Original Article

Stereotypes shape response competition when forming impressions

Neil Hester,[1] 🆔 **Sally Y. Xie,[1] Jeannine Alana Bertin[2]** 🆔
and Eric Hehman[1]

Abstract

Dynamic models of impression formation posit that bottom-up factors (e.g., a target's facial features) and top-down factors (e.g., perceiver knowledge of stereotypes) continuously interact over time until a stable categorization or impression emerges. Most previous work on the dynamic resolution of judgments over time has focused on either categorization (e.g., "is this person male/female?") or specific trait impressions (e.g., "is this person trustworthy?"). In two mousetracking studies—exploratory (N = 226) and confirmatory (N = 300)—we test a domain-general effect of cultural stereotypes shaping the process underlying impressions of targets. We find that the trajectories of participants' mouse movements gravitate toward impressions congruent with their stereotype knowledge. For example, to the extent that a participant reports knowledge of a "Black men are less [trait]" stereotype, their mouse trajectory initially gravitates toward categorizing individual Black male faces as "less [trait]," regardless of their final judgment of the target.

In North America, Black men are often stereotyped as less intelligent than White men (Devine & Elliot, 1995; Williams & Mohammed, 2013), even by individuals who do not explicitly endorse such a viewpoint. Imagine separate situations in which both a Black and a White man are encountered and evaluated as equally intelligent. Although the ultimate conclusion about each man is the same—that they are "very intelligent"—do perceivers reach this conclusion through the same underlying cognitive process? How a perceiver arrives at "very intelligent" in each scenario may be different because of perceivers' awareness of cultural stereotypes. Stereotypic knowledge about gender, race, and other

[1]Department of Psychology, McGill University, Canada
[2]Department of Psychology, New York University, USA

*Neil Hester is now affiliated to Department of Psychology, University of Waterloo, Canada; Sally Y. Xie is now affiliated to Department of Psychology, Princeton University, USA

Corresponding author:
Neil Hester, Department of Psychology, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada.
Email: neilrhester@gmail.com

characteristics shapes our expectations, impacting which people are regarded as intelligent (or trustworthy, or any other trait). Thus, North American perceivers may initially be inclined to perceive the very intelligent-looking Black man as "less intelligent" given their stereotypical expectations, even if this stereotype is corrected or updated milliseconds later in the impression formation process as more information is integrated. Here, we investigate whether *stereotype–feature congruence*—the degree of agreement between stereotype knowledge and a target's overall features—has an impact on the way in which the psychological process of impression formation unfolds over time. This research extends existing work on impression formation by examining the interplay between perceiver characteristics and target characteristics throughout the *process* of impression formation (i.e., unfolding over time), rather than merely examining how these effects influence the final product of the impression formation process.

Background

Impressions formed from others' appearance are multiply determined. Theoretical models have generally broken down contributions as being from (a) the target's characteristics, (b) the perceiver's characteristics, and (c) interplay between target and perceiver characteristics (Funder, 1995; Hehman et al., 2019; Hester et al., 2021; Hönekopp, 2006; Kenny et al., 2006). The first part is intuitive—what targets look like influences perceivers' impressions of them. The second part includes ways in which perceiver differences influence their impressions of targets (Kawakami et al., 2017; Kunda & Thagard, 1996; Macrae & Bodenhausen, 2000). For example, perceivers' moods could have an impact on their impressions, or they might have a stable tendency to give more positive or negative impressions overall. The third part of variability in impression formation—the interaction between perceiver and target—is explained by perceiver and target characteristics together. For example, many stereotype findings involve the combination of perceiver-level endorsements of stereotypes and target-level social identity. Research using variance decomposition techniques

has consistently revealed that this is the largest source of variance in impression formation (Hehman et al., 2017; Hester et al., 2021; Hönekopp, 2006; Xie et al., 2019). Whereas the target and perceiver characteristics alone seem to each contribute roughly 10%–20% of the variance in a given impression, perceiver × target interactions account for between 30%–40% (Hehman et al., 2017; Xie et al., 2019).

Perceiver × target interactions are often at the heart of social cognitive questions and research, encompassing topics such as how perceivers' cognitions influence behaviors differently when they interact with members of different social groups. For example, prejudiced people rate own- and other-group individuals differently, whereas unprejudiced people do so to a lesser extent (Hugenberg & Bodenhausen, 2003; Hutchings & Haddock, 2008). Impression formation theories present models for how perceiver × target interactions occur (Freeman & Ambady, 2011; Rhodes, 2006; Todorov et al., 2015; Webster & MacLeod, 2011; Zebrowitz & Montepare, 2005) throughout the impression formation process, resulting in a final impression. A modern example is the dynamic interactive model of impression formation, which suggests that the processing of bottom-up facial features is constrained by top-down cognition (e.g., emotion, stereotypes, goals) (Freeman et al., 2020).

Although models of person perception commonly implicate perceiver × target interactions in the impression formation process (i.e., how impressions unfold over time), formal tests of these interactions are still uncommon, as most perceiver × target interactions address the final impression rather than the underlying process. This is partially due to methodological challenges. However, understanding the nature of these interactions is central to such models and their predictions. In this article, we examine the role of perceiver × target interactions in the impression formation process using a novel mousetracking approach.

Tapping Into the Process of Impression Formation

Researchers face a difficult challenge when measuring the specific processes underlying impression

formation: impression formation unfolds very quickly over time and involves difficult-to-measure psychological factors. Much of the work examining the process underlying impression formation relies on indirect inferences from outcomes such as reaction time or error rate. For example, when people are asked to classify threatening versus non-threatening objects accurately, racial information has an impact on both reaction times and error rates (e.g., tools versus guns) (Payne, 2001; Todd et al., 2016), which allows for indirect inferences about the underlying process, but not for direct measurement of this process. Other measures, such as event-related potentials, can provide more direct insight by recording fluctuations in brain activity (Correll et al., 2015; Kubota & Ito, 2014); however, these techniques can be difficult to use and interpret.

The limitations of these techniques are at least partly addressed by mousetracking, which provides high-resolution, millisecond-level information about people's decision process when deciding between two or more competing response options (Dale et al., 2007; Freeman & Ambady, 2010). Mousetracking has been employed in multiple domains to investigate and advance models of cognitive theory (Stillman et al., 2018). Outside of the social perception domain, mousetracking research has explored the real-time resolution of response competition in the domain of self-control (Gillebaart et al., 2016; Schneider et al., 2015; Stillman et al., 2017) and moral decision-making (Koop, 2013). With regard to person perception, mousetracking research has provided evidence of cognitive competition in the process of gender and race categorization of atypical faces (Freeman et al., 2010; Freeman & Ambady, 2011; Hehman et al., 2014). Finally, mousetracking research has broadly illustrated the dynamic role of stereotypes in social cognition. For instance, race and gender cues in the face interact with stereotype information to drive the categorization of targets' race, gender, and emotions (Brooks & Freeman, 2018; Johnson et al., 2012).

The vast majority of tests that have used mousetracking to test person perception models have focused on the process of forming social categorizations (e.g., "is this person male or female?") rather than the process of forming trait impressions (e.g., "how attractive is this person?"), which is the focus of the present research. Social categorization and trait impressions share some similarities (Freeman et al., 2020) and one might assume that they involve the same processes, such that research on the process of forming social categorizations accurately generalizes to the process of forming trait impressions. However, there are reasons to doubt this assumption, necessitating research specifically on the process of forming trait impressions. In the next section, we outline differences between social categorization and trait impressions. Our goal here is not to set up a direct experimental comparison between social categorization and trait impressions, but rather to justify the importance of specifically examining the process underlying trait impressions.

Differences Between Forming Social Categorizations and Trait Impressions

There is evidence suggesting that the processes underlying social categorization are distinct from those underlying trait impressions. To start, categories are usually discrete (it is intuitive to say that someone is male *or* female), whereas trait impressions are usually continuous (someone can be *more* or *less* attractive). Although processes underlying discrete categorizations are continuous, as ongoing response competition between multiple choices resolves itself over time to arrive at a categorization (Dale et al., 2007; Freeman et al., 2008), the nature of the judgment itself is nevertheless different. The determinants of these patterns of fluctuation might differ depending on whether choices reflect discrete categories (e.g., "male" versus "female") or continuous trait impressions (e.g., "more" versus "less" attractive).

In addition, impressions (e.g., trustworthy, immoral) are typically more valenced than categorizations (e.g., male, female), just as perceiver knowledge about trait concepts is more valenced than categorizations. As a result, categorization

is more likely to leverage perceiver knowledge that is grounded in base rates (e.g., "on average, women have longer hair than men"), whereas impressions are more likely to leverage perceiver knowledge or endorsement of valenced stereotypes (e.g., "women are less competent" or "Black people are lazier"). These differences suggest that idiosyncratic perceiver variability might play a larger role in forming impressions than in categorization processes. Indeed, past work has found that observers have a high degree of consensus when inferring the sex or race category of faces, but a much lower degree of consensus when inferring traits (e.g., attractive) from faces (Bjornsdottir et al., 2021; Hehman et al., 2017). Given our focus on how top-down perceiver characteristics interact with bottom-up target features to influence impressions, a clear understanding of this perceiver variability is important.

Another critical distinction between social categorization and first impressions is temporal precedence. If we assume that race and gender have an impact on trait impressions of faces via stereotypes, then this necessitates the categorization of faces into different races and genders prior to impression formation (Cloutier et al., 2005; Funder, 1995; Macrae & Martin, 2007). This categorization of race (occurring within 100–120 ms) and gender (occurring within 150–200 ms) emerges early in the perceptual process (Ito & Urland, 2003, 2005; Zhang et al., 2018). Because the morphological facial features necessary for an individualized trait impression are processed later in the impression formation pipeline than race and gender categorization (Ambrus et al., 2019; Cloutier et al., 2005; Dobs et al., 2019), evidence for the integration of race and gender stereotypes with morphological features (as captured by mousetracking) would be gravitation toward stereotypical responses early in the impression formation process, followed by subsequent adjustment as individual facial features are incorporated.

Together, these distinctions support the usefulness of research specifically examining the process of forming trait impressions. Next, we review existing work on the process of forming trustworthiness impressions specifically, then describe how our research approach attempts to form more generalizable conclusions about the process of forming trait impressions.

Uncovering the Process of Forming Trait Impressions

Although little research has explored the process of forming trait impressions, there have been some notable exceptions studying impressions of trustworthiness specifically. Some research has examined these perceptions across the lifespan (Cassidy et al., 2019), finding that older adults exhibited a tendency to rate faces as more trustworthy. Other research focusing on face–context integration asked participants to view faces in the presence of threatening, negative but unthreatening, or neutral contexts (Brambilla et al., 2018). Results were consistent with predictions of their model in that contextual information was integrated into both the process, and the final trustworthiness ratings of the faces.

Building on this previous research, we conducted a broader examination of the process of forming impressions across numerous traits. Trustworthiness is an important impression and the focus of much research, but—relative to many other trait impressions—it is particularly informed by the facial appearance of a target (Hehman et al., 2017), especially the extent to which the target appears to be smiling or frowning (presumably due to overgeneralization processes) (Zebrowitz et al., 2010). Consistent with the dynamic interactive model (Freeman et al., 2020), we posit that impressions are formed through the integration of top-down perceiver information and bottom-up stimuli appearance. To support this broad hypothesis, we examine impression formation processes across (a) traits, (b) target race, and (c) target gender. This approach, which emphasizes generalizability and replicability across varied conditions, is critical for achieving a fuller understanding of impression formation.

"Top-down perceiver information" could refer to any individual difference, but given the close relationship between stereotypes and trait impressions, we focus on perceivers' knowledge of cultural stereotypes, regardless of whether they personally endorse those stereotypes. For example, take the finding that the association of "Asian" with "femininity" leads to a North American cultural stereotype of Asian women as attractive and Asian men as less attractive (gendered race hypothesis) (Galinsky et al., 2013; Johnson et al., 2012; Schug et al., 2015). When there is high congruence between the stereotypes activated by a face's race and gender category and a face's morphological features (e.g., an Asian female face with attractive facial features), we predicted that participants' evolving impression—reflected by a mouse trajectory toward their choice—would take a relatively straight path, indicating little attraction toward the alternate choice (i.e., lower response competition). However, when there is low congruence between the stereotypes activated by a face and the face's overall features (e.g., an Asian male face with attractive facial features), we predicted that participants' impressions would initially exhibit greater attraction toward the unselected choice (i.e., higher response competition). We refer to this predicted pattern of outcomes as *stereotype–feature congruence*.

Although our example specifically focuses on attractiveness, we expect stereotype–feature congruence to be a trait-general phenomenon, as predicted by the dynamic interactive model of person perception (Freeman et al., 2020). In other words, the model does not predict the pattern of effects to be moderated by the specific trait according to which perceivers form impressions. Traits do differ in their characteristics in several ways, for example, the morphological characteristics they are based upon (Hehman et al., 2015; Oosterhof & Todorov, 2008), how they correlate with other traits (Oh et al., 2020; Sutherland et al., 2013; Xie et al., 2021), and the amount of agreement between perceivers with regard to a trait impression (Hehman et al., 2017; Xie et al., 2019). However, to the extent that an individual has

stereotype knowledge about a given trait and its association with a target's category (e.g., race and gender), we predict this stereotype information will influence the impression formed no matter the trait, because these aspects of specific traits should not be relevant to the process by which impressions unfold.

Sharp Versus Graded Response Curves

A secondary aim in our analyses is to examine the trajectories of perceivers' trait impressions over time. There are several possible ways in which impressions can be formed, and the high-resolution data that mousetracking provides can help distinguish between mechanisms underlying impressions. These can reflect either sequential processes or the simultaneous integration of information. Past research has attempted to demarcate the perceptual pipelines leading to an impression (Macrae & Bodenhausen, 2000). From a target's appearance, category information informs a social categorization (e.g., White); then, stereotypes associated with this categorization (e.g., trustworthy) are activated. The classic sequential view of this process is that social categorization must be completed prior to the stereotype information being accessed. In contrast, modern work informed by dynamic models (Freeman et al., 2008, 2020; McKinstry et al., 2008) has argued that these processing stages can be continuous and simultaneous rather than discrete and sequential. In this model, partially processed information about a target's appearance continuously updates a category impression (e.g., White), which in turn continually updates partially active impressions (e.g., trustworthy) that are informed by stereotypes. This relationship is bidirectional, as the activation of stereotype information (e.g., trustworthy) further constrains categorization (e.g., White).

The patterns of trajectories in a mousetracking paradigm can reveal which of these processes is occurring. The sequential processing possibility would predict a mouse trajectory in which participants are initially attracted to a

response, but then abruptly "correct" this tendency mid-flight as they integrate additional stereotype information sequentially. Patterns indicative of this response often have a straight line toward one response that abruptly shifts at a sharp angle toward the other response, resulting in a bimodal distribution of area under the curve (AUC). On the other hand, top-down stereotypes and bottom-up morphological information might instead be simultaneously and continuously integrated throughout the response. Participants demonstrating such a pattern would show smooth competition and a slow shift from one response category toward the other, with the mouse trajectory indicating a curve instead of sharp angles (Dale et al., 2007). In this case, the distribution of AUC scores would be unimodal and relatively normal. These two patterns are indicative of distinct underlying processes of impressions.

Previous research has found more evidence for graded responses (Freeman et al., 2008, 2010; Johnson et al., 2012) in categorization processes. We test whether these results generalize to trait impressions, which would provide further evidence for a dynamic interactionist perspective.

The Present Studies

We conducted two studies. The first was an in-lab exploratory study using MouseTracker (Freeman & Ambady, 2010). The second was a preregistered direct replication and extension using an online sample and a mousetracking task programmed using MinnoJS (Zlotnick et al., 2015). Prior to COVID-19, we had preregistered a second in-lab direct replication, but changed our plan when in-person data collection became unfeasible.

Open Data, Open Materials, and Preregistration

Our complete data and the analysis script, as well as an anonymized copy of our preregistration, are available at https://osf.io/mgsza. In both studies, we report all measures, manipulations, and exclusions.

Study 1

Participants and Design

We recruited 238 participants through a participant pool. Each participant completed 300 main trials (150 faces, two ratings each) in the mouse-tracking task in a 3 (Target race: East Asian, Black, White) $\times$ 2 (Target gender: female, male) $\times$ 6 (Trait: attractive, dominant, friendly, intelligent, physically strong, trustworthy) mixed design, with repeated measures on the first two factors. These six traits were chosen based on previous work that identifies trustworthiness, dominance, and youthfulness/attractiveness as primary dimensions of face perception (Sutherland et al., 2013). Fourteen of these participants were excluded during the data cleaning process (details below) and four asked that their data not be included, leaving us with 220 participants for analysis (81% women, 18% men, 0.5% other gender identity, 0.5% no response, $M_{age} = 20.34$ years, $SD_{age} = 2.47$ years).

Using *simr* (Green & MacLeod, 2016), we conducted a sensitivity power analysis to estimate the minimum effect size for the key effect that we could reasonably detect using this sample. The estimate suggested that this sample is sufficient to detect an effect size of $b = -0.040$ with 80% power.

Materials and Procedure

Stimuli. Stimuli were Asian, Black, and White male and female faces with neutral expressions pseudo-randomly selected from the Chicago Face Database (Ma et al., 2015). Twenty-five stimuli for each group were presented in order randomized by participant. Participants rated each face twice for a single trait.

Mousetracking task. Participants first received instructions for a mousetracking task programmed in MouseTracker (Freeman & Ambady, 2010). Instructions were as follows: "We are interested in how people evaluate faces. Today, you will be evaluating faces on their [TRAIT]. For every face that appears, your task

is to indicate whether they are more or less [TRAIT]. You will do so by clicking the selections at the top corners of the screen with your mouse. It is important to respond AS QUICKLY AS POSSIBLE. If you don't respond quickly enough, you will receive a warning."

Following practice trials to familiarize themselves with the task, participants completed the main task. They were encouraged to take a short break at 50-trial intervals, for a total of five breaks. For every trial, the cursor was re-centered at the bottom of the screen. Target faces appeared slightly above this starting position. If participants took longer than 500 ms to begin moving their mouse at the beginning of a trial, they received a warning message after the trial asking them to move more quickly.

The two response choices—"Less [TRAIT] than average" and "More [TRAIT] than average"—were displayed in the upper left and upper right corners of the screen, respectively. We anticipated variation in what perceivers considered "average," and interpret these responses as above or below participants' idiosyncratic perceptions of average. Given our theoretical interest in top-down and bottom-up integration of information, we considered it ideal that this design naturally adjusted for participants' individual perceptions.

Participants' responses to the MouseTracker task were recorded as raw mouse coordinates, which measure the x and y coordinates of the mouse every ~15 ms (exact time varies according to the Hz of the screen). These coordinates were time-normalized, each trajectory having 101 equally spaced time units (Freeman & Ambady, 2010), and then used to derive the AUC, which measures the degree of spatial attraction of the mouse cursor toward the unselected choice. AUC is calculated as the area between the actual trajectory and an idealized straight-line trajectory from the origin to the final choice. AUC is the outcome on which we focused our analyses, although we also report analyses using spatial entropy in the markdowns provided on the OSF page (Hehman et al., 2015). The spatial entropy analyses produced similar patterns of results.

Stereotype knowledge task. After completing the mousetracking task, participants completed a set of questionnaires asking them about North American stereotypes with regard to various race × gender groups. Participants indicated the extent to which they felt that each of the six race × gender groups was stereotypically perceived as each of the six traits, for a total of 36 questions. The wording was as follows: "To what extent are [RACE] [GENDER] stereotyped as [TRAIT]?" Participants gave their responses on a seven-point Likert-type scale ranging from 1 = "not at all" to 7 = "completely". Previous research has measured stereotype knowledge in a similar manner (e.g., Stolier & Freeman, 2016; Xie et al., 2021) and found it corresponds with impressions both behaviorally and at the neural level. Furthermore, work measuring stereotype knowledge in other ways supports its validity as a construct (e.g., Devine & Elliot, 1995). In Study 2, we further examine the construct validity evidence for this measure by comparing it with stereotype endorsement.

Data Cleaning

Data were processed and cleaned, consistent with previous research (Freeman et al., 2010). In order to compare mouse trajectories across all participant response choices, we time-normalized trajectories to 101 timesteps and horizontally flipped trajectories so that both response options (e.g., "less intelligent," "more intelligent") were in the same direction regardless of response selection. Any trials with an initiation time outside the range (*min* = 1 ms, *max* = 500 ms) and any trials with a response time outside the range (*min* = 100 ms, *max* = 3000 ms) were removed. On visual inspection of individual trajectories, we removed outlier mouse trajectories that either (a) showed initial horizontal deviation from the cursor origin or (b) contained looping. Additionally, data from 14 participants were excluded from analysis because they consisted uniquely or largely of trajectories with frequent looping, indicative of participants not participating in the task. In total, 94% of trials were retained for analysis.

Results

Data and code are available at https://osf.io/mgsza.

Analysis Strategy

For each participant, each trial was assigned a *stereotype knowledge score*, which was the participant's stereotype knowledge for the target's race-by-gender category for the trait currently being rated. For example, if a participant answered the question "To what extent are Black women stereotyped as intelligent?" with a "3," in all mousetracking trials rating the intelligence of Black women, this participant's stereotype knowledge score would be 3. The use of this score incorporates race, gender, and stereotype knowledge ratings in a single variable, improving the interpretability of our findings and simplifying the statistical model. Put differently, this stereotype knowledge score allows us to test our primary hypothesis via a two-way interaction between response and stereotype knowledge score, rather than a five-way interaction between response, stereotype knowledge, trait, race, and gender.

By combining this stereotype knowledge score with participants' responses to a given trial, we captured the extent to which their response trajectories were congruent with stereotype knowledge. To account for the cross-classified structure of the data, we nested responses within (a) participant, (b) stimulus, (c) social category,[1] and (d) participant × social category in a multilevel framework. This specification is necessary because the stereotype knowledge score is a function of both participant and the specific race-by-gender social category being referenced. Participant response is a trial-level variable, making the interaction between participant response and stereotype a cross-level interaction. For cross-level interactions, the slope of the lower-level predictor must be specified as random at the level of the higher-level predictor to avoid severe Type I error (Heisig & Schaeffer, 2019). Accordingly, we allowed the slope of participant response to vary at the random intercept of participant × social category. Thus, the final model is as follows:[2]

$$AUC_i \sim N\left(\mu, \sigma^2\right)$$

$$\mu = \alpha_{j[i],k[i],l[i],m[i]} + \beta_{1j[i]}(\text{response}) + \beta_2(\text{stereotype}) + \beta_3(\text{stereotype} \times \text{response})$$

$$\begin{pmatrix} \alpha_j \\ \beta_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\alpha_j} \\ \mu_{\beta_{1j}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_j} & \rho_{\alpha_j\beta_{1j}} \\ \rho_{\beta_{1j}\alpha_j} & \sigma^2_{\beta_{1j}} \end{pmatrix} \right),$$

for participant : socialcat $j = 1, \dots, J$

$$\alpha_k \sim N\left(\mu_{\alpha_k}, \sigma^2_{\alpha_k}\right), \text{for participant} = 1, \dots, K$$

$$\alpha_l \sim N\left(\mu_{\alpha_l}, \sigma^2_{\alpha_l}\right), \text{for participant} = 1, \dots, L$$

$$\alpha_m \sim N\left(\mu_{\alpha_m}, \sigma^2_{\alpha_m}\right), \text{for participant} = 1, \dots, M$$

For any models that did not converge in lme4, confirmation of the parameter estimates was done using a Bayesian approach in brms (Bürkner, 2017).

AUC

In line with our stereotype–feature congruence hypothesis, the interaction between response and stereotype knowledge score was significant, $b = -0.10$, $t(1414) = -6.85$, $p < .001$, 95% CI [−0.13, −0.07]. When participants responded that a target was "less" of a trait, a higher stereotype knowledge score was associated with a larger AUC. When participants responded that a target was "more" of a trait, a greater stereotype knowledge score was associated with a smaller AUC. In Figure 1, we visualize this effect by using a graph to depict time-normalized trajectories for "less" and "more" responses at the lower and upper tertiles of stereotype knowledge score.

The main effect of stereotype knowledge score was not significant, $b = -0.01$, $t(695) = -0.99$, $p = .32$, 95% CI [-0.02, .001]. Unrelated to our hypotheses, we did observe a main effect of response, such that "more" responses showed higher AUC scores than "less" responses, $b = 0.17$, $t(1252) = 11.10$, $p < .001$, 95% CI [0.14, 0.20], which was qualified by a response × trait interaction, detailed below. The full model (with model statistics and intraclass correlations) is available in the Markdown file on the OSF page under "Main Analyses."

Figure 1. Time-Normalized Trajectories by Stereotype Knowledge and Response
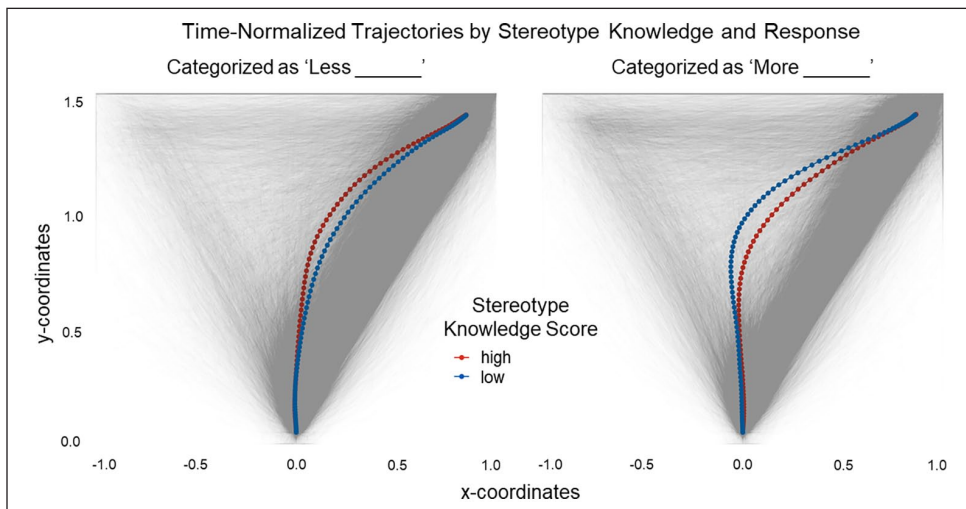
Note. Red and blue lines represent the average trajectories for trials in the upper and lower tertiles of stereotype scores. Tertiles are for visualization purposes only. Points along the red and blue lines represent the 101 time-normalized steps. Gray lines represent individual trajectories.

Generalizability of Effects Across Traits

We sampled six different traits between subjects in order to improve the generalizability of our findings—for any single trait, it could be the case that any effect was specific to that trait, rather than applicable to all traits more broadly. Accordingly, we tested whether effects were dependent on trait. We found no significant interaction between trait and our key interaction of response × stereotype knowledge score according to AUC, $F(5, 1358) = 1.20$, $p = .31$, suggesting that evidence for stereotype–feature congruence is consistent across different traits.

We did observe a significant interaction between trait and response, $F(5, 1458) = 61.32$, $p < .001$, such that the way response choice predicted AUC varied considerably depending on the trait: for attractive, $b = 0.58$ (95% CI [0.52, 0.64]); for dominant, $b = 0.21$ (95% CI [0.13, 0.30]); for friendly, $b = 0.19$ (95% CI [0.10, 0.28]); for intelligent, $b = –0.05$ (95% CI [–0.14, 0.03]); for physically strong, $b = 0.05$ (95% CI [–0.04, 0.14]); for trustworthy, $b = –0.16$ (95% CI [–0.26, –0.06]). See Figure 2. In addition, this heterogeneity suggests that people's initial tendency to begin evaluating someone as "more" or "less" of a trait, independent of stereotype knowledge scores or race × gender, varies depending on the trait. For example, these results suggest that people tend to move toward "less" for attractiveness impressions but toward "more" for trustworthiness impressions (this is visible in Figure 2 by focusing on the "overall" trajectories for "less" and "more" by trait). See the Markdown file for full statistics by trait.

Distributional Analyses

Next, we examined the patterns of mouse trajectories to conclude whether they support sequential processing of appearance and stereotype information versus simultaneous integration of this information. The mouse trajectories evident in Figures 1 and 2 are apparently smooth and graded. As described in the introduction, such a pattern is evidence of continuous and simultaneous processing of top-down and bottom-up information. However, such a conclusion would be preliminary, as these apparently smooth trajectories could be the result of averaging together trials showing very little attraction to the

Figure 2. Time-Normalized Trajectories by Stereotype Knowledge and Response
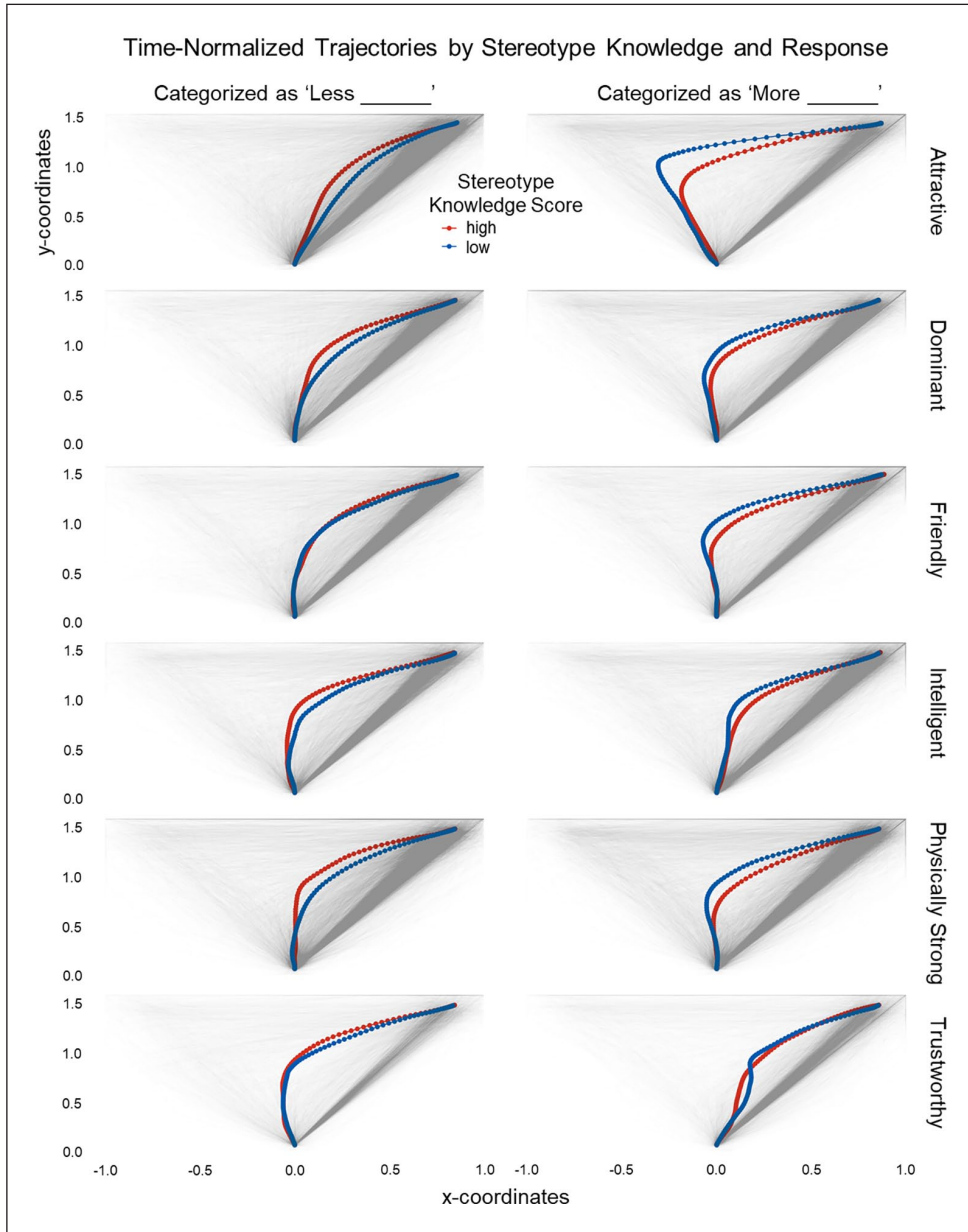
Note. Red and blue lines represent the trajectories for trials in the upper and lower tertiles of stereotype scores. Tertiles are for visualization purposes only. Points along the red and blue lines represent the 101 normalized time steps. Gray lines represent individual trajectories.

category not selected, and trials showing discrete mid-flight corrections (i.e., evidence of sequential processing).

To examine this possibility, and as in previous research (Freeman et al., 2008), we plotted the distributions of AUC across different levels of

Figure 3. Density Plots for Area Under the Curve Across Stereotype Knowledge Scores
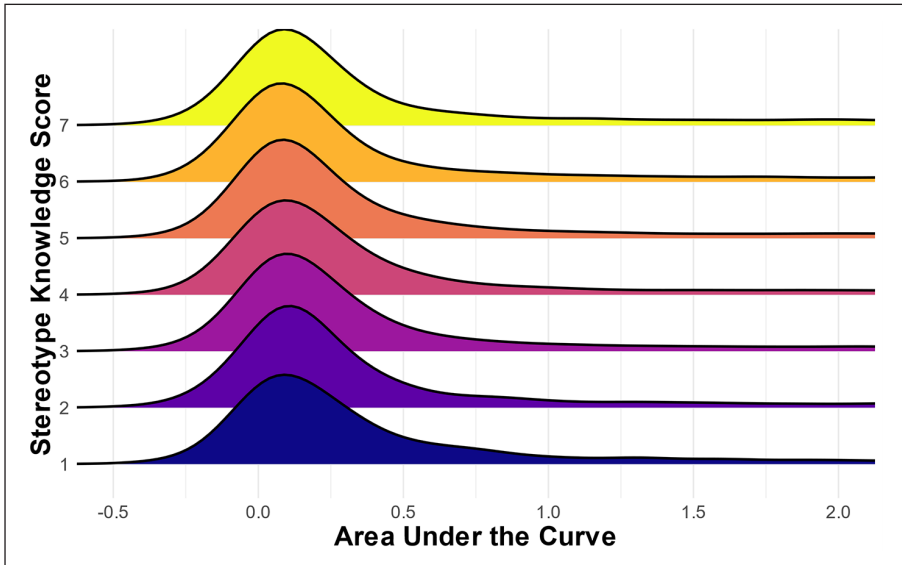
Note. Distributions of area under the curve scores were approximately the same across levels of stereotype knowledge score (i.e., the participants' reported stereotype knowledge for a given race-by-gender category).

stereotype knowledge scores in order to assess whether there are two distinct subpopulations of trajectories (i.e., bimodality) or one single distribution (Figure 3). Results clearly show that there was no bimodality at any level of stereotype knowledge. Accordingly, we interpreted these graded responses as evidence of simultaneous integration of stereotype knowledge and target appearance throughout the impression formation process, consistent with a dynamic interactionist perspective (Freeman et al., 2020).

### Participant Stereotype Knowledge

Although the calculation of stereotype knowledge scores is useful for parsimoniously modeling our hypothesized effect, this approach does remove the descriptive component of the specific patterns in participants' stereotypes. Therefore, we modeled participants' reported stereotype knowledge as a function of race and gender to illustrate how the stereotype–feature congruence results map onto stereotype knowledge scores. Figure 4 indicates that participants' stereotype knowledge shows interactions between race and gender, in line with work on

intersectional stereotyping (Galinsky et al., 2013; Ghavami & Peplau, 2013; Johnson et al., 2012)—all six traits showed a significant interaction between race and gender, although the specific patterns varied by trait. The patterns observed align with previous work sampling from a North American population. We see that Asian men and Black women are stereotyped as especially low in attractiveness (gendered race hypothesis) (Galinsky et al., 2013; Johnson et al., 2012; Schug et al., 2015), Black people are stereotyped as high in physical strength and low in intelligence and trustworthiness (Devine & Elliot, 1995; Wilson et al., 2017), Asian people are stereotyped as low in physical strength and high in intelligence (Wong et al., 2012; Yee, 1992), and White people are stereotyped as high in attractiveness, friendliness, and trustworthiness (Fiske et al., 2002).

Discussion

This study offers clear support for our hypothesized stereotype–feature congruence effect. Perceivers' knowledge of race × gender stereotypes predicted their spatial attraction to the response choice

Figure 4. Stereotype Knowledge by Race and Gender
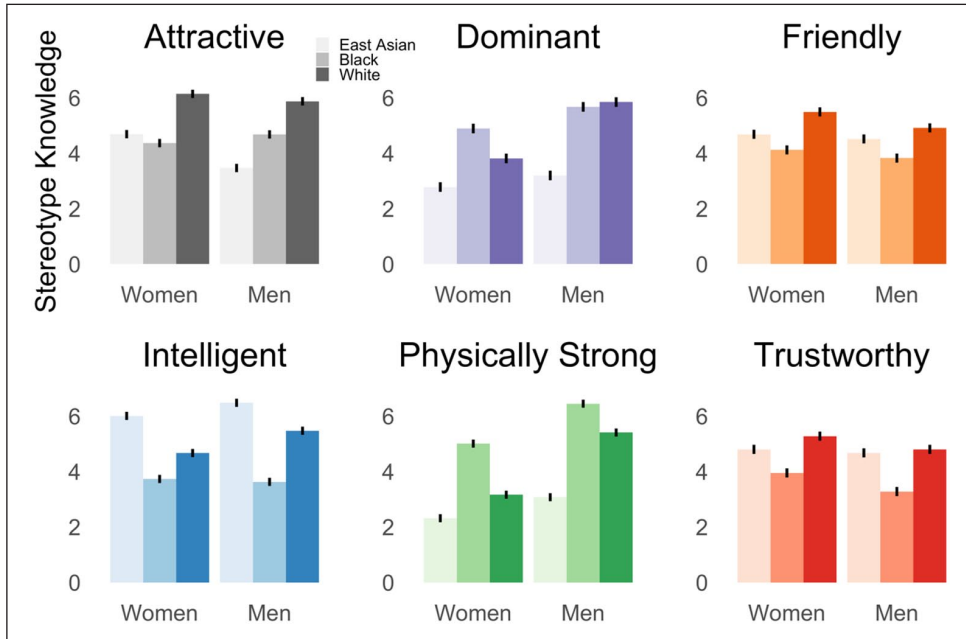
Note. Scores show the extent to which participants indicated that each race × gender group was stereotypically perceived as each of the six traits. Black error bars represent 95% confidence intervals.

congruent with the stereotype, regardless of their final response choice. This finding suggests that within milliseconds in the decision-making process, participants' impressions of faces are guided by top-down knowledge of stereotypes. This result supports previous theory and research implicating top-down knowledge of stereotypes in impressions and, critically, provides novel evidence that this top-down knowledge interacts with the content of impressions (i.e., the specific trait being rated), influencing how the process underlying these impressions unfolds.

Study 2

We conducted Study 2 as a direct preregistered replication: https://osf.io/jmfz2. For Study 2, we collected data online, providing a more diverse sample. In addition to participants' stereotype knowledge, we also collected information about their personal endorsements of stereotypes. We aimed to contrast the role of stereotype endorsement with stereotype knowledge in the process

of forming face impressions. To address power and participant fatigue concerns arising from our use of an online sample, we focused on three of the six traits from Study 1: attractiveness, intelligence, and trustworthiness. Again, these reflect the primary dimensions of face perception and stereotype content identified in previous work (Sutherland et al., 2013).

Method

Participants and Design

Three hundred MTurk workers (M_{age} = 37.2 years, SD_{age} = 10.9 years)[3] participated in a 3 (Target race: East Asian, Black, White) x 2 (Target gender: female, male) x 3 (Trait: attractive, intelligent, trustworthy) mixed design, with repeated measures for the first two factors. Each participant completed 120 trials.

We conducted a sensitivity power analysis to estimate the minimum effect size for the key effect that we could reasonably detect using this

sample. The estimate suggested that this sample is sufficient to detect an effect size of $b = -0.027$ with 80% power.

Materials and Procedure

Stimuli. Stimuli were identical in source and composition to Study 1, but with only 10 faces from each group due to participant attrition concerns with regard to the online sample. Again, each face was rated twice.

Mousetracking. We performed online mousetracking using the platform MinnoJS (Zlotnick et al., 2015), which is designed to measure temporal data in online studies. MinnoJS captures mouse movement at a similar high resolution to Mouse-Tracker (~15 ms, exact time varies according to the Hz of the screen), making it an ideal online substitute.

The mousetracking task was identical to the in-lab task, with the following exceptions: trials were reduced to 120 per participant due to concerns about online participant fatigue and data quality; the response area was bound to a set-size rectangle to account for different screen sizes and, thus, smaller than the in-lab response area; and the unit of measurement was pixels, rather than a standardized scale. We maintained this pixel scale rather than converting to x and y coordinates. Finally, our hypotheses are independent of any specific trait; Study 1 found that the trait being evaluated did not moderate any effects. Accordingly, we narrowed our focus to three traits instead of six for the replication. We chose attractiveness, trustworthiness, and intelligence to reflect broadly the three dimensions some have hypothesized underlie face impressions and to maximize variance among the traits.

Stereotype knowledge and endorsement. After the mousetracking task, participants completed the same stereotype knowledge questions as in Study 1. Then, they answered questions about the extent to which they personally *endorse* these stereotypes. We used both sets of responses to calculate separate congruence scores for stereotype knowledge

(our direct replication of Study 1) and stereotype endorsement (an extension of Study 1).

Data Cleaning

Data cleaning varied from Study 1 due to the different collection methods. We filtered trials outside the bounds of the study area ($xmin = -400$, $xmax = 400$, $ymin = -1$). As defined in our preregistration, we excluded trials with a latency of less than 400 ms and we excluded participants who gave a repeat response over 80% of the time. Finally, we removed outlier AUCs ($+3SD$) from analysis. All exclusions are documented in the Markdown on the OSF page. Data cleaning syntax is provided on the OSF page.

Results

Main analyses were identical to Study 1. We first report results of our direct replication, followed by the extension.

AUC

Results replicated those in Study 1. We found a significant interaction between response and stereotype knowledge score in line with our stereotype–feature congruence hypothesis, $b = -0.03$, $t(1298) = -2.80$, $p = .005$, 95% CI [-0.05, -0.01]. Consistent with the idea that stereotypes were constraining impressions as the process of impression formation unfolded, when participants responded that a target was "less" of a trait, greater stereotype knowledge scores were associated with larger AUCs. When participants responded that a target was "more" of a trait, greater stereotype knowledge scores were associated with smaller AUCs. The main effect of stereotype knowledge score was not significant, $b = -0.01$, $t(1081) = -1.14$, $p = .25$, 95% CI [-0.02, 0.01]. Replicating Study 1, we again observed a main effect of response, such that "more" responses showed lower AUC scores than "less" responses, $b = -0.06$, $t(1296) = -5.92$, $p < .001$, 95% CI [-0.08, -0.04], which was qualified by a response × trait interaction, detailed below. The full model (with model statistics and intraclass

correlations) is available in the Markdown file under "Main Analyses."

Generalizability of Effects Across Traits

To explore the generalizability of our effects, we again fit models testing effects of trait on AUC. Replicating Study 1, we found no significant interaction between trait and our key interaction of response × stereotype, $F(2, 1296) = 0.14$, $p = .87$. As in Study 1, we did observe a significant interaction between trait and response, $F(2, 1294) = 14.69$, $p < .001$, such that the way response choice predicted AUC varied by trait: for attractive, $b = 0.02$ (95% CI [-0.02, 0.06]); for intelligent, $b = -0.10$ (95% CI [-0.15, -0.05]); for trustworthy, $b = -0.10$ (95% CI [-0.15, -0.05]). See the Markdown file on the OSF page for all output.

Distributional Analyses

As in Study 1, we again plotted the distributions of AUC across levels of congruence to assess evidence for sequential processing or simultaneous integration of information. Replicating the results of Study 1 across all levels of congruence, there was no evidence of bimodality, indicating simultaneous integration of stereotype knowledge and target appearance throughout the impression formation process.

Stereotype Endorsement

Finally, we regressed AUC onto an identical model, but using stereotype endorsement instead of stereotype knowledge. Although knowledge of stereotypes and personal endorsement of those stereotypes share some variance ($R^2 = .22$ in this study), evidence suggests some distinction between the psychological constructs (see Hester et al., 2022, for further evidence). For this reason, we chose to interpret these analyses separately as converging evidence.

Nevertheless, the pattern of results was near-identical to stereotype knowledge. We found a significant interaction between response and stereotype endorsement consistent with our stereotype–feature congruence hypothesis, $b = -0.04$, $t(1295) = -3.91$, $p < .001$, 95% CI [-0.06, -0.02]. When participants responded that a target was "less" of a trait, greater stereotype endorsement was associated with larger AUCs. When participants responded that a target was "more" of a trait, greater stereotype endorsement was associated with smaller AUCs. We observed a similar main effect of response compared with stereotype knowledge, $b = -0.06$, $t(1298) = -5.92$, $p < .001$, 95% CI [-0.08, -0.04].

As with stereotype knowledge, we saw no interaction between trait and the key response × stereotype endorsement effect, $F(2, 1292) = 1.97$, $p = .14$. The trait × response interaction is similar to the one reported for stereotype knowledge. See the Markdown file for all output.

General Discussion

Two studies provide evidence that top-down perceiver knowledge of stereotypes interacts with bottom-up target facial features to influence the process underlying impression formation. For six traits—attractiveness, dominance, friendliness, intelligence, physical strength, and trustworthiness—we found that, to the extent that perceiver stereotype knowledge and target facial features were congruent, mouse trajectories toward perceivers' final impressions of the target were more direct. Conversely, incongruence between stereotype knowledge and facial features resulted in mouse trajectories that showed initial attraction toward the response option not chosen. This result is robust according to a number of sample and methodological variations between Study 1 and 2, and we interpet it as supporting a dynamic interactive model of person construal.

Various models have theorized that stereotypes constrain the process by which both categorizations and impressions are formed. However, empirical tests of impression formation processes are limited in both number and scope, focusing mostly on impressions of trustworthiness. We conducted the broadest examination of impression formation processes to date across trait, target

race, and target gender, and our results provide support for dynamic models of social perception. The work addresses a gap in this literature, given that most research demonstrating perceiver × target interactions over time focuses on categorization, rather than impression formation (Freeman et al., 2020). Because (a) the traits we have examined are theorized to cover the conceptual "space" in which impressions are formed broadly (Oosterhof & Todorov, 2008; Stolier et al., 2018; Sutherland et al., 2013), and (b) we saw little variation as a function of trait, we conclude that our results would likely generalize to other traits for which perceivers form impressions. We also found that these results generalize across stereotype knowledge and stereotype endorsement. This finding suggests that, at least for the process of forming trait impressions, stereotype effects are concerned with whether specific stereotypes are cognitively accessible, rather than the particular reason they are accessible. Future research might further test these possibilities.

Furthermore, the impact of stereotypes on impressions specific to individual traits necessarily requires that the target be partially categorized prior to the formation of the impression. In this way, the present work elaborates on the back-and-forth nature of dynamic interactive models: top-down and bottom-up processes interact dynamically to yield a stable categorization of the target, which facilitates the activation of specific top-down race × gender stereotypes; in turn, these dynamically interact with bottom-up facial features to facilitate a stable impression of the target. Analyses of the distributions of the response trajectories corroborate this argument. Rather than supporting a discrete and sequential processing pipeline, the graded trajectories presented are instead more consistent with partial and parallel activations of stereotype knowledge associated with the morphological characteristics of the faces. These multiple cues influence and constrain one another over time until settling on a final impression (Freeman et al., 2020).

The present work also clarifies why accelerated judgments enable stereotype-guided responses. To the extent that stereotype information influences the trajectory of impression formation early in the process, this stereotype information might exert greater influence on responses that are accelerated, which may constrain the integration of information from other facial features. Notably, the results suggest that mere stereotype *knowledge* is sufficient for early attraction toward the stereotypical impression, such that stereotype endorsement may not be necessary for time pressure to increase the stereotypicality of impressions. To the extent that stereotype knowledge and implicit stereotypes are thought of as separate latent constructs, this pattern of results resists a straightforward dual-process interpretation, in which the automatic or implicit stereotype influences judgments when there is a lack of ability or motivation to exert cognitive control (see Chaiken & Trope, 1999). It might be more accurate to simply characterize the process underlying impressions as a time course in which "stereotype" information shapes the trajectory of judgments earlier in impression formation, and is then integrated with or "corrected" by other facial features, in line with a dynamic interactive modeling approach (Freeman et al., 2020).

The conclusions of the present work are also consistent with other work on the structure of impressions in different contexts and around the world. Here we show that stereotypes influence the process of impression formation for any given trait. Different demographic groups have different stereotypic associations, so we would not expect to find the same factor structure for different groups or in different contexts, as different stereotypes would have an impact on the impressions formed. This is exactly the conclusion of recent work examining impressions formed around the world (Jones et al., 2021) and of different racial groups (Xie et al., 2021).

Limitations

We highlight a few limitations of the present study. First, we note that our sample is drawn from a Canadian undergraduate population in Study 1 and a broader North American population in Study 2, both of which, according to Henrich et al. (2010), are WEIRD (Western, Educated, Industrialized,

Rich, and Democratic). Furthermore, although we compare the process of forming trait impressions with the process of forming social categorizations as a means for justifying the theoretical importance of the present work, we do not empirically compare the two processes. Future work might pursue this direct comparison, although the exact nature of the top-down perceiver questions (e.g., stereotype knowledge versus social category knowledge) is unclear.

Another limitation is the set of traits for which we collected data. These six traits are by no means exhaustive in representing the full universe of traits. However, given the centrality of these traits to processes of face perception (Jones et al., 2021; Oosterhof & Todorov, 2008; Vernon et al., 2014), we find it unlikely that the revealed interaction between top-down stereotypes and bottom-up facial features would be different in nature for other traits. Furthermore, in both Study 1 and 2, we did not find that trait interacted with the stereotype–feature congruence effect. Nevertheless, future research might examine impression formation processes across a broader set of traits.

Conclusion

Psychologists increasingly appreciate the complex and dynamic process underlying final judgments with regard to person perception. This process has been mapped out in some detail for the social categorization of targets; however, we still know little about this process for *trait impressions* of targets, which might be thought of as the "second half" of forming a global perception of a target after categorization occurs. In this way, the present work is a natural extension of existing work on dynamic interactive models of social perception, elaborating on an essential part of the back-and-forth process that ultimately leads to deciding whether someone looks attractive, trustworthy, or friendly.

Acknowledgements

Funding

ORCID iDs

Neil Hester [iD] https://orcid.org/0000-0003-3025-766X
Jeannine Alana Bertin [iD] https://orcid.org/0000-0001-9296-8091

Supplemental material

Supplemental material for this article is available online. Markdowns, materials, and data for this experiment are available at https://osf.io/mgsza. The preregistration for Study 2 is available at https://osf.io/jmfz2.

Notes

1. Social category is a categorical variable with values 1 through 6 (representing the six race-by-gender combinations in the study). Because we assigned each trial a stereotype knowledge with information about race and gender built into it, social category does not need to be represented as a predictor in the regression, but does need to be represented as a nesting variable.
2. The corresponding *lmer* code is as follows: lmer(AUC ~ response*stereotype + (1|participant) + (1|stimulus) + (1|social_category) + (response|participant × social_category), data=data).
3. Due to a programming error, participants' responses to the gender question in the demographics section were not recorded for this study.

References

Ambrus, G. G., Kaiser, D., Cichy, R. M., & Kovács, G. (2019). The neural dynamics of familiar face recognition. *Cerebral Cortex*, *29*, 4775–4784. https://doi.org/ 10.1093/cercor/bhz010

Bjornsdottir, R. T., Hehman, E., & Human, L. J. (2021). Consensus enables accurate social judgments. *Social Psychological and Personality Science*, *13*, 1010–1021. https://doi.org/10.1177/19485506211047095

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of

facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34–42. https://doi.org/10.1016/j.jesp.2018.04.011

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour*, *2*, 581–591. https://doi.org/10.1038/s41562-018-0376-6

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Cassidy, B. S., Boucher, K. L., Lanie, S. T., & Krendl, A. C. (2019). Age effects on trustworthiness activation and trust biases in face perception. *The Journals of Gerontology: Series B*, *74*, 87–92. https://doi.org/10.1093/geronb/gby062

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*, *88*, 885–894. https://doi.org/10.1037/0022-3514.88.6.885

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of Personality and Social Psychology*, *108*, 219–233. https://doi.org/10.1037/pspa0000015

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*, 15–28. https://doi.org/10.3758/BF03195938

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, *21*, 1139–1150. https://doi.org/10.1177/01461672952111002

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*, article 1258. https://doi.org/10.1038/s41467-019-09239-1

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902. https://doi.org/10.1037//0022-3514.82.6.878

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. 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*, 247–279. https://doi.org/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*, 673–690. https://doi.org/10.1037/a0013875

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. https://doi.org/10.1016/j.jesp.2009.10.002

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (*Vol. 61*, pp. 237–287). Elsevier. https://doi.org/10.1016/bs.aesp.2019.09.005

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

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*, 498–506. https://doi.org/10.1177/0956797612457783

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly*, *37*, 113–127. https://doi.org/10.1177/0361684312464203

Gillebaart, M., Schneider, I. K., & Ridder, D. T. D. D. (2016). Effects of trait self-control on response conflict about healthy and unhealthy food. *Journal of Personality*, *84*, 789–798. https://doi.org/10.1111/jopy.12219

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/10.1111/2041-210X.12504

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, *41*, 1123–1134. https://doi.org/10.1177/0146167215591495

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The face-time continuum: Lifespan changes in facial width-to-height ratio impact aging-associated perceptions. *Personality & Social*

Psychology Bulletin, *40*, 1624–1636. https://doi.org/10.1177/0146167214552791

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

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*, e12431. https://doi.org/10.1111/spc3.12431

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*, 513–529. https://doi.org/10.1037/pspa0000090

Heisig, J. P., & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review*, *35*, 258–279. https://doi.org/10.1093/esr/jcy053

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Hester, N., Axt, J. R., Siemers, N., & Hehman, E. (2022). Evaluating validity properties of 25 race-related scales. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01873-w

Hester, N., Xie, S. Y., & Hehman, E. (2021). Little between-region and between-country variance when forming impressions of others. *Psychological Science*, *32*, 1–11. https://doi.org/10.1177/09567976211019950

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 199–209. https://doi.org/10.1037/0096-1523.32.2.199

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*, 640–643. https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology*, *44*, 1418–1420. https://doi.org/10.1016/j.jesp.2008.05.002

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

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 21–36. https://doi.org/10.3758/CABN.5.1.21

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*, 116–131. https://doi.org/10.1037/a0025335

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., . . . Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*, 159–169. https://doi.org/10.1038/s41562-020-01007-2

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology* (*Vol. 55*, pp. 1–80). Elsevier. https://doi.org/10.1016/bs.aesp.2016.10.001

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*, 527–539.

Kubota, J. T., & Ito, T. a. (2014). The role of race and expression in weapons identification. *Emotion*, *14*, 1115–1124. https://doi.org/10.1037/a0038214

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308.

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, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

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

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

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*, 22–24. https://doi.org/10.1111/j.1467-9280.2008.02041.x

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*, 323–342. https://doi.org/10.1037/xge0000638

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092. https://doi.org/10.1073/pnas.0805664105

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192. https://doi.org/10.1037//0022-3514.81.2.181

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Schneider, I. K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., & Koole, S. L. (2015). The path of ambivalence: Tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology*, *6*, article 996. https://doi.org/10.3389/fpsyg.2015.00996

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. https://doi.org/10.1016/j.jesp.2014.09.012

Stillman, P. E., Medvedev, D., & Ferguson, M. J. (2017). Resisting temptation: Tracking how self-control conflicts are successfully resolved in real time. *Psychological Science*, *28*, 1240–1258. https://doi.org/10.1177/0956797617705386

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive

theory. *Trends in Cognitive Sciences*, *22*, 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*, 795–797. https://doi.org/10.1038/nn.4296

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, *115*, 9210–9215. https://doi.org/10.1073/pnas.1807222115

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*, 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science*, *27*, 384–393. https://doi.org/10.1177/0956797615624492

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, *111*, E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1702–1725. https://doi.org/10.1098/rstb.2010.0360

Williams, D. R., & Mohammed, S. A. (2013). Racism and health I: Pathways and scientific evidence. *American Behavioral Scientist*, *57*, 1152–1173. https://doi.org/10.1177/0002764213487340

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*, 59–80. https://doi.org/10.1037/pspi0000092

Wong, Y. J., Owen, J., Tran, K. K., Collins, D. L., & Higgins, C. E. (2012). Asian American male college students' perceptions of people's stereotypes about Asian American men. *Psychology of Men &*

Masculinity, *13*, 75–88. https://doi.org/10.1037/a0022800

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*, 364–385. https://doi.org/10.1037/pspi0000160

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science*, *32*, 1979–1993. https://doi.org/10.1177/09567976211024259

Yee, A. H. (1992). Asians as stereotypes and students: Misperceptions that persist. *Educational Psychology Review*, *4*, 95–132. https://doi.org/10.1007/BF01322396

Zebrowitz, L. A., Kikuchi, M., & Fellous, J.-M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology*, *98*, 175–189. https://doi.org/10.1037/a0017990

Zebrowitz, L. A., & Montepare, J. M. (2005). Appearance DOES matter. *Science*, *308*, m1565–1566. https://doi.org/10.1126/science.1114170

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2018). The time course from gender categorization to gender-stereotype activation. *Social Neuroscience*, *13*, 52–60. https://doi.org/10.1080/17470919.2016.1251965

Zlotnick, E., Dzikiewicz, A. J., & Bar-Anan, Y. (2015). *Minno.js* (Version 0.3) [Computer software].