making the task more difficult, the facilitating or inhibiting effect of race-gender stereotypes plays a larger role in decision making. This is in line with our results, which showed an increase in response time (latency), and uncertainty (AUC, x-flips) when categorizing faces of higher androgyny as male or female.

In our study, stereotype incongruency had a larger effect on decision conflict when the face stimuli being categorized were more androgynous. In other words, the shared stereotypes between race and gender had a larger influence on gender categorization when the task may have imposed higher cognitive load on the participant. High cognitive load has been shown to increase stereotype effects (Biernat et al., 2003). For example, judgements made under high cognitive load have been shown to be biased by stereotypes (e.g., Burgess, 2009; Dijksterhuis & Van Knippenberg, 1995; Gilbert & Hixon, 1991; Macrae et al., 1993; Stangor & Duan, 1991; Wigboldus et al., 2004).

One experiment by Van Knippenberg et al. (1999) studied the effects of stereotypes and cognitive load on the judgment and memory of criminal acts. In this study, participants read information about a criminal case (including alleged police report with factual evidence, witness testimonies, and a suspect declaration) either at their own pace (low load), or under a quick time restraint (high load). Participants then had to free recall the case information, make a judgement of guilt, and propose a prison sentence (Van Knippenberg et al., 1999). Half the participants were given a positive stereotypical description of the suspect (a bank employee, trustworthy, respectable), while the other half were given a negative stereotypical description (a drug addict who served a sentence for burglary). Results found that judgement of guilt, punishment (prison sentence), and memory were only affected by the stereotype information in high-load conditions; negative stereotypes of the suspect evoked better memory of incriminating evidence, higher estimates of guilt, and harsher punishments.

In our experiment, androgyny presumably made the typically low-load gender categorization task a high-load task. In line with the previous research described above, the effect of stereotypes (i.e., stereotype congruency) was greater when there were high levels of androgyny (i.e., high-load) compared to low levels of androgyny (i.e., low-load).

Implications

The results of our experiment indicate - through higher measures of decision conflict - that more androgynous faces are harder to categorize as male or female. In addition, androgynous faces that are stereotype incongruent (predominantly male and Asian or predominantly female and Black) elicit even more decision conflict. Stern & Rule (2018) found that more androgynous transgender people were evaluated more negatively due to the increased amount of effort necessary to categorize them as either male or female. Similar processes may occur for more androgynous targets, as greater categorization difficulty and more decision conflict translate into lower interpersonal evaluations. In the real world, if an individual appears to be physically androgynous and their race happens to be stereotype incongruent, this may mean that they are also evaluated by others even more negatively due to the impaired gender categorization demonstrated in our study.

Limitations

One clear limitation of this study is the sample of participants. The majority of the participants were female (66.5%), and White (68.8%). This makes the results of the study less generalizable to the general population. Previous gender categorization research has found that participants categorize same-sex targets faster than those of the opposite sex (Zarate & Smith, 2011). From this we can infer that the majority female participants could have enhanced the gender categorization of all female targets compared to male targets. If we had a more representative sample, with more male participants, we may have found slightly less of a facilitating effect of Asian and Female targets, and slightly more of a facilitating effect of Black and male targets.

Also, worth mentioning, is that the participants were voluntarily participating through Project Implicit. Recent research has found that Project Implicit participants tend to be more liberal than the general US population (Charlesworth & Banaji, 2019). Those who have more conservative ideology show greater endorsement of binary gender beliefs and are more likely to display prejudice against gender non-conformists (Prusaczyk & Hodson, 2019). Believing that gender is binary could alter the participants ability of differentiating a gender ambiguous face into binary categories. It may be more difficult for someone who believes gender is a continuum to gender categorize an androgynous face, but easier for someone who has strong binary gender beliefs. If we had a more representative sample – with more conservative participants – the decision conflict may be lower for androgynous faces than what we measured.

Our study extends the work of Johnson et al.'s (2012) findings of race-gender stereotype congruency and their effect on gender categorization. The Johnson et al. (2012) study used only White participants. It is a limitation that our sample is not demographically the same in terms of recreating the results, however, our more diverse sample pool demonstrates that the "race is gendered" effect is not limited to White participants, which increases the generalizability of this and prior work.

Future Directions

To explore how cognitive load effects gender categorization of androgynous and non-androgynous faces, a future experiment could have participants complete the same gender categorization task with a working memory task condition. The working memory condition could instruct the participants to complete a digit span task (recite a six-digit string of numbers) while completing the gender categorization trials (Dillen et al., 2013). This would add cognitive load by using the working memory to recite the digits but would not directly impair the participants ability to complete the mouse-tracking task. If the digit span trials increased decision conflict more for androgynous faces relative to non-androgynous faces, this would support that androgyny increases cognitive load. If stereotype congruency impacted decision conflict more on digit reciting trials, this would support that stereotypes bias gender categorization when there is high-load.

Another study could assess if stereotype priming for gender role stereotypes and racial stereotypes strengthens the effect of stereotype congruency on gender categorization. Previous research has demonstrated that stereotype priming facilitates responses to stereotypic trials (Blair & Banaji, 1996). In a follow-up study, participants could be randomly assigned to either a gender role stereotype prime condition, an Asian/Black racial stereotype prime condition, a gender and racial stereotype prime condition, or a control prime condition (prime for semantically unrelated concept). Participants would be primed by attending to images that endorse the conditions particular stereotype. For example, the gender role stereotype condition could be primed by showing the participant a photograph of a women cooking, and a man in a corporate office. Participants would then complete the same mouse-tracking gender categorization task as described in our experiment. This design can then assess whether the effects of stereotype priming carry over into the gender categorization task. The priming manipulation of relevant gender and race stereotypes could lead to an increase effect of stereotype congruency on the mouse-tracking gender categorization task. Such a study would provide important information on how mouse-tracking performance is related to which stereotypes are activated, and more broadly, how gender categorization is related to the degree to which stereotypes are cognitively accessible.

A final follow up study could assess how participants counterstereotype intention changes the effect of stereotype congruency on the gender categorization task. Counterstereotype intention is an individual’s conscious will to think or act counter to a stereotype (Blair & Banaji, 1996). Previous research has shown that counterstereotype intention reduced the effect of stereotype priming under high cognitive restraints and reversal of stereotype priming under low cognitive restraints (Blair & Banaji, 1996). Thus, future research could assess how informing participants of the "race is gendered" effect (Johnson et al., 2012) and instructing the participants to counter it would affect decision conflict in the gender categorization task. These results could then shed light on the degree to which mouse-tracking performance can be consciously controlled and changed via participants' goals.

References

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. https://doi.org/10.1037/h0036215

Biernat, M., Kobrynowicz, D., & Weber, D. L. (2003). Stereotypes and Shifting Standards: Some Paradoxical Effects of Cognitive Load. *Journal of Applied Social Psychology*, 33(10), 2060–2079. doi:10.1111/j.1559-1816.2003.tb01875.x

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, 70(6), 1142–1163. https://doi.org/10.1037/0022-3514.70.6.1142

Brewer, M. B. (1988). *A dual process model of impression formation.* In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition, Vol. 1. A dual process model of impression formation* (p. 1–36). Lawrence Erlbaum Associates, Inc.

Burgess, D. J. (2009). Are Providers More Likely to Contribute to Healthcare Disparities Under High Levels of Cognitive Load? How Features of the Healthcare Setting May Lead to Biases in Medical Decision Making. *Medical Decision Making*, 30(2), 246–257. doi:10.1177/0272989x09341751

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, *8*(2), 237–262. https://doi.org/10.1080/13506280042000072

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233–246. doi:10.1111/j.2044-8279.1992.tb01017.x

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. long-term change and stability from 2007 to 2016. *Psychological Science*, 30(2), 174–192. https://doi.org/10.1177/0956797618813087

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. European Economic Review, 78, 97-119.

Dijksterhuis, A., & Van Knippenberg, A. (1995). Timing of schema activation and memory: Inhibited access to inconsistent information. *European Journal of Social Psychology*, 25(4), 383–390. doi:10.1002/ejsp.2420250403

Van Dillen, L. F., Papies, E. K., & Hofmann, W. (2013). Turning a blind eye to temptation: how cognitive load can facilitate self-regulation. *Journal of Personality and Social Psychology*, 104(3), 427.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in experimental social psychology*, *23*, 1-74.

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

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, 27(5), 315-323.

Freeman, Jonathan B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*, 27(4), 502–517. https://doi.org/10.1177/0956797615627418

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60(4), 509–517. https://doi.org/10.1037/0022-3514.60.4.509

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85 (2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

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

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. https://doi.org/10.1037/a0025335

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational psychology review*, 21(1), 31-42.

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

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, 23(1), 77–87. doi:10.1002/ejsp.2420230107

Melnikoff, D. E., Mann, T. C., Stillman, P. E., Shen, X., & Ferguson, M. J. (2021). Tracking prejudice: A mouse-tracking measure of evaluative conflict predicts discriminatory behavior. *Social Psychological and Personality Science*, *12*(2), 266-272.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358–368. doi:10.1037/0022-0663.91.2.358

Nosek, B. A. (2005). Moderators of the Relationship Between Implicit and Explicit Evaluation. *Journal of Experimental Psychology: General*, 134(4), 565–584. https://doi.org/10.1037/0096-3445.134.4.565

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). Psychology Press.

Nowak, K. L., & Denes, A. (2016). Androgyny. In A. Wong, M. Wickramasinghe, renee hoogland, & N. A. Naples (Eds.), *The Wiley Blackwell encyclopedia of gender and* *sexuality studies* (pp. 1–3). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118663219.wbegss282

Prusaczyk, E., Hodson, G. The Roles of Political Conservatism and Binary Gender Beliefs in Predicting Prejudices Toward Gay Men and People Who Are Transgender. *Sex Roles* 82, 438–446 (2020). https://doi.org/10.1007/s11199-019-01069-1

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351–371. doi:10.1007/bf02213420

Stangor, C., & Duan, C. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology*, 27(4), 357–378. doi:10.1016/0022-1031(91)90031-z

Stillman, P. E., Krajbich, I., & Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences. *Proceedings of the National Academy of Sciences*, *117*(50), 31738-31747.

Stroessner, S. J. (1996). *Social Categorization by Race or Sex: Effects of Perceived Non-Normalcy on Response Times. Social Cognition, 14(3), 247–276.* doi:10.1521/soco.1996.14.3.247

Stern, C., & Rule, N. O. (2018). Physical androgyny and categorization difficulty shape political conservatives' attitudes toward transgender people. *Social Psychological and Personality Science*, 9(1), 24–31. https://doi.org/10.1177/1948550617703172

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making–the effect of time pressure and cognitive load on moral judgment and altruistic behavior. PloS one, 11(10), e0164012.

Van Knippenberg, A., Dijksterhuis, A., & Vermeulen, D. (1999). Judgement and memory of a criminal act: the effects of stereotypes and cognitive load. *European Journal of Social Psychology*, 29(2-3), 191–201. doi:10.1002/(sici)1099-0992(199903/05)29:2/3<191::aid-ejsp923>3.0.co;2-o

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, 5(1-2), 39–43. doi:10.1002/jip.82

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & Knippenberg, A. van. (2004). Capacity and Comprehension: Spontaneous Stereotyping Under Cognitive Load. Social Cognition, 22(3), 292 309. doi:10.1521/soco.22.3.292.35967

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social cognition*, 8(2), 161-185.