# AI-Generated Human Stimuli for Experimental Social Science

Chandler Robinson* Ian T. Adams† Matthew Logan‡ Pete Blair‡

Draft as of September 26, 2025

**Abstract**

Social science experiments commonly rely on visual stimuli, yet available images are often scarce, inconsistent, and hard to reproduce. We test whether AI-generated human profiles offer a scalable, valid alternative. In a $4 \times 2 \times 3$ factorial design that varies race, gender, and somatotype, we generate multiple exemplars per condition and evaluate them in a nationally representative U.S. sample ($n = 513$; $n = 2{,}565$ profile ratings). Respondents accurately identified intended traits and reliably detected one-at-a-time manipulations across otherwise identical images. Impression ratings reproduced well-established patterns in social perception, including gender differences in competence versus warmth and body-type effects on perceived status. These results show that AI-generated stimuli both convey intended attributes and recover predictable patterns of social judgment. The approach provides a transparent, reproducible workflow for creating large, systematically varied image sets, with direct implications for studying how the public perceives police officers and, more broadly, for research in psychology, criminology, political science, and other domains that require precise, scalable visual treatments.

**Key Words:** AI-generated images, experimental stimuli, synthetic methods, perceptions of police

*University of Utah. Email: Chandler.Robinson@utah.edu. *Draft is preliminary and incomplete.*

†University of South Carolina

‡Texas State University

# 1 Introduction

Visual experiments across the social and behavioral sciences depend on images that are consistent, transparent, and reproducible. Yet most available photo sets are scarce, idiosyncratic, and difficult to share, which constrains design space and limits cumulative inference.

Recent reviews articulate both the promise and the peril of AI in social science. Grossmann et al. (2023) argues that large language models can simulate human judgments at scale, which accelerates hypothesis testing, but also heightens concerns about transparency and bias management. Bail (2024) likewise contends that generative AI can furnish realistic, scalable stimuli for experiments and surveys, yet cautions about inherited data biases, replication challenges, and unresolved ethical issues. We respond to this agenda by validating AI-generated images of police officers as experimental stimuli. Here we show that synthetic profiles varying race, gender, and body composition pass perceptual validity checks under stimulus sampling, and they elicit theory-consistent patterns on competence, agency, and warmth across 2,565 profile ratings in a U.S. sample. The result is a transparent, reproducible workflow for large, systematically varied image sets.

Across politics, the military, policing, and organizational hierarchies, thin-slice judgments from faces systematically shape who rises to leadership. In elections, rapid impressions of competence from candidate portraits predict outcomes above chance in the United States and across Europe (Todorov et al. 2005; Antonakis and Dalgas 2009; Berggren, Jordahl and Poutvaara 2010; Laustsen 2014; Olivola and Todorov 2010*a*,*b*), and attractiveness or appearance explains vote shares even net of party and incumbency in proportional systems (Berggren, Jordahl and Poutvaara 2017). In hierarchical organizations, classic evidence from West Point graduation portraits shows that facial dominance forecasts rank decades later, controlling for early performance (Mueller and Mazur 1996). Together, these findings establish that embodied visual cues exert replicable effects on leadership selection across institutions.

Policing provides a critical extension of this logic and a stringent test case for adding what prior studies often omit: body composition as a core cue alongside race and gender. Adams et al. (2024) demonstrate that officers' embodiment has concrete institutional consequences. Using archival academy headshots, they show that public judgments of facial traits and perceived leadership predict above chance which cadets later become sergeants and lieutenants. Because those leaders set policy, hiring, discipline, and organizational culture, the study links individual-level impression formation to the reproduction of institutional leadership in one of society's most consequential organizations.

At the same time, experimental work that manipulates officer profiles yields mixed results. Some studies find that Black officers are evaluated more favorably than White officers under similar conditions (Levin and Thomas 1997; Pica et al. 2020), others report the opposite (Castaneda 2018; Ewanation and Maeder 2023), and some find no main effects of officer race (Salerno and Sanchez 2020). More nuanced designs show conditionality: perceptions depend on intersections with professional conduct records (Dunbar, Hanink and Kyle 2024) and, in some contexts, the blue identity of policing may overshadow race (Weitzer 2000). Heterogeneity may reflect contextual moderators and methodological limits, as many studies rely on vignettes or small, idiosyncratic photo sets that constrain consistency, generalizability, and reproducibility.

We build on the leadership-perception literature by addressing the identification and reproducibility problems directly. A critical omission in this comparative literature is body composition. Social and organizational research shows that variation in somatotype systematically biases evaluations of competence, agency, and desirability (Levine and Schweitzer 2015; Pingitore et al. 1994; Roehling 1999; Puhl and Brownell 2006), while formidability cues are central to status allocation and dominance in groups (Blaker and van Vugt 2014; Lukaszewski et al. 2016; Zeng, Cheng and Henrich 2022). We therefore treat body composition as a first-order visual attribute on par with race and gender. We develop and validate AI-generated officer profiles that (i) can be systematically varied to isolate race, gender, and body composition; (ii) pass perceptual validity checks under stimulus sampling; and

(iii) elicit impression patterns consistent with established theories of social perception. This framework moves the field from observational forecasting with archival images to controlled experimentation with transparent, scalable visual stimuli, enabling cumulative social science and clarifying how embodied cues shape public perceptions in policing and, by extension, other sectors of public life.

## 2 Characterizing Officer Profiles in Visual Experiments

Visual stimuli enable experiments to manipulate social cues with high ecological face validity. In policing applications, images can capture subtle, context-specific signals that shape perceptions of professionalism, competence, and accountability. The challenge is measurement: stimuli must convey intended attributes without introducing uncontrolled confounds, and they must be reproducible across studies.

Recent advances in AI image generation (Bail 2024; Bird and Lotfi 2024) make it feasible to produce realistic (Cooke et al. 2025; Park et al. 2024; Epstein et al. 2023; Velásquez-Salamanca, Martín-Pascual and Andreu-Sánchez 2025; Bail 2024), systematically varied human profiles under tight experimental control (Zamudio, Grigsby and Michelsen 2025). Unlike stock or staged photos, synthetic profiles allow researchers to fix nuisance features while manipulating focal traits (e.g., race, gender, body composition). This substantially expands the design space and improves consistency across treatment conditions. Because construct validity cannot be assumed, we explicitly validate that participants (i) recognize intended traits and (ii) detect targeted changes across matched images. In this study we manipulate race, gender, and *body composition*—operationalized as somatotype (ectomorph, mesomorph, endomorph)—as core visual attributes.

**Why include body composition?** Much work on social evaluation focuses on race and gender, but evidence from social psychology shows that body composition systematically shapes judgments of competence, agency, and warmth. Studies document penalties associated with weight and related morphology in evaluative settings (Levine and Schweitzer 2015; Pingitore et al. 1994; Roehling 1999; Puhl and Brownell 2013, 2006), alongside status-linked stereotypes tied to gender and race (Mize 2025; Heiserman 2023; Mize and Manago 2018; Ridgeway 2019; Melamed et al. 2019; Jimeno-Ingrum, Berdahl and Lucero-Wagoner 2009). For example, men are often perceived as more competent than women across contexts (Mize 2025; Heiserman 2023; Ridgeway 2019; Ebert, Steffens and Kroth 2014), and perceived competence varies across racial/ethnic categories (Heiserman 2023). Treating body composition as a first-order cue (rather than a nuisance) aligns the design with this broader evidence.

**Physical formidability as a status cue.** Contemporary social psychology links perceived physical formidability to status allocation and dominance. We use somatotype strictly as a visible morphology for perception[1] (Blaker and van Vugt 2014; Lukaszewski et al. 2016; von Rueden 2014), and documents strong implicit associations between perceived size and rank (Holbrook and Fessler 2015; Duguid and Goncalo 2011; Hwang and Matsumoto 2014; Yap, Mason and Ames 2013; Zanolie et al. 2012). These mechanisms are germane to policing, where physical presence and fitness are salient in public expectations and task demands (Marins, Rombaldi and Del Vecchio 2018). Notably, objective fitness levels of officers do not consistently exceed those of the general public across core domains (Marins, David and Del Vecchio 2019), which heightens the relevance of *perceived* formidability for social judgments.

**Organizational norms and perceptions.** Although police leaders overwhelmingly endorse the importance of fitness and healthy body weight, policy requirements are uneven: in a survey of 401 chiefs, nearly all emphasized fitness, yet a majority of departments lacked maintenance mandates (Wagner et al. 2023). Reported concerns about obesity included functional limits (e.g., stamina, self-protection) and potential loss of public confidence—channels

[1]The criminological tradition has long attended to physique and social evaluation (Sheldon, Stevens and Tucker 1940; Sheldon 1949; Maddan, Walker and Miller 2008; McCandless, Persons III and Roberts 1972).

that map directly onto perceived competence and agency. These organizational beliefs underscore why body composition belongs alongside race and gender in experiments on officer perception.

In sum, AI-generated imagery provides a reproducible means to manipulate race, gender, and body composition while holding other features constant. Our validation tests whether these synthetic profiles transmit intended attributes and whether observers register the manipulations as theoretically meaningful differences.

# 3 Significance of this Study

This study advances experimental social science by validating a scalable method to create systematically varied, reproducible human stimuli. Across politics, organizations, and the military, appearance-based judgments forecast who attains leadership; in policing, similar perceptual processes plausibly shape both public evaluations and the internal reproduction of organizational elites. Yet much evidence to date relies on small, idiosyncratic photo sets or observational designs that limit causal inference and replication. By validating AI-generated profiles that reliably transmit intended attributes (race, gender, body type) and pass manipulation checks, we provide an identification-ready stimulus toolkit that generalizes beyond policing to any domain where visual cues are theoretically central.

Methodologically, the contribution is threefold. First, we establish construct validity for AI-generated human images: participants recognize intended traits and detect targeted changes across matched sets. Second, we implement stimulus sampling (multiple exemplars per cell) to avoid idiosyncratic-image artifacts, improving generalization over stimuli and supporting cumulative science. Third, we document a transparent generation protocol that others can reuse, enabling preregistration of treatments, precise replication, and multi-site pooling—features rarely feasible with archival photographs or bespoke imagery.

Substantively, the framework enables cleaner tests of how embodied cues structure core dimensions of social perception such as competence, agency, and warmth. More broadly, validated synthetic stimuli open avenues for comparative and cross-domain experiments (e.g., politics, workplaces, courts) that isolate visual mechanisms under controlled conditions. In short, we move the field from constrained, non-replicable imagery toward a reproducible, scalable, and causally informative approach to studying how visual information organizes social judgment.

As recent work in Science and PNAS observes, the promise of generative AI for social science lies in producing scalable, realistic stimuli—but only if researchers can ensure validity, transparency, and replicability (Grossmann et al. 2023; Bail 2024; Nature Human Behaviour Editorial Board 2023). This study reflects on this challenge, providing evidence that AI-generated human profiles reliably convey intended traits and elicit predictable patterns of social judgment.

# 4 Methods

## Research Design and Procedure

This study uses original data from an online survey experiment, conducted in September 2025, designed to assess whether participants can accurately identify the intended attributes (race, gender, and body type) of police officer profiles in AI-generated images. The experiment employs a $4 \times 2 \times 3$ factorial design, systematically varying the visual stimuli along race, gender, and body-composition, resulting in 24 unique treatment cells (see Table 1). Following Judd, Westfall and Kenny (2017), our experiment is a CNC design (participants crossed with condition; targets nested within condition; participants and targets crossed). Accordingly, we fit mixed-effects models with random intercepts for targets and random intercepts/slopes for participants, estimating variance components with crossed random effects for participants and targets.

Table 1: Experimental Design Factors

| **Factor** | **Levels** |
|---|---|
| Race | White<br>Black<br>Asian<br>Latino |
| Gender | Male<br>Female |
| Somatotype | Ectomorph (slim, low body fat)<br>Mesomorph (muscular, athletic)<br>Endomorph (larger rounded body shape, high body fat) |

To enhance the robustness of our method, we implemented a stimulus sampling strategy (Wells and Windschitl 1999; Judd, Westfall and Kenny 2012; Monin and Oppenheimer 2014) to mitigate potential idiosyncratic effects tied to any single image. Within each treatment cell (e.g., White male mesomorph), we generated multiple distinct AI officer profiles using structured ChatGPT 4o and eventually ChatGPT 5 prompts. While the prompts consistently specified race, gender, and body type, we introduced minor variations to produce perceptually distinct but trait-consistent exemplars. This approach reduces the likelihood that results are driven by peculiarities of individual images rather than the intended experimental manipulation. We include a generation parameter example to support transparency and facilitate replication (see Table 3).

We recruited 513 participants via Prolific to ensure a diverse and representative U.S.-based sample, and we report the full sample's descriptive results in Table 2. We derived this sample size from a simulation based power analysis. Eligibility criteria included being at least 18 years old, fluent in English, and residing in the United States. Each participant was presented with a randomized subset of the images—ensuring that no respondent viewed multiple exemplars from the same treatment cell—to avoid intra-condition carryover effects. Participants were compensated for their time.

Table 2: Respondent Descriptives

| Variable | Category | N | % |
|---|---|---|---|
| **Gender** | Female | 257 | 50.1 |
| | Male | 254 | 49.5 |
| | Other | 2 | 0.4 |
| **Race/Ethnicity** | American Indian or Alaskan Native | 7 | 1.4 |
| | Asian | 35 | 6.8 |
| | Black or African American | 65 | 12.7 |
| | Hispanic or Latino | 55 | 10.7 |
| | Middle Eastern or North African | 3 | 0.6 |
| | Native Hawaiian or Pacific Islander | 1 | 0.2 |
| | Other | 18 | 3.5 |
| | White | 329 | 64.1 |
| **Partisanship** | Democrat | 196 | 38.2 |
| | Independent | 159 | 31.0 |
| | Republican | 158 | 30.8 |
| **Education** | Associate's degree (two-year) | 66 | 12.9 |
| | Bachelor's degree (four-year) | 199 | 38.8 |
| | Doctoral degree (PhD) | 6 | 1.2 |
| | High school graduate (highschool diploma or GED) | 72 | 14.0 |
| | Less than high school degree | 2 | 0.4 |
| | Master's degree | 57 | 11.1 |
| | Professional degree (JD, MD, etc.) | 8 | 1.6 |
| | Some college, but no degree | 103 | 20.1 |
| **Age** | 18-29 | 97 | 18.9 |
| | 30-44 | 142 | 27.7 |
| | 45-64 | 210 | 40.9 |
| | 65 and over | 64 | 12.5 |
| **Income** | $100,000-$149,999 | 78 | 15.2 |
| | $20,000-$39,999 | 89 | 17.3 |
| | $40,000-$59,999 | 103 | 20.1 |
| | $60,000-$79,999 | 75 | 14.6 |
| | $80,000-$99,999 | 69 | 13.5 |
| | Less than $20,000 | 51 | 9.9 |
| | More than $150,000 | 48 | 9.4 |

## Image Generation and Trait Manipulation

The AI-generated officer profiles were designed to closely resemble real-world police headshots in uniform, background, and lighting. We used precise LLM prompts to create a specific combination of race, gender, and body type which we adapt from Sheldon, Stevens and Tucker (1940)'s concept of somatotype, with image quality and realism verified through internal checks before inclusion[2]

Table 3: Generic image-generation prompt with brackets marking manipulated traits.

| **Example Image-Generation Prompt** |
| --- |
| Produce a professional police officer headshot photograph (head to waist) for experimental research. The officer is [RACE], [GENDER], and has a [BODY TYPE] build. The officer wears a standard U.S. police uniform with neutral facial expression, plain background, even lighting, and realistic photographic style. No accessories, no text, no background objects. |

Image 1: White Male Ectomorph

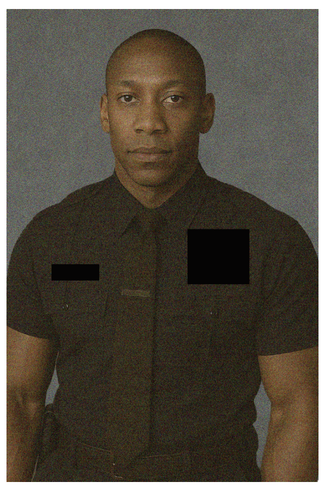


Image 2: Black Male Mesomorph

Image 3: Latino Male Endomorph

[2]Minor adjustments were sometimes necessary after image generation. For example, when an image did not properly reflect the intended trait, we applied a reprompt to correct it. In addition, all backgrounds were standardized using photo-editing software, and badges or nameplates were covered to reduce potential sources of bias.

## Observable Traits Coding (Task 1)

To validate whether the generated images reliably conveyed the intended traits, participants were asked to classify a random subset of the officer images across the following dimensions:

- **Race Identification:** "What race does this officer appear to be?" (White, Black, Asian, Hispanic)
- **Gender Identification:** "What gender does this officer appear to be?" (Male, Female)
- **Somatotype Identification:** "How would you describe this officer's body type?" (Ectomorph, Mesomorph, Endomorph)

Responses were recorded using a standardized multiple-choice format. Each image was independently rated by multiple participants to allow for assessment of inter-rater consistency. We computed inter-rater reliability using Krippendorf's $\alpha$ and percentage agreement.

To measure confoundedness across trait dimensions, we conducted a Principal Components Analysis (PCA) with direct oblimin rotation on the trait responses (Adams et al. 2024). This analysis tested the empirical independence of race, gender, and somatotype classifications (see A.3.2).

### Trait Recognition Hypotheses (Task 1)

- **H1a**: Participants will accurately identify the race of AI-generated officer profiles at rates significantly above chance.
- **H1b**: Participants will accurately identify the gender of AI-generated officer profiles at rates significantly above chance.
- **H1c**: Participants will accurately identify the body type (somatotype) of AI-generated officer profiles at rates significantly above chance.

## Observable Traits Change (Task 2)

The second task asked participants to compare sets of officer images presented side by side. Within each set, one officer differed on a single manipulated trait while all other features remained constant. Respondents were asked to identify which trait differed, with options including race or ethnicity, gender, body type, facial expression, uniform, or age. This task provided a direct test of whether participants could perceive intended manipulations across otherwise identical profiles. Importantly, the number of observations (N) is not constant across comparison types because the factorial design generated unequal numbers of available image sets for each trait. For example, there were more unique race and body type contrasts than gender contrasts, which produced a higher N for those dimensions. In addition, random assignment of image sets to participants meant that not every respondent evaluated every type of change, further contributing to variation in the total number of observations by trait.

### Trait Change Hypotheses (Task 2)

- **H2a**: Participants will accurately identify changes in officer race or ethnicity at rates significantly above chance.
- **H2b**: Participants will accurately identify changes in officer gender at rates significantly above chance.
- **H2c**: Participants will accurately identify changes in officer body type (somatotype) at rates significantly above chance.

## Officer Impressions (Task 3)

Building on Carrier et al. (2014); Brambilla et al. (2011), who disentangled agency from competence by restricting agency to its motivational component, we adopted validated short scales for competency, agency, and warmth. In the final task, respondents evaluated each officer on a set of impression measures that capture three fundamental dimensions of social judgment: agency, competence, and warmth. Respondents rated officers on nine trait adjectives using a seven-point scale ranging from 1 (Not at all) to 7 (Extremely). Competence was measured with competent, capable, and efficient. Agency was measured with ambitious, self-confident, and assertive. Warmth was measured with warm, friendly, and likable. This structure allows us to separately assess whether officers convey the motivation to advance the self (agency), the ability to effectively pursue goals (competence), and concern for others (warmth).

We independently assessed the internal consistency of the three impression scales using McDonald's $\omega$, which provides a more general reliability estimate than Cronbach's $\alpha$ (McNeish 2018). Reliability was excellent across all three scales: competence ($\omega$ = .95), agency ($\omega$ = .90), and warmth ($\omega$ = .96). These values indicate that the items within each scale cohered strongly and justify treating them as composite measures of the intended constructs.

Table 4: Impression Rating Descriptives

| Scale | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Competence | 2565 | 4.98 | 1.29 | 1 | 7 |
| Agency | 2565 | 4.88 | 1.30 | 1 | 7 |
| Warmth | 2565 | 4.35 | 1.40 | 1 | 7 |

**Impressions Hypotheses (Task 3)**

- **H3a**: Participants will rate male officers as more competent and agentic, but less warm than female officers (Mize 2025; Mize and Manago 2018; Ebert, Steffens and Kroth 2014)

- **H3b**: Participants will rate endomorphic officers as less competent and agentic than ectomorphic officers (Levine and Schweitzer 2015; Puhl and Brownell 2013, 2006; Roehling 1999), but there will be no effect for warmth (Levine and Schweitzer 2015)

- **H3c**: Participants will rate mesomorphic officers as more competent, agentic, and warm than ectomorphic officers. (Lukaszewski et al. 2016; Zeng, Cheng and Henrich 2022)

**H3d**: Partisipants will rate White and Asian officers as more competent than latino or black officers (Heiserman 2023), but White officers will be rated as less warm (Heiserman 2023; Levin and Thomas 1997; Dunbar, Hanink and Kyle 2024; Pica et al. 2020)

# 5 Results

## 5.1 Observable Traits Agreement

The first set of analyzes tested the trait recognition hypotheses (H1a-H1c). Table 5 reports the proportion of participants who correctly classified each officer profile's race, gender, and body type, disaggregated by the true[3] gender[4] of the officer image. Across all traits, agreement between the intended (true) characteristics and participant classifications was consistently high, with by-chance tests indicating performance significantly above random guessing (all $p < 0.001$).

Participants were especially accurate in identifying officer gender, with agreement rates exceeding 92% for female profiles and 99% for male profiles. This strong performance provides clear support for H1b, demonstrating that the AI-generated profiles reliably conveyed gender. Similarly, officer body type was accurately perceived, with agreement rates of 81.7% for female officers and 88.6% for male officers, strongly supporting H1c. Finally, classification of race was also well above chance, with agreement rates of 86.8% for female officers and 94.2% for male officers, consistent with H1a. Here, we demonstrate that the AI-generated

[3] "True" characteristics are determined by the LLM prompt used to generate the image

[4] See Table A.4.2 for accuracy by officer race and Table A.4.3 for accuracy by officer body type

officer profiles conveyed the intended demographic and somatotype characteristics with high fidelity. Additionally, these findings provide the necessary foundation for subsequent tasks: assessing differentiation across manipulated traits (Task 2) and evaluating impressions of competence, agency, and warmth (Task 3). Additionally, we find that results slightly improve upon attention check implementation (see Table A.4.1).

In addition to percent agreement, we assessed inter-rater reliability using Krippendorff's $\alpha$, which accounts for chance agreement, multiple raters, and missing data (Krippendorff 2019). Reliability was high for gender ($\alpha$ = .87), strong for race ($\alpha$ = .82), and moderate for body type ($\alpha$ = .72). These values indicate that respondents not only performed above chance but also reached satisfactory levels of consistency in their classifications, particularly for gender and race. According to Krippendorff (2019), our $\alpha$ for body type is within the range for the lower bound for tentative conclusions.

Table 5: Agreement Between True and Perceived Traits

| Trait | N | Correct | Percent Agreement | By-chance p | Krippendorff's $\alpha$ |
|---|---|---|---|---|---|
| **Female Officers** | | | | | |
| Gender | 1326 | 1230 | 92.8 | 0 | |
| Body Type | 1326 | 1082 | 81.6 | 0 | |
| Race | 1326 | 1151 | 86.8 | 0 | |
| **Male Officers** | | | | | |
| Gender | 1239 | 1233 | 99.5 | 0 | |
| Body Type | 1239 | 1099 | 88.7 | 0 | |
| Race | 1239 | 1168 | 94.3 | 0 | |
| **Overall** | | | | | |
| Gender | 2565 | 2463 | 96.0 | 0 | 0.872 |
| Body Type | 2565 | 2181 | 85.0 | 0 | 0.718 |
| Race | 2565 | 2319 | 90.4 | 0 | 0.816 |

**Mixed-Effects Models of Trait Recognition** To further evaluate the perceptual validity of the AI-generated officer profiles, we estimated mixed-effects logistic regression models (Judd, Westfall and Kenny 2017) predicting correct identification of gender, race, and body type (Table A.4.6). Each model included fixed effects for the officer's true race, gender, and body type, with random intercepts for both respondents and profiles.

Baseline accuracy was very high across outcomes (intercepts: gender = 16.149, $p < 0.001$; race = 4.883, $p < 0.001$; body type = 3.018, $p < 0.001$). Female officers were significantly less likely to be correctly identified than male officers for gender ( = -4.447, $p < 0.001$) and race ( = -0.989, $p = 0.008$), with no reliable difference for body type ( = -0.569, $p = 0.150$). For race recognition, respondents were more accurate for Black officers ( = 1.903, $p = 0.007$), but less accurate for Asian officers ( = -1.538, $p = 0.002$) and Latino officers ( = -1.798, $p < 0.001$). For body-type recognition, mesomorphic officers were less likely to be correctly recognized than ectomorphs ( = -1.298, $p = 0.006$), whereas endomorphic officers were more likely ( = 1.088, $p = 0.028$).

## 5.2 Trait Change Agreement

The second set of analyses tested the trait-change recognition hypotheses (H2a - H2c). Table 6 reports results from the trait change task (task 2), where participants were presented with sets of officer profiles in which only one characteristic (race, gender, or body type) was systematically varied. Respondents were asked to identify which trait differed across the images. Across all categories, accuracy was substantially higher than chance, with by-chance tests confirming reliable detection of intended manipulations (all $p < 0.001$).

Notably, participants were comparably effective at detecting changes across the three focal traits. Agreement rates were 71.1% for body type, 71.4% for gender, and 70.9% for race or ethnicity. This shows clear support for the differentiation component of our validation, demonstrating that respondents could reliably perceive when an officer's profile differed on a targeted attribute.

Extending the evidence from Section 5.1. Not only did participants accurately recognize the intended traits of individual images, they also consistently detected manipulated changes across otherwise identical profiles. This reinforces the conclusion that the AI-generated officer profiles systematically convey the intended demographic and physical characteristics and that participants perceive these manipulations as meaningful and distinct. In other words, we show that AI generated images can precisely generate certain traits while controlling for others. Again, we find that results are robust and slightly improve when we implement attention check analyses (see Table A.4.4)

Although Krippendorff's $\alpha$ is a useful index of inter-rater reliability in many content-analysis settings, it is not the best measure for the change-detection task in Study 2. This task differs fundamentally from a coding exercise in which multiple raters independently classify the same units. Instead, each respondent was presented with a unique set of triads and asked to identify which trait changed. Because the task has an objectively correct answer, accuracy relative to chance is the more direct test of our hypotheses. In this context, $\alpha$ is misleading: it penalizes variability in the specific incorrect categories chosen by respondents, even when overall accuracy is well above chance. Thus, while $\alpha$ estimates were low ( .23), this does not indicate poor construct validity. Rather, it reflects that when participants erred, they did not consistently select the same wrong category. For this reason, we rely on percent agreement relative to chance as the most appropriate indicator of performance in this task.

Table 6: Agreement Between True Change and Perceived Change

| Trait Change | N | Correct | Respondent Accuracy | By-chance p |
|---|---|---|---|---|
| Gender | 311 | 222 | 71.4% | 0 |
| Body Type | 1320 | 938 | 71.1% | 0 |
| Race or ethnicity | 934 | 662 | 70.9% | 0 |

## 5.3 Impressions: Main Effects

The third set of analyses tested the impressions hypotheses (H3a–H3d) by examining whether respondents evaluated officers differently on competence, agency, and warmth as a function of race, gender, and body type. Table 7 presents the main effects from respondent and image random-effects models, and Figures 1 and 2 display conditional means and average effect sizes for each trait dimension.

**Gender.** Consistent with the direction of H3a but not statistically significant, female officers were rated lower than male officers on competence ($-0.038$, n.s.) and agency ($-0.108$, n.s.). At the same time, female officers were evaluated as warmer (+0.173, $p < 0.01$). The pattern aligns with established gender stereotyping: female officers are perceived as warmer and, directionally, slightly less competent and agentic, indicating that AI-generated profiles elicit theoretically coherent evaluations even though only warmth reached significance.

**Race.** Relative to White officers (reference), Black and Latino officers did not differ significantly on competence or agency (competence: +0.067 and +0.051; agency: +0.113 and +0.038; all n.s.). Asian officers were also similar to Whites on competence and agency (competence: +0.156, n.s.; agency: 0.012, n.s.). In contrast, warmth ratings were consistently higher for minority officers: Black (+0.272, $p < .05$), Latino (+0.220, $p < .05$), and Asian (+0.312, $p < .01$). These results provide support for the warmth component of H3d, while competence and agency show no reliable racial differences.

**Body Type.** Body type produced some of the largest and most consistent effects across outcomes, in line with H3b and H3c. Relative to ectomorphic officers (the reference category), mesomorphic officers were rated substantially higher on competence (+0.376, $p < .001$), agency (+0.555, $p < .001$), and warmth (+0.215, $p < .05$). In contrast, endomorphic officers were penalized on competence (-0.625, $p < .001$) and agency (-0.608, $p < .001$), but showed no significant difference in warmth (-0.037, n.s.). These results replicate long-standing stereotypes that reward muscular body types with perceptions of capability and motivation, while heavier builds face sharp penalties on status-linked traits. Importantly, warmth evaluations were largely unaffected by endomorphic profiles, highlighting that penalties were specific to competence and agency rather than extending to interpersonal likability.

**Overall Patterns.** Figure 1 displays conditional means across officer categories, highlighting the strong role of body type in shaping evaluations. Mesomorphic officers consistently received the most favorable ratings, while endomorphic officers faced the steepest penalties on competence and agency. In contrast, racial differences were modest on competence and agency but pronounced on warmth, with all minority groups rated significantly warmer than White officers. Gender effects followed a similar pattern: women were rated warmer, with small and nonsignificant disadvantages on competence and agency. Figure 2 summarizes these patterns by effect size, showing that body type exerted the largest influence on competence and agency, whereas race and gender were more consequential for warmth. Together, these results demonstrate that respondents systematically AI produced officer profiles along theoretically coherent dimensions, with body type driving perceptions of capability and race/gender shaping interpersonal judgments. Notably, all results are robust to attention check analyses (see Table A.4.7).

Table 7: Key Outcomes

| | Competence | Agency | Warmth |
|---|---|---|---|
| Intercept | 5.017*** | 4.917*** | 3.999*** |
| | (0.093) | (0.102) | (0.111) |
| Black officer | 0.067 | 0.113 | 0.272* |
| | (0.091) | (0.103) | (0.109) |
| Latino officer | 0.051 | 0.038 | 0.220* |
| | (0.091) | (0.103) | (0.109) |
| Asian officer | 0.156 | −0.012 | 0.312** |
| | (0.090) | (0.103) | (0.109) |
| Female officer | −0.038 | −0.108 | 0.173* |
| | (0.064) | (0.073) | (0.077) |
| Mesomorph | 0.376*** | 0.555*** | 0.215* |
| | (0.079) | (0.089) | (0.094) |
| Endomorph | −0.625*** | −0.608*** | −0.037 |
| | (0.078) | (0.089) | (0.094) |
| Num.Obs. | 2565 | 2565 | 2565 |
| R2 Marg. | 0.105 | 0.134 | 0.017 |
| R2 Cond. | 0.617 | 0.572 | 0.595 |
| ICC | 0.6 | 0.5 | 0.6 |
| RMSE | 0.72 | 0.77 | 0.80 |
| Random intercept: respondent | Yes | Yes | Yes |
| Random intercept: profile | Yes | Yes | Yes |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Reference categories: White, Male, Ectomorph. Models include random intercepts for respondent and profile.

**Conditional Means**

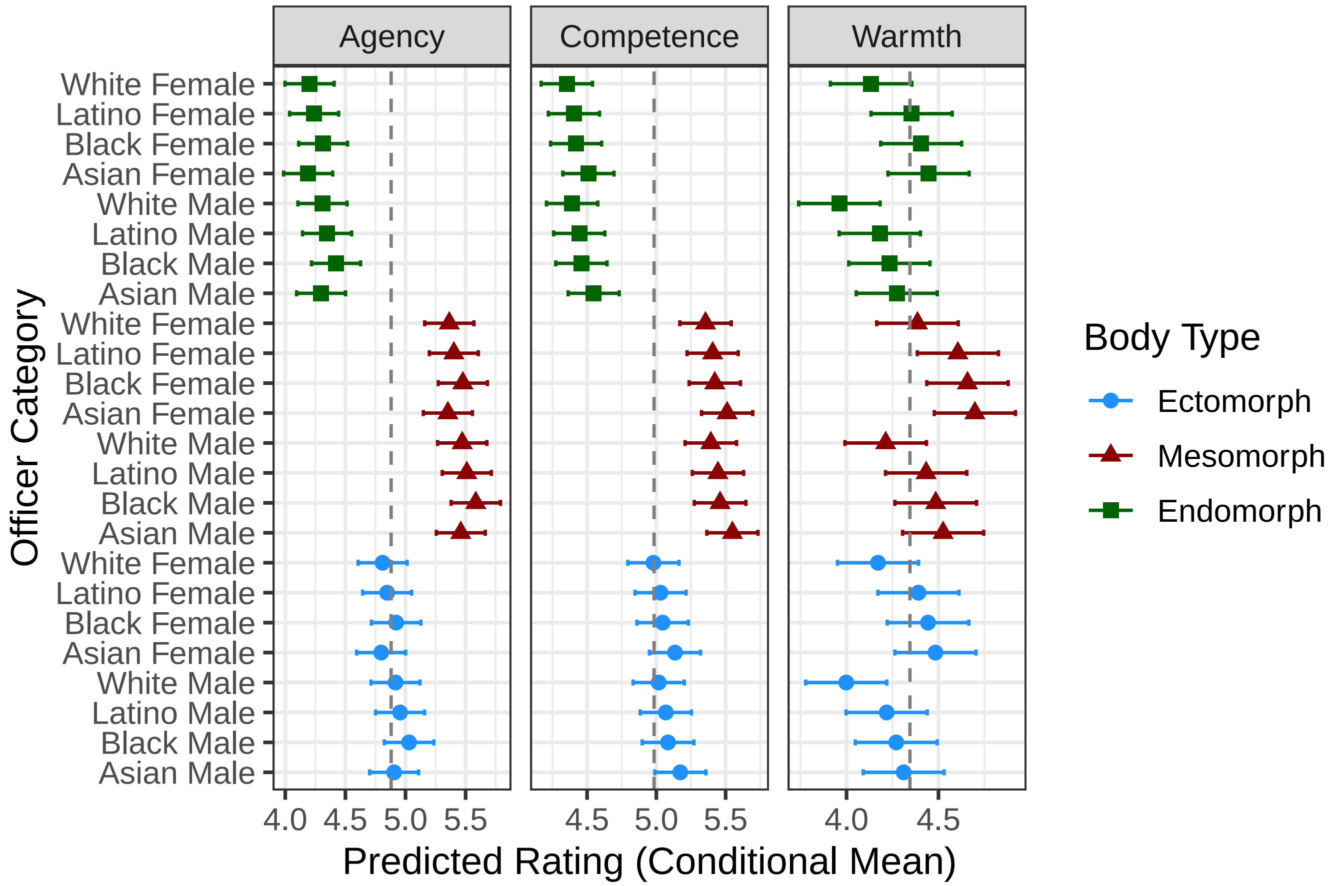



Figure 1

**Marginal Effects**

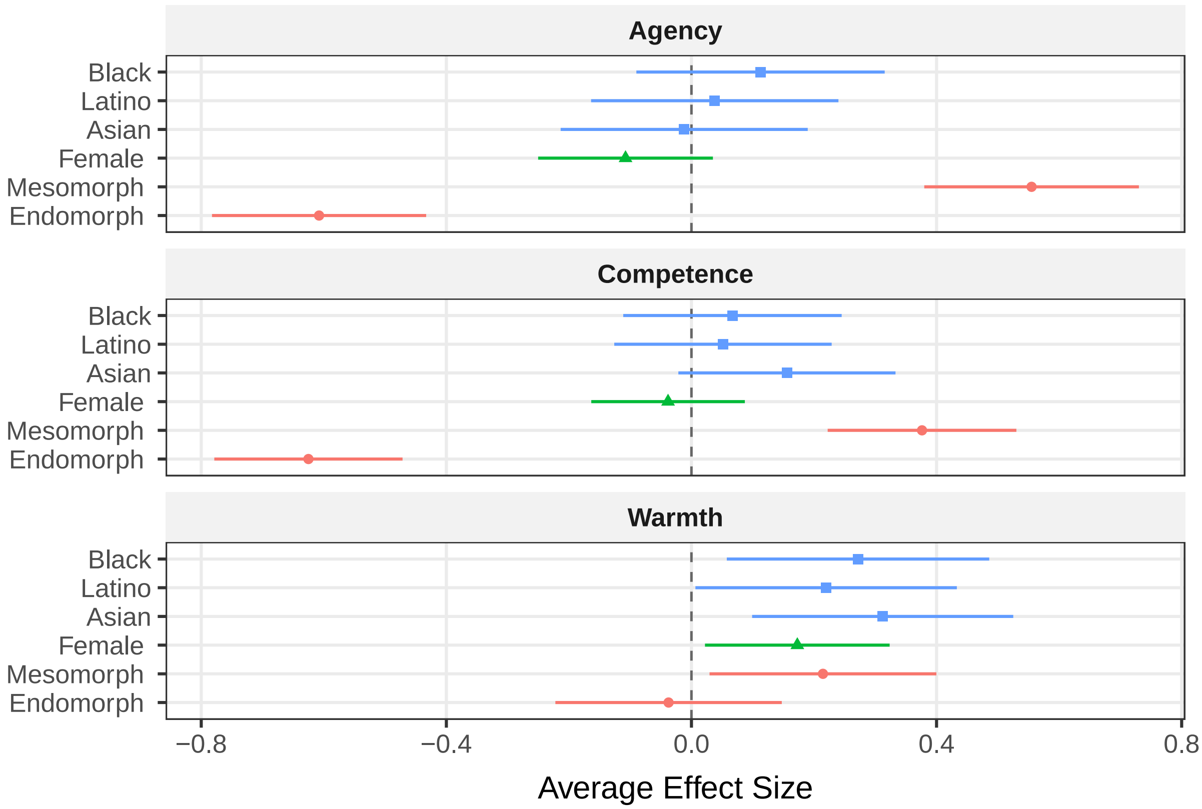



Figure 2

# 6 Limitations and Future Research

Although the study demonstrates the value of AI-generated officer profiles, several limitations should be recognized. Most importantly, the study relied on a U.S.-based sample, which may limit the generalizability of results to other cultural or national contexts where perceptions of police differ, or where different normative ideals of sex, race, and body types may hold. Additionally, the current design focused on three core attributes (race, gender, and body type) while other potentially important characteristics such as age, facial expression, or uniform details remain untested. While our design is robust to unobserved confounders, those same confounds may provide additional explanatory power and can likely serve as additional focal concerns in work building on this method.

Future research can build on this study in several ways. First, expanding the set of validated traits to include age, facial expression, rank, and uniform features would provide an even broader toolkit for studying perceptions of police. Second, extending validation to dynamic formats such as AI-generated video could allow researchers to examine how traits interact with movement, speech, or behavior. Third, combining officer profiles with situational contexts such as narratives of encounters or trial transcripts would allow for richer tests of how individual and contextual cues interact in shaping perceptions of law enforcement. Finally, replication across diverse populations and cross-national settings would strengthen confidence in the generalizability of these findings and support comparative studies of policing and social perception.

# 7 Conclusion

This study addresses two linked problems in experimental social science: the scarcity of reproducible visual stimuli and the systematic neglect of *body composition* as a first-order visual cue. Building on evidence that appearance-based judgments shape leadership selection across institutions (Todorov et al. 2005; Antonakis and Dalgas 2009; Berggren, Jordahl and Poutvaara 2010; Laustsen 2014; Olivola and Todorov 2010*a*,*b*; Mueller and Mazur 1996) and the policing application that connects thin-slice facial judgments to real promotion outcomes (Adams et al. 2024), we develop and validate AI-generated officer profiles that permit controlled manipulation of race, gender, and somatotype.

Empirically, three results matter. First, the synthetic profiles exhibit strong construct validity: observers reliably identify intended traits. Second, observers detect targeted changes across matched images, indicating that manipulations register as theoretically meaningful differences rather than idiosyncratic artifacts. Third, impressions along competence, agency, and warmth reproduce well-established patterns from social perception and status research, including gender differences and sizable penalties/advantages tied to body composition (Mize

2025; Mize and Manago 2018; Ebert, Steffens and Kroth 2014; Levine and Schweitzer 2015; Puhl and Brownell 2013, 2006; Roehling 1999; Lukaszewski et al. 2016; Zeng, Cheng and Henrich 2022). Together with stimulus sampling (multiple exemplars per cell), these findings show that AI-generated imagery can deliver identification-ready treatments while generalizing over stimuli.

Substantively, treating body composition as co-equal with race and gender clarifies a mechanism often sidelined in prior work: cues of weight and formidability function as status signals with downstream implications for who is seen as competent, agentic, and ultimately leadership material (Blaker and van Vugt 2014; Lukaszewski et al. 2016; Zeng, Cheng and Henrich 2022). In policing, where leaders set policy, hiring, and discipline, such perceptual inputs can scale into institutional consequences (Adams et al. 2024). Methodologically, the contribution is a transparent, reusable protocol that shifts the field from observational forecasting with archival photos to causal tests using reproducible, shareable visual treatments.

The practical takeaways are straightforward: (i) validated synthetic images let researchers isolate visual mechanisms with high control; (ii) body composition should be modeled as a core attribute; and (iii) stimulus sampling is essential for credible generalization beyond any single image. Ultimately, the approach is portable to other domains where previous work has demonstrated visual judgments matter, including elections, workplaces, courts. In sum, validated, systematically varied synthetic profiles provide a scalable foundation for experiments on how visual information organizes social judgment across institutions.

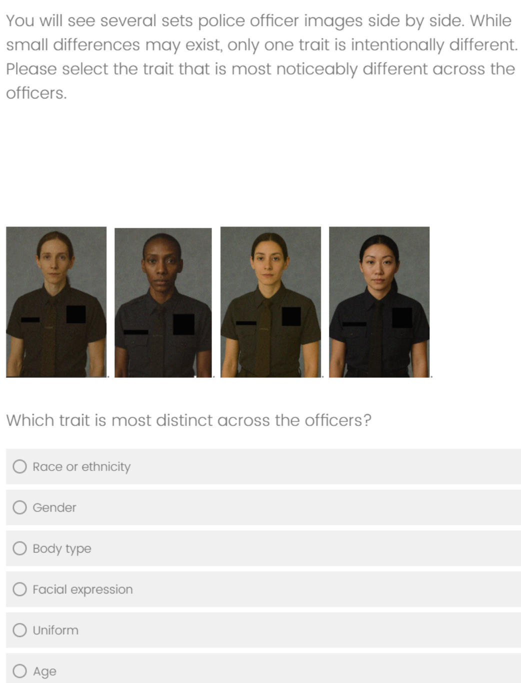
# Appendix Items

## A.1 Task 1

Principal components analysis with direct oblimin rotation produced a three-factor solution. Race loaded cleanly on a single factor (loading = .91), while gender and body type showed moderate cross-loadings. This indicates that the two traits were generally distinguishable but not fully independent: when respondents misclassified one, they were somewhat more likely to misclassify the other.

## A.2 Task 2

**Example: Task 2**



You will see several sets police officer images side by side. While small differences may exist, only one trait is intentionally different. Please select the trait that is most noticeably different across the officers.

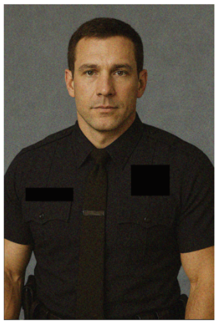
Which trait is most distinct across the officers?

- ○ Race or ethnicity
- ○ Gender
- ○ Body type
- ○ Facial expression
- ○ Uniform
- ○ Age

Figure A.2.1

## A.3 Task 3

### Example: Task 3



Based only on this image, how much do you think each of the following words describes this officer?

| | 1 (Not at all) | 2 | 3 | 4 | 5 | 6 | 7 (Extremely) |
|---|---|---|---|---|---|---|---|
| Competent | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Capable | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Efficient | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Ambitious | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Self-Confident | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Assertive | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Warm | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Friendly | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Likeable | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Figure A.3.2

## A.4 Results

### A.4.1 Observable Traits Agreement

Table A.4.1: Agreement Between True and Perceived Traits: Attention Check

| Trait | N | Correct | Percent Agreement | By-chance p | Krippendorff's $\alpha$ |
|---|---|---|---|---|---|
| **Female Officers** | | | | | |
| Gender | 1267 | 1177 | 92.9 | 0 | |
| Body Type | 1267 | 1036 | 81.8 | 0 | |
| Race | 1267 | 1103 | 87.1 | 0 | |
| **Male Officers** | | | | | |
| Gender | 1178 | 1172 | 99.5 | 0 | |
| Body Type | 1178 | 1050 | 89.1 | 0 | |
| Race | 1178 | 1111 | 94.3 | 0 | |
| **Overall** | | | | | |
| Gender | 2445 | 2349 | 96.1 | 0 | 0.874 |
| Body Type | 2445 | 2086 | 85.3 | 0 | 0.726 |
| Race | 2445 | 2214 | 90.6 | 0 | 0.820 |

Table A.4.2: Agreement Between True and Perceived Traits: By Race

| Officer Race | N | Gender Accuracy | Race Accuracy | Body Type Accuracy |
|---|---|---|---|---|
| White | 703 | 96.7% | 95.3% | 84.6% |
| Black | 651 | 94.6% | 99.1% | 90.2% |
| Asian | 632 | 97.6% | 81.8% | 81.0% |
| Latino | 579 | 95.0% | 84.1% | 84.1% |

Table A.4.3: Agreement Between True and Perceived Traits: By Body Type

| Officer Body Type | N | Gender Accuracy | Race Accuracy | Body Type Accuracy |
|---|---|---|---|---|
| Ectomorph | 863 | 96.6% | 91.8% | 87.3% |
| Mesomorph | 823 | 96.4% | 92.8% | 72.1% |
| Endomorph | 879 | 95.1% | 86.8% | 95.0% |

### A.4.2 Trait Change Agreement

Table A.4.4

Table A.4.5

| Trait Change | N | Correct | Respondent Accuracy | By-chance p |
|---|---|---|---|---|
| Race or ethnicity | 889 | 640 | 72.0% | 0 |
| Body Type | 1261 | 907 | 71.9% | 0 |
| Gender | 295 | 212 | 71.9% | 0 |

### A.4.3 Impressions

Table A.4.6: Recognition Hererogeneity

| | Gender recognition | Race recognition | Body type recognition |
|---|---|---|---|
| Intercept | 16.149*** | 4.883*** | 3.018*** |
| | (2.186) | (0.560) | (0.517) |
| Female | −4.447*** | −0.989** | −0.569 |
| | (1.050) | (0.379) | (0.392) |
| Black | −0.546 | 1.903** | 0.296 |
| | (1.201) | (0.680) | (0.550) |
| Asian | 0.998 | −1.538** | −0.069 |
| | (1.303) | (0.504) | (0.555) |
| Latino | −0.327 | −1.798*** | −0.195 |
| | (1.293) | (0.503) | (0.546) |
| Mesomorph | −0.961 | 0.031 | −1.298** |
| | (1.126) | (0.471) | (0.466) |
| Endomorph | −1.256 | −0.747 | 1.088* |
| | (1.143) | (0.455) | (0.501) |
| Num.Obs. | 2565 | 2565 | 2565 |
| R2 Marg. | 0.067 | 0.303 | 0.157 |
| R2 Cond. | 0.960 | 0.616 | 0.517 |
| ICC | 1.0 | 0.4 | 0.4 |
| RMSE | 0.09 | 0.22 | 0.28 |
| Random intercept: respondent | Yes | Yes | Yes |
| Random intercept: profile | Yes | Yes | Yes |

* p <0.05, ** p <0.01, *** p <0.001
Reference categories: White, Male, Ectomorph. Models include random intercepts for respondent and profile.

Table A.4.7: Key Outcomes: Attention Check

| | Competence | Agency | Warmth |
|---|---|---|---|
| Intercept | 5.010*** | 4.920*** | 3.981*** |
| | (0.096) | (0.104) | (0.113) |
| Black officer | 0.080 | 0.128 | 0.254* |
| | (0.094) | (0.105) | (0.111) |
| Latino officer | 0.051 | 0.041 | 0.200 |
| | (0.093) | (0.104) | (0.111) |
| Asian officer | 0.174 | 0.004 | 0.288** |
| | (0.093) | (0.104) | (0.111) |
| Female officer | −0.034 | −0.106 | 0.171* |
| | (0.066) | (0.074) | (0.078) |
| Mesomorph | 0.371*** | 0.540*** | 0.239* |
| | (0.081) | (0.091) | (0.096) |
| Endomorph | −0.651*** | −0.628*** | −0.019 |
| | (0.081) | (0.090) | (0.096) |
| Num.Obs. | 2445 | 2445 | 2445 |
| R2 Marg. | 0.108 | 0.134 | 0.017 |
| R2 Cond. | 0.622 | 0.577 | 0.599 |
| ICC | 0.6 | 0.5 | 0.6 |
| RMSE | 0.72 | 0.77 | 0.80 |
| Random intercept: respondent | Yes | Yes | Yes |
| Random intercept: profile | Yes | Yes | Yes |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Reference categories: White, Male, Ectomorph. Models include random intercepts for respondent and profile.

# References


Adams, Ian T, Scott M Mourtgos, Christopher A Simon and Nicholas P Lovrich. 2024. "If the face fits: predicting future promotions from police cadets' facial traits." *Journal of experimental criminology* 20(3):937–972.

Antonakis, John and Olaf Dalgas. 2009. "Predicting elections: Child's play!" *Science* 323(5918):1183–1183.

Bail, Christopher A. 2024. "Can Generative AI improve social science?" *Proceedings of the National Academy of Sciences* 121(21):e2314021121.

Berggren, Niclas, Henrik Jordahl and Panu Poutvaara. 2010. "The looks of a winner: Beauty and electoral success." *Journal of public economics* 94(1-2):8–15.

Berggren, Niclas, Henrik Jordahl and Panu Poutvaara. 2017. "The right look: Conservative politicians look better and voters reward it." *Journal of Public Economics* 146:79–86.

Bird, Jordan J. and Ahmad Lotfi. 2024. "CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images." *IEEE Access* 12:15642–15650.

Blaker, Nancy M. and Mark van Vugt. 2014. The Status-Size Hypothesis: How Cues of Physical Size and Social Status Influence Each Other. In *The Psychology of Social Status*, ed. Joey Cheng, Jessica L. Tracy and Cameron Anderson. New York, NY: Springer pp. 119–137.

Brambilla, Marco, Patrice Rusconi, Simona Sacchi and Paolo Cherubini. 2011. "Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering." *European Journal of Social Psychology* 41(2):135–143.

Carrier, Antonin, Eva Louvet, Bruno Chauvin and Odile Rohmer. 2014. "The Primacy of Agency over Competence in Status Perception." *Social Psychology* 45(5):347–356.

Castaneda, G. 2018. Public Perception of Police Behavior During a Traffic Stop Doctoral dissertation Alliant International University. Publication no. 10935723.

Cooke, Di, Abigail Edwards, Sophia Barkoff and Kathryn Kelly. 2025. "As Good As A Coin Toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli.". **URL:** *https://arxiv.org/abs/2403.16760*

Duguid, Michelle M. and Jack A. Goncalo. 2011. "Living Large: The Powerful Overestimate Their Own Height." *Psychological Science* 23(1):36–40. Original work published 2012. **URL:** *https://doi.org/10.1177/0956797611422915*

Dunbar, A., P. A. Hanink and B. Kyle. 2024. "Good Cop, Bad Cop, Black Cop, White Cop: Unraveling Public Perceptions of Police Use of Force." *Crime & Delinquency* pp. 1–23.

Ebert, Irena D, Melanie C Steffens and Alexandra Kroth. 2014. "Warm, but maybe not so competent?—Contemporary implicit stereotypes of women and men in Germany." *Sex roles* 70(9):359–375.

Epstein, David C., Ishan Jain, Oliver Wang and Richard Zhang. 2023. Online Detection of AI-Generated Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops.* pp. 382–392.

Ewanation, L. and E. M. Maeder. 2023. "The influence of race on jurors' perceptions of lethal police use of force." *Law and Human Behavior* 47(1):53–67.

Grossmann, Igor, Matthew Feinberg, Dawn C. Parker, Nicholas Christakis, Philip E. Tetlock and William A. Cunningham. 2023. "AI and the transformation of social science research." *Science* 380(6650):1108–1109.

Heiserman, Nicholas. 2023. "Intersectional Complexity in Stereotype Content." *American Journal of Sociology* 129(3):753–795. **URL:** *https://doi.org/10.1086/728051*

Holbrook, Colin and Daniel M. T. Fessler. 2015. The same, only different: Threat management systems as homologues in the tree of life. In *Handbook of Personal Security*, ed. Patrick J. Carroll, Robert M. Arkin and Alana L. Wichman. New York, NY: Psychology Press pp. 95–109.

Hwang, H. C. and David Matsumoto. 2014. "Dominance threat display for victory and achievement in competition context." *Motivation and Emotion* 36:208–214.

Jimeno-Ingrum, Diana, Jennifer L Berdahl and Brennis Lucero-Wagoner. 2009. "Stereotypes of Latinos and Whites: Do they guide evaluations in diverse work groups?" *Cultural Diversity & Ethnic Minority Psychology* 15(2):158.

Judd, C. M., J. Westfall and D. A. Kenny. 2012. "Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem." *Journal of Personality and Social Psychology* 103(1):54–69.

Judd, Charles M., Jacob Westfall and David A. Kenny. 2017. "Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power." *Annual Review of Psychology* 68:601–625.

Krippendorff, Klaus. 2019. *Content Analysis: An Introduction to Its Methodology*. 4 ed. SAGE Publications, Inc.

Laustsen, Lasse. 2014. "Decomposing the relationship between candidates' facial appearance and electoral success." *Political Behavior* 36(4):777–791.

Levin, J. and A. R. Thomas. 1997. "Experimentally manipulating race: Perceptions of police brutality in an arrest: A research note." *Justice Quarterly* 14(3):577–586.

Levine, Emma E. and Maurice E. Schweitzer. 2015. "The affective and interpersonal consequences of obesity." *Organizational Behavior and Human Decision Processes* 127:66–84.
**URL:** *https://www.sciencedirect.com/science/article/pii/S0749597815000035*

Lukaszewski, Adam W., Zachary L. Simmons, Cameron Anderson and James R. Roney. 2016. "The Role of Physical Formidability in Human Social Status Allocation." *Journal of Personality and Social Psychology* 110(3):385–406.

Maddan, Sean, Jeffery T. Walker and J. Mitchell Miller. 2008. "Does size really matter?: A reexamination of Sheldon's somatotypes and criminal behavior." *The Social Science Journal* 45(2):330–344.
**URL:** *https://www.sciencedirect.com/science/article/pii/S036233190800030X*

Marins, E, A Rombaldi and F Del Vecchio. 2018. "Occupational physical fitness tests for police officers–a narrative review.".

Marins, Eduardo F., Gabriela B. David and Fabrício B. Del Vecchio. 2019. "Characterization of the Physical Fitness of Police Officers: A Systematic Review." *Journal of Strength and Conditioning Research* 33(10):2860–2874.

McCandless, Boyd R, W Scott Persons III and Albert Roberts. 1972. "Perceived opportunity, delinquency, race, and body build among delinquent youth." *Journal of Consulting and Clinical Psychology* 38(2):281.

McNeish, Daniel. 2018. "Thanks coefficient alpha, we'll take it from here." *Psychological Methods* 23(3):412–433.

Melamed, David, Christopher W Munn, Leanne Barry, Bradley Montgomery and Oneya F Okuwobi. 2019. "Status characteristics, implicit bias, and the production of racial inequality." *American Sociological Review* 84(6):1013–1036.

Mize, Trenton D. 2025. "Competence Perceptions at the Intersection of Gender and Race-Ethnicity." *Social Psychology Quarterly* 88(2):271–284.
**URL:** *https://journals.sagepub.com/home/spq*

Mize, Trenton D and Bianca Manago. 2018. "The stereotype content of sexual orientation." *Social Currents* 5(5):458–478.

Monin, B. and D. M. Oppenheimer. 2014. "The limits of direct replications and the virtues of stimulus sampling." *Social Psychology* 45(4):299–300.

Mueller, Ulrich and Allan Mazur. 1996. "Facial dominance of West Point cadets as a predictor of later military rank." *Social forces* 74(3):823–850.

Nature Human Behaviour Editorial Board. 2023. "Living in a brave new AI era." *Nature Human Behaviour* 7(11):1799.

Olivola, Christopher Y and Alexander Todorov. 2010*a*. "Elected in 100 milliseconds: Appearance-based trait inferences and voting." *Journal of nonverbal behavior* 34(2):83–110.

Olivola, Christopher Y and Alexander Todorov. 2010*b*. "Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences." *Journal of Experimental Social Psychology* 46(2):315–324.

Park, H., G. Kim, D. Lee and H.K. Kim. 2024. Can You Spot the AI-Generated Images? Distinguishing Fake Images Using Signal Detection Theory. In *Cross-Cultural Design. HCII 2024*, ed. P.L.P. Rau. Vol. 14702 of *Lecture Notes in Computer Science* Springer, Cham pp. 287–302.
**URL:** *https://doi.org/10.1007/978-3-031-60913-8*$_2$1

Pica, E., C. L. Sheahan, J. Pozzulo and C. Bennell. 2020. "Guns, gloves, and tasers: Perceptions of police officers and their use of weapon as a function of race and gender." *Journal of Police and Criminal Psychology* 35:348–359.

Pingitore, Regina, Bernard L. Dugoni, R. Scott Tindale and Bonnie Spring. 1994. "Bias Against Overweight Job Applicants in a Simulated Employment Interview." *Journal of*

*Applied Psychology* 79(6):909–917.
**URL:** *https://www.researchgate.net/publication/15347435*

Puhl, Rebecca and Kelly D Brownell. 2013. "Bias, discrimination and obesity." *Health and human rights in a changing world* pp. 581–606.

Puhl, Rebecca M and Kelly D Brownell. 2006. "Confronting and coping with weight stigma: an investigation of overweight and obese adults." *Obesity* 14(10):1802–1815.

Ridgeway, Cecilia L. 2019. *Status: Why is it everywhere? Why does it matter?* Russell Sage Foundation.

Roehling, Mark V. 1999. "Weight-based discrimination in employment: Psychological and legal aspects." *Personnel Psychology* 52(4):969–1016.

Salerno, J. M. and J. Sanchez. 2020. "Subjective Interpretation of "Objective" Video Evidence: Perceptions of Male Versus Female Police Officers' Use-of-Force." *Law and Human Behavior* 44(2):97–112.

Sheldon, William Herbert. 1949. "Varieties of delinquent youth: an introduction to constitutional psychiatry.".

Sheldon, William Herbert, S. S. Stevens and W. B. Tucker. 1940. *The Varieties of Human Physique: An Introduction to Constitutional Psychology*. New York: Harper Brothers.

Todorov, Alexander, Anesu N Mandisodza, Amir Goren and Crystal C Hall. 2005. "Inferences of competence from faces predict election outcomes." *Science* 308(5728):1623–1626.

Velásquez-Salamanca, D., M. Á. Martín-Pascual and C. Andreu-Sánchez. 2025. "Interpretation of AI-Generated vs. Human-Made Images." *Journal of Imaging* 11(7):227.
**URL:** *https://doi.org/10.3390/jimaging11070227*

von Rueden, Christopher. 2014. The Roots and Fruits of Social Status in Small-Scale Human Societies. In *The Psychology of Social Status*, ed. Joey Cheng, Jessica L. Tracy and Cameron Anderson. New York, NY: Springer pp. 179–200.

Wagner, Michael, Matthew Harper, Ashley R. Rockwell and William Wells. 2023. "Police Chief Perceptions of Officer Physical Fitness and Barriers to Better Fitness." *International Journal of Police Science* 2(1).
**URL:** *https://doi.org/10.56331/ijps.v2i1.7613*

Weitzer, Ronald. 2000. "White, black, or blue cops? Race and citizen assessments of police officers." *Journal of Criminal Justice* 28(4):313–324.
**URL:** *https://doi.org/10.1016/S0047-2352(00)00043-X*

Wells, G. L. and P. D. Windschitl. 1999. "Stimulus Sampling and Social Psychological Experimentation." *Personality and Social Psychology Bulletin* 25(9):1115–1125. Original work published 1999.

Yap, Andy J., Malia F. Mason and Daniel R. Ames. 2013. "The powerful size others down: The link between power and estimates of others' size." *Journal of Experimental Social Psychology* 49:591–594.
**URL:** *http://dx.doi.org/10.1016/j.jesp.2012.10.003*

Zamudio, Cristian, John L. Grigsby and Matthew Michelsen. 2025. "RAISE: A New Method to Develop Experimental Stimuli for Advertising Research with Image Generative Artificial Intelligence." *Journal of Advertising* pp. 1–16.

Zanolie, Kiki, Saskia van Dantzig, Inge Boot, Jasper Wijnen, Thomas W. Schubert, Steffen R. Giessner and Diane Pecher. 2012. "Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension." *Brain and Cognition* 78:50–58.
**URL:** *http://dx.doi.org/10.1016/j.bandc.2011.10.006*

Zeng, T. Chen, Joey T. Cheng and Joseph Henrich. 2022. "Dominance in Humans." *Philosophical Transactions of the Royal Society B: Biological Sciences* 377(1845):20200451.