ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202511.00145

Mental Representations and Inference Patterns of Facial Social Exclusion

Authors: Hou Chunna, Ma Yisheng, Wu Lin, Liu Zhijun, Hou Chunna, Wu Lin

Date: 2025-11-23T00:00:00+00:00

Abstract

Based on trait attribution theory, this study employed reverse correlation image classification technique to investigate the mental representation content and trait inference patterns of faces that trigger social exclusion. With Chinese university students as participants, Study 1 primed social exclusion contexts and implemented a two-image forced-choice task, revealing that mental representations of excluded individuals' faces incorporate trustworthiness and dominance trait information, wherein the nose and mouth (trustworthiness) and eyebrows (dominance) constitute core diagnostic information regions. Study 2, through objective measurements and subjective evaluations, demonstrated consistent results indicating that low-trustworthiness, low-dominance mental representations represent typical trait characteristics of social exclusion, with both playing crucial roles in the trait inference process of social exclusion. Importantly, low trustworthiness exhibited greater weight of influence, providing strong support for the trustworthiness priority hypothesis. This study elucidates the role of facial mental representations in social exclusion processes, extends the trustworthiness priority mechanism from verbal materials to facial stimuli, and furnishes novel theoretical foundations for understanding exclusion processes at the level of higher-order social cognition.

Full Text

The Mental Representation and Inference Patterns of Facial Social Exclusion

HOU Chunna[1], MA Yisheng[1], WU Lin[2], LIU Zhijun[3]
([1]School of Psychology, Northeast Normal University, Changchun 130024, China)
([2]School of Sociology, Wuhan University, Wuhan 430072, China)

([3]Department of Sociology, Changchun University of Science and Technology, Changchun 130022, China)

Abstract

Drawing on trait attribution theory, this study employed reverse correlation image classification techniques to investigate the psychological representation content and trait inference patterns underlying facial cues that elicit social exclusion. Using Chinese college students as participants, Study 1 utilized a two-image forced-choice task following a social exclusion priming procedure. The findings revealed that mental representations of excluded faces contain trait information related to trustworthiness and dominance, with the nose and mouth (trustworthiness) and eyebrows (dominance) serving as core diagnostic regions. Study 2 demonstrated, through both objective measurement and subjective evaluation, that low-trustworthiness and low-dominance mental representations constitute the typical trait profile of social exclusion, with both dimensions playing crucial roles in the trait inference process. Notably, low trustworthiness carried greater weight, providing robust support for the trustworthiness primacy hypothesis. This research illuminates the role of facial mental representations in social exclusion processes, extends the trustworthiness primacy mechanism from verbal materials to facial stimuli, and offers new theoretical insights into understanding exclusion at the level of higher-order social cognition.

Keywords: face, social exclusion, mental representation

As social beings, humans depend on establishing and maintaining social relationships to ensure survival and reproductive success. Individuals who adapt well to their social environment gain acceptance, whereas those who fail may face social exclusion. This phenomenon encompasses dual experiential dimensions, touching not only on the feelings of the excluded but also on the psychological states of those who exclude. Research from the perspective of the excluded has primarily focused on behavioral responses following ostracism, such as prosocial behavior, aggressive behavior, and withdrawal (Casini et al., 2022; Chen et al., 2023; Dickerson & Quas, 2024). While exclusion causes significant distress and threatens the survival interests of targets, it can serve adaptive functions for perpetrators. By selectively excluding or ignoring specific individuals, actors aim to obtain potential adaptive benefits and strengthen group cohesion (Wyer & Schenke, 2016). However, malicious and unjustified exclusion readily elicits condemnation and punishment from others, and baseless exclusion can cause discomfort to the excluder. Consequently, people do not readily exclude others without sufficient justification (Rudert et al., 2020).

Among studies examining the excluder' s perspective, investigating the mechanisms underlying exclusion behavior remains a central academic concern. In subtle interpersonal interactions, facial traits constitute a critical element shaping first impressions. Most evolutionary psychologists believe that natural selection has equipped humans with systems to monitor and respond to cues signaling

social exclusion. These systems link such cues to appropriate warning and response mechanisms, enabling selective exclusion of others and detection of one' s own actual or potential exclusion (Over & Uskul, 2016). This evolved psychological adaptation mechanism may be related to trait inferences drawn from facial cues (Rudert et al., 2020).

Trait Attribution Theory posits that perceivers infer personality characteristics from unfamiliar faces along two trait dimensions: trustworthiness and dominance (Oosterhof & Todorov, 2008; Todorov et al., 2015). This process occurs extremely rapidly (within 100ms) and shows cross-cultural consensus (Todorov & Oosterhof, 2011). These represent evolved human capacities for evaluating environmental stimuli, nearly fully explaining how people perceive others (Fiske et al., 2007). Trustworthiness involves intentions to help or harm, providing information about potential trust or distrust, while dominance reflects the capacity to execute these intentions (Todorov et al., 2015). These correspond to the "warmth" (trustworthiness) and "competence" (dominance) dimensions in the Stereotype Content Model (SCM; Fiske et al., 2007) and constitute two fundamental dimensions of social cognition (Park & Baumeister, 2015).

How do facial trait inferences trigger social exclusion? Scholars hypothesize that this involves socially consensual facial stereotypes about exclusion targets that people hold in their minds (Rudert et al., 2020). Stereotypes, as beliefs about the personalities, characteristics, and behaviors of certain group members, essentially represent mental representations of actual (or perceived) differences (Chattalas, 2005). Mental representations, as schematic knowledge structures, are the forms through which information or knowledge is manifested and recorded in mental activities (Hou & Liu, 2019). The content of mental representations forms the basis for information interpretation, generating evaluations of specific targets (individuals or groups) through matching with input facial information to produce adaptive behaviors (Jack & Schyns, 2017). To date, experimental research on this topic remains scarce, primarily because methods for obtaining facial trait mental representations largely exceed the scope of traditional research approaches (Jack & Schyns, 2017).

The SCM theoretical framework provides an important reference for facial social exclusion research. SCM posits that positioning "warmth" and "competence" within quadrants of a two-dimensional space forms different stereotype contents about individuals (or groups), which constitutes the main cause of social exclusion (Fiske et al., 2007). Rudert et al. (2017) were the first to test whether facial stimuli could elicit social exclusion within the "warmth-competence" framework. Using the Basel Face Model (BFM), they manipulated vectors along the "warmth" and "competence" dimensions to generate synthetic faces in four quadrants, asking participants to imagine excluding an outgroup or ingroup member. Results showed that exclusion tendency was unrelated to group identity and depended solely on the face' s position in the warmth-competence space: cold-incompetent faces elicited the strongest exclusion, followed by cold-competent faces, while warm-incompetent faces were least likely to be excluded—consistent

with SCM. However, the mediating role of emotion differed from SCM predictions: cold-incompetent faces indeed evoked strong disgust but also pity, while warm-incompetent faces did not elicit the sympathy predicted by SCM but were instead characterized by an "absence of disgust."

Despite being the first to confirm that facial traits can elicit social exclusion, Rudert et al. (2017) had limitations. First, methodologically, they followed Oosterhof and Todorov' s (2008) data-driven modeling logic, extracting physical features individually associated with each "warmth" and "competence" dimension in face space, then linearly averaging these feature sets to synthesize four types of faces with different levels, finally mapping them onto four quadrants in a two-dimensional coordinate system for trait evaluation under the SCM framework. However, this face modeling approach directly contradicts SCM' s original method for identifying a two-dimensional trait space, which derived the quadrant space from numerous judgments combining both dimensions. Scholars have questioned whether this modeling approach truly reflects how facial traits are integrated (Oliveira et al., 2019). Second, in terms of task design, the experiment required participants to actively implement exclusion. Given the experimenter' s authority, participants might question the legitimacy of excluding others and recognize that deliberate exclusion would cause pain, potentially inducing psychological distress (Wirth & Wesselmann, 2018).

Building on this, Rudert et al. (2020) improved their methodology by employing Reverse Correlation Image Classification (RCIC) technology based on random vectors to explore the personality characteristics of facial mental representations of social exclusion. RCIC is a purely data-driven psychophysical method, uninfluenced by researchers' a priori assumptions, capable of visualizing stereotypes or mental representations of specific facial traits. During this process, participants must first activate corresponding mental representations and then compare them with input stimuli, making RCIC highly sensitive to mental representations both as a method and in its results (Oliveira et al., 2019). Using RCIC, Rudert et al. (2020) simultaneously obtained mental representations of both social exclusion and social inclusion without explicitly mentioning exclusion, confirming that people hold socially consensual beliefs about which faces are prone to exclusion. Subjective evaluations showed that mental representations of facial social exclusion combined low conscientiousness and low agreeableness, while trait ratings revealed low trustworthiness and low dominance. In this study, Rudert et al. questioned SCM' s explanatory power, proposing that facial traits are a more direct cause of social exclusion, while emotion (disgust) is a relatively distal cause. Furthermore, they argued that exclusion based on mixed stereotypes (high agreeableness and low conscientiousness) could not be explained through compensation effects. In SCM, compensation effects refer to the tendency to compensate for negative evaluation on either the warmth or competence dimension through positive evaluation on the other dimension, thereby canceling each other out (Cuddy et al., 2009).

While advancing research on social exclusion in facial stimuli, Rudert et

al. (2020) still had two issues. First, to create stimuli for the image classification task, they used shape parameters (length, roundness, etc.) from BFM as random vectors, using BFM' s average real face as the base face, and generated a set of novel real faces as stimulus materials through computational manipulation of base face shape parameters. This approach has been criticized for potentially compromising ecological validity: since each real face has a unique identity including specific morphology (shape and structure), this could trigger multiple facial inferences (emotion, identity, gender) that confound results, making it difficult to obtain the mental representations needed by experimenters (Jack & Schyns, 2017). Second, and more importantly, facial trait inference depends not on the entire face but on a few key regions. Only by precisely identifying these diagnostic information regions can we fully and necessarily recognize trait inferences from faces (Jack & Schyns, 2017). Although diagnostic regions for trustworthiness and dominance have been identified (Dotsch & Todorov, 2012; Mo et al., 2022), Rudert et al. (2020) did not report whether these trait diagnostic regions existed in mental representations of facial social exclusion. Additionally, with fewer than 100 trials, this RCIC study may have reduced the image signal-to-noise ratio, affecting mental representation image quality (Giacomin et al., 2022).

Based on this, the present study employed RCIC technology based on random noise to improve upon Rudert et al.' s (2020) methodology. This technique generates two alternative face stimuli that are mirror images of each other by superimposing opposite random sinusoidal noise onto the same base face. Mirror images refer to two face stimuli that are visually symmetrical (Hou & Liu, 2019; Okazawa et al., 2018). For example, if a region is brightened in one image, the corresponding region is darkened in the other. This design maximizes differences, capturing the maximum signal in a single trial (Dotsch et al., 2008). Since superimposing random noise distorts the base face, creating a blurred, non-real-person visual effect, this approach offers better ecological validity. Moreover, this technique does not depend on the base face and can extract mental representations through purely random noise. Most importantly, this RCIC technology enables direct analysis of pixel data from mental representations to precisely locate diagnostic information regions for facial inference (Oliveira et al., 2019).

Using this technique, the present study conducted two investigations to reveal previously undiscovered content: Study 1a aimed to obtain visualized mental representation images of Chinese facial social exclusion while ensuring ecological validity. Based on this, we attempted to answer two scientific questions: (1) What unique characteristics along trait dimensions does the mental representation of facial social exclusion possess (Study 1b)? (2) How do the two facial trait dimensions of trustworthiness and dominance influence mental representations of facial social exclusion—that is, what is the trait inference pattern (Study 2)?

Study 1a: Visualization of Mental Representation Content in Facial Social Exclusion

Study 1 aimed to generate visualized mental representation images of Chinese facial social exclusion using random-noise-based RCIC technology with more trials (300) among Chinese participants.

2.1.1 Method

Design. A single-factor (scenario type: exclusion vs. inclusion) between-subjects experimental design was employed. The independent variable was the scenario type primed in participants, including social exclusion and social inclusion levels; the dependent variable was selection results for face stimuli.

Participants. RCIC technology uses image pixel brightness values rather than single samples as the unit of analysis, ensuring statistical power through massive pixel samples. Previous research has shown that 20 participants achieve optimal results (Oliveira et al., 2019). Based on this, the present study recruited 81 participants with a mean age of 19.60 $\pm$ 0.83 years, including 42 in the social exclusion group (N_{female} = 29) and 39 in the social inclusion group (N_{female} = 24), meeting statistical power requirements.

All participants had normal or corrected-to-normal vision and had not previously participated in similar experiments. The study was approved by the Northeast Normal University Ethics Committee (2023054). Participants provided written informed consent before the experiment and received compensation afterward.

Materials. Facial stimuli were constructed by superimposing random noise onto the same base face in each trial. The base face was a male neutral-expression average face, which helps ensure ecological validity (Hou, 2017). Random noise consisted of truncated sinusoidal curve segments 叠加 from 2 cycles across 6 orientations (0°, 30°, 60°, 90°, 120°, and 150°), forming noise dot patterns containing 4092 random contrast parameters across 5 spatial scales (2, 4, 8, 16, and 32 cycles) $\times$ 2 phases (0, $\pi/2$). The generation process is illustrated in Figure 1

. The study generated 300 pairs of positive and negative face stimuli. Facial stimulus images were 512 $\times$ 512 pixels, displayed at 9.70 cm $\times$ 13.00 cm at a viewing distance of 70 cm, approximating the size of a real face (13.80 cm $\times$ 18.50 cm) viewed from 1 meter.

Procedure. The experiment consisted of two phases:

Phase 1: Social Exclusion Priming. After obtaining informed consent, the experimenter asked participants to imagine themselves in the excluder's position while watching a Cyberball game video (Wesselmann et al., 2009). Specifically, participants viewed a first-person perspective video where "I" initiated and dominated a three-player ball-tossing game. Player 1 rarely received passes

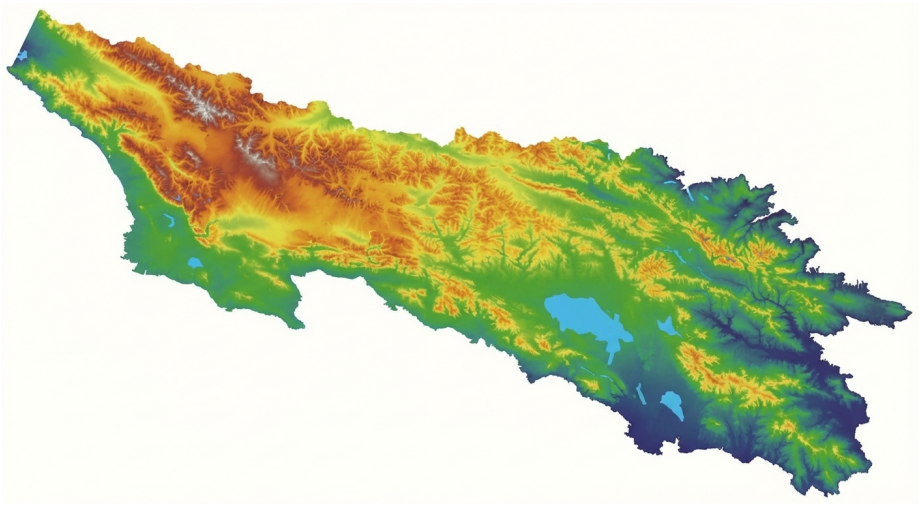

Figure 1: Figure 1

from "me," serving as the excluded role, while Player 2 received more passes, representing the socially included role.

Phase 2: Two-Image Forced Choice. After the video, participants were randomly assigned to two groups (Player 1 = social exclusion group; Player 2 = social inclusion group) to perform the RCIC two-image forced-choice task (2IFC), selecting the face that matched their assigned player. In each trial, a fixation cross appeared at the center of the screen for 500 ms, followed by two alternative face stimuli displayed side-by-side horizontally. Group-specific instructions at the top of the screen asked: "Which of the two faces better matches Player 1 (/Player 2)?" Participants responded via keypress (F key = left image; J key = right image). The 2IFC comprised 300 trials, which scholars believe significantly improves image quality (Giacomin et al., 2022). Trials were divided into 6 blocks with brief rests between blocks. The positions of the two alternative noise stimuli were balanced throughout the experiment, and face stimulus pairs appeared in random order.

After all trials, participants rated their feelings about the player identities on a 7-point scale (-3 = this person is excluded by others; +3 = this person is accepted by others) and received a small gift (approximately ¥10 value). The experimenter explained the study's purpose and provided debriefing to eliminate any negative effects.

2.1.2 Results

Data Screening. First, data were screened based on participants' ratings of player identity feelings. One participant who failed to correctly identify the excluded player and six who failed to identify the included player were excluded,

leaving 41 participants in the social exclusion group and 33 in the social inclusion group for RCIC data analysis.

Second, to improve signal-to-noise ratio and obtain higher-quality group classification images (groupCI), individual classification images (individual CI) were screened, and one participant in the social exclusion group whose individual CI negatively correlated with groupCI was excluded. Ultimately, valid RCIC data were obtained from 73 participants (40 in the social exclusion group, 33 in the social inclusion group).

Data Consistency Check. The academic consensus holds that averaging individual CI images to create group CI images is meaningful only when noise patterns across individuals achieve good consistency. Therefore, this study tested data consistency among individual CIs. Significant regions in the group CI face were used as calculation ranges for each individual CI. Specifically, non-informative standardized regions outside the effective areas of the group CI face were selected as baselines (to determine means and standard deviations). The group CI image was Z-transformed using this baseline, with $Z \geq 2.58$ as the threshold criterion to construct significant regions.

Intraclass Correlation Coefficient (ICC) analysis using a two-way random model showed ICC = 0.972 (95% CI [0.972, 0.973]) among individuals in the social exclusion group and ICC = 0.938 (95% CI [0.935, 0.942]) in the social inclusion group, indicating excellent consistency (ICC > 0.90 indicates excellent consistency; Koo & Li, 2016). Supplementary Pearson correlation analysis revealed that the average correlation coefficient between individual CI and groupCI noise patterns was 0.954 (range: 0.894–0.983, quartiles: 0.947, 0.960, 0.965) in the social exclusion group and 0.817 (range: 0.675–0.919, quartiles: 0.741, 0.830, 0.873) in the social inclusion group, with all correlations positive. These results indicate satisfactory RCIC data consistency within each group, providing a foundation for generating mental representation images of social exclusion and inclusion.

Mental Representation Images of Facial Social Exclusion. Following RCIC data analysis procedures, the forced-choice results from all trials for each participant in the social exclusion group—that is, the averaged combination of random noise patterns from all selections matching the "excluded person" image—were used to generate the social exclusion group' s group CI (see Figure 2b [FIGURE:2]). Superimposing this onto the base face yielded the mental representation image of the excluded person in the excluder' s mind (see Figure 2a). A similar process generated the mental representation image of the included person.

To better present the diagnostic information effects of social exclusion mental representations, this study attempted visual information reconstruction of facial noise on the original mental representation data. This could more "realistically" display the diagnostic information characteristics of social exclusion mental representations, while also reconstructing social inclusion mental representation

effects for comparison (Figure 3

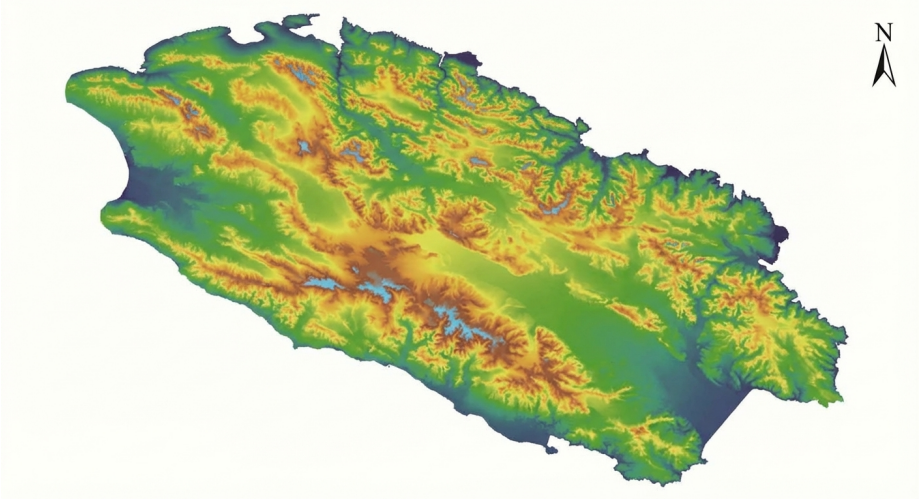

Figure 2: Figure 3

).

Supplementary Analysis. This study additionally used the Structural Similarity index (SSIM) to examine similarity between social exclusion and social inclusion mental representations, as its advantage lies in high consistency with observers' subjective judgments. SSIM results showed that noise patterns between social exclusion and social inclusion mental representations were significantly negatively correlated (SSIM = -0.334, $p < 0.001$), consistent with conventional understanding. Further 3D heat maps visually presented the negative correlation results through light-dark contrast (see Figure 2c), thereby ruling out the possibility that random noise produced artifacts.

Study 1b: Diagnostic Information Regions in Mental Representations of Facial Social Exclusion

2.2.1 Purpose

Study 1b aimed to examine the diagnostic information regions in mental representations of facial social exclusion and compare them with known diagnostic regions of trait mental representations, thereby revealing the trait characteristics of facial social exclusion mental representations.

2.2.2 Hypotheses

Based on existing research on social exclusion, faces, and SCM, Study 1b proposed the following predictions:

(1) **Trustworthiness.** SCM posits that people are extremely sensitive to distinctions between trust and distrust in the social world, and such distinctions are difficult to change (Fiske et al., 2007). Unfriendly, untrustworthy individuals are subject to exclusion (Hales et al., 2016). Evidence from mental representations of facial traits shows that the eye, nose, and mouth regions are diagnostic information for trustworthiness trait inference, with cross-cultural universality independent of race and gender (Dotsch & Todorov, 2012; Ma et al., 2015; Mo et al., 2022). The cheekbone, near the eye region, is also diagnostic for facial trustworthiness judgments (prominent/flat) (Todorov et al., 2008).

(2) **Dominance.** People desire dominance capabilities, believing that being in control yields personal benefits, and produce exclusion through two pathways. Active exclusion occurs when others' high dominance threatens one' s status, prompting active exclusion. Passive exclusion manifests as neglect of individuals with low capability who cannot provide adaptive benefits (Chen et al., 2017). Evidence from facial trait mental representations shows that the eyebrow region (eyebrows/supraorbital ridge) and facial contour are primary diagnostic information sources for dominance (Dotsch & Todorov, 2012).

In summary, according to SCM, untrustworthy and high/low dominance traits can all produce facial social exclusion. If mental representations of facial social exclusion are indeed generated through trait information, their diagnostic information regions should reflect these areas. Recent literature indicates that eyes and mouth are core cue regions for forming trustworthiness inferences (Mo et al., 2022). However, Chinese people avoid direct eye contact in social interactions (Wang et al., 2020), preferring to use the lower face for information diagnosis, with the mouth carrying the most information in the face (Blais et al., 2012). Therefore, **Hypothesis 1** predicted: If mental representations of facial social exclusion produce exclusion through untrustworthy trait information, then the mouth is a diagnostic information region.

Meanwhile, eyebrow features are relatively stable. Although not fixed facial structures (like brow ridges), eyebrows serve as reference points dividing the forehead and eye socket into two facial regions. Losing eyebrows reduces efficiency in extracting facial information and makes recognizing others difficult (Sekuler et al., 2004). Therefore, **Hypothesis 2** predicted: If mental representations of facial social exclusion produce exclusion through dominance (high/low) trait information, then eyebrows are a diagnostic information region.

2.2.3 Method

Participants. Study 1b required only data analysis of mental representation images obtained in Study 1a, without recruiting additional participants.

Materials and Data Analysis Procedure. Study 1b used three images as research materials: the 512 × 512 pixel RGB-format social exclusion and social

inclusion groupCIs generated in Study 1a, plus the base face. The first two were used for data analysis, while the base face was used for visual presentation of results. Data analysis used pixels as the unit of analysis, applying Gaussian filtering with a standard deviation of σb = 4 pixels for smoothing, with a significance threshold of Zcrit $\geq |2.3|$, p $< .05$ for two-sided pixel tests and diagnostic difference image analysis (Dotsch & Todorov, 2012).

Pixel Test Results. Study 1b used pixel test analysis to reveal diagnostic information regions in social exclusion mental representations, shown in the left (a) portion of Figure 4

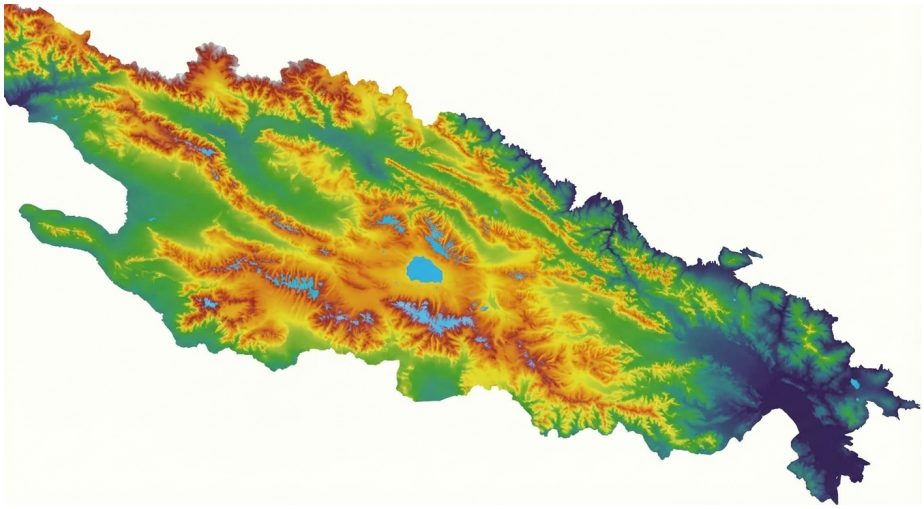

Figure 3: Figure 4

. Red and green regions represent significant pixel test areas, with green clusters indicating pixel brightness positively correlated with social exclusion (i.e., when these pixels are brighter, participants are more likely to make social exclusion judgments). Conversely, red clusters indicate pixel brightness negatively correlated with social exclusion (darker pixels lead to greater tendency for social exclusion judgments). Results showed that diagnostic information for social exclusion encompassed most key facial information regions: the forehead, eye region, nose region, mouth region, and peripheral facial contour regions.

The eye and mouth regions are considered key diagnostic information regions for trustworthiness trait inference (Dotsch & Todorov, 2012). The nose region is also diagnostic for trustworthiness trait inference (Dotsch & Todorov, 2012). This suggests that social exclusion mental representations may contain trustworthiness trait information. Meanwhile, eyebrows are important diagnostic information regions for dominance (Dotsch & Todorov, 2012), while trustworthiness cannot be distinguished through eyebrows (Todorov et al., 2015). Significant pixel clusters scattered at facial edges showed low distinction between

face and background contours, which is diagnostic information for low dominance (Dotsch & Todorov, 2012). This indicates that social exclusion mental representations may simultaneously contain dominance trait information.

Diagnostic Difference Image Analysis Results. In the Cyberball paradigm, revealing characteristics of excluded individuals requires comparison with included individuals, with differences obtained through this comparison revealing unique features of social exclusion (Syrjämäki & Hietanen, 2020). Therefore, Study 1b performed diagnostic difference image analysis (process shown in the lower portion of Figure 4, where d1 -d2 = d). Results showed that after eliminating significant regions overlapping with social inclusion, unique diagnostic information regions for social exclusion mental representations concentrated primarily in the eyebrow, nose, and mouth regions. Given that significant pixel regions did not show left-side concentration, results could rule out effects of face processing lateralization, confirming that information differences originated from facial information itself.

Previous literature on facial trait inference mental representations indicates that brighter mouth regions (green) correspond to lower trustworthiness (Liu, 2021; Mo et al., 2022). The mouth region outlined by white dashed lines in Figure 4c is consistent with this, suggesting that social exclusion mental representations may contain "untrustworthy" trait information, aligning with SCM perspectives, thus confirming Hypothesis 1. Darker eyebrow regions indicate higher dominance (red eyebrow region circled on left side of Figure 4c), while brighter eyebrow regions indicate lower dominance (green eyebrow region circled on right side of Figure 4c). This indicates that social exclusion mental representations may simultaneously contain trait information from both ends of the dominance dimension, consistent with SCM perspectives. Hypothesis 2 was also confirmed.

Additionally, the nose diagnostic information region was relatively large, essentially encompassing the entire nose length (region outlined by white trapezoid in Figure 4c). Cross-race research shows that the nose is a core cue for trustworthiness trait inference (Todorov et al., 2008), with longer noses perceived as more trustworthy (Ma et al., 2015). However, this result only indicates that social exclusion mental representations may contain trustworthiness trait information, without distinguishing between trustworthy and untrustworthy specifics. Notably, the nose region reaching significance may also relate to East Asians' special face processing strategies: East Asians tend to fixate more on the central face region, extracting more information from the nose area as a compensatory strategy for the eye (and mouth) region, thereby avoiding direct eye contact. This attention to the nose is a robust cultural fixation bias in East Asian face processing (Wang et al., 2020). Therefore, this result should be interpreted cautiously.

Study 2: Trait Inference Patterns in Mental Representations of Facial Social Exclusion

3.1 Purpose

Study 1 generated people' s mental representation images of social excluders and revealed that eyebrows, nose, and mouth are key diagnostic information regions, which overlap substantially with known diagnostic regions for trustworthiness and dominance traits, answering Question 1. However, trustworthiness and dominance are not encoded as single dimensions but are integrated together in facial mental representations to influence specific mental representation images (Gunaydin & DeLong, 2015). How these two traits operate on mental representations of facial social exclusion remained unclear in Study 1.

Study 2 aimed to examine the predictive effects of four single traits (trustworthiness: high vs. low; dominance: high vs. low) on social exclusion mental representations through both objective measurement (Study 2a) and subjective evaluation (Study 2b), exploring the specific influences of these two fundamental facial trait dimensions and revealing trait inference patterns in facial mental representations to address Question 2.

3.2 Hypotheses

Currently, two main perspectives explain how trait information influences facial mental representations to elicit social exclusion: the trustworthiness primacy hypothesis (Oliveira et al., 2019) and the negative halo effect hypothesis (Rudert et al., 2020).

(1) **Trustworthiness Primacy Hypothesis.** Oliveira et al. (2019) proposed that at the mental representation level, trait dimensions also form a two-dimensional space similar to SCM. Integrated trait information maps onto this face space, ultimately forming stereotypes (mental representations) of others. In this process, both dimensions operate in parallel on specific facial content in an integrated manner, following a principle of unequal weighting, with trustworthiness carrying greater weight than dominance—hence, trustworthiness primacy. From an evolutionary psychology perspective, to increase survival and reproductive chances, identifying whether others have harmful intentions (trustworthiness) is more important than judging their capacity to execute these intentions (dominance) (Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008). Therefore, regardless of dominance level (high/low), once others are identified as untrustworthy, individuals will choose to distance or exclude them based on approach-avoidance principles. Evidence shows that integration centered on low trustworthiness readily elicits social exclusion, with low trustworthiness-low dominance (cold-incompetent) evoking the most severe social exclusion, followed by low trustworthiness-high dominance (cold-competent), while integration centered on low dominance with high trustworthiness (warm-incompetent) does not produce exclusion (Rudert et al., 2017).

(2) **Negative Halo Effect Hypothesis.** Rudert et al. (2020) proposed that trustworthiness and dominance are not linear functions—that is, negative evaluation on one dimension cannot be compensated by positive evaluation on the other dimension. Instead, they follow a nonlinear negative weighting enhancement, where negative evaluation on one trait dimension leads to more negative overall evaluation, a phenomenon they termed the negative halo effect. Even when the other trait dimension is evaluated positively, this contrast effect causes the negative trait dimension to have an even stronger negative impact. Taking trustworthiness as an example, from a utilitarian perspective, the function of social exclusion is to screen out potentially poor cooperators. Once trustworthiness is negatively evaluated (untrustworthy, unable to cooperate), dominance loses almost all weight; even high ability cannot compensate for lack of trust, ultimately leading to exclusion.

Both perspectives agree that trustworthiness and dominance operate in parallel on facial mental representations, with unequal effects across the two trait dimensions. However, they differ in that the trustworthiness primacy effect identifies trustworthiness as the key trait dimension, proposing that only untrustworthy trait information produces social exclusion, whereas the negative halo effect hypothesis rejects any assumed key status for a single trait dimension, emphasizing instead the impact of negative trait information itself on social exclusion. The negative halo effect is premised on denying the existence of compensation effects for facial traits (Rudert et al., 2020). However, existing evidence suggests the evidentiary premise of the negative halo effect hypothesis is difficult to sustain, as recent laboratory evidence shows that compensation effects appear not only in word stimuli but also in facial trustworthiness and dominance (Schmitz et al., 2024). Therefore, this study proposes **Hypothesis 3**: Trustworthiness and dominance traits operate in parallel on mental representations of facial social exclusion, with trustworthiness carrying greater weight than dominance.

3.3 Study 2a: Objective Measurement of Facial Trait Effects on Social Exclusion Mental Representations

3.3.1 Pilot Study: Generation of Four Single-Trait Facial Mental Representations This pilot study aimed to obtain mental representation images of four types from the two basic dimensions of trait attribution theory (trustworthiness: high vs. low; dominance: high vs. low) for subsequent pixel regression model analysis of social exclusion mental representations.

Participants. Using the same RCIC paradigm as Study 1, 200 college student volunteers aged 17-22 were recruited (mean age = 19.41 ± 1.18 years, N_{female} = 124).

Materials and Procedure. Facial noise stimuli came from Study 1. The 2IFC task procedure was identical to Study 1, except that instructions were tailored to four trait conditions. Specifically, the high trustworthiness 2IFC task asked

participants to choose "which face looks more trustworthy" (35 participants); the high dominance 2IFC task asked "which face looks more capable or leader-like" (38 participants); the low trustworthiness 2IFC task asked "which face looks less trustworthy" (62 participants); and the low dominance 2IFC task asked "which face looks less capable or lacking leadership" (65 participants). Trait instructions were derived from trait attribution theory descriptions (Dotsch & Todorov, 2012; Fiske et al., 2007), with 300 trials per condition.

Data Screening and Consistency Check. Participants whose individual CI negatively correlated with groupCI were excluded from each group. Final valid RCIC data for analysis were: high trustworthiness group n = 30, low trustworthiness group n = 58, high dominance group n = 31, low dominance group n = 58. Consistency checks showed: high trustworthiness group ICC = 0.951 (95% CI [0.949, 0.953]), average Pearson correlation between individual CI and groupCI noise patterns = 0.927 (range: 0.847–0.971, quartiles: 0.897, 0.933, 0.956); low trustworthiness group ICC = 0.981 (95% CI [0.980, 0.981]), average correlation = 0.958 (range: 0.900–0.984, quartiles: 0.945, 0.963, 0.976); high dominance group ICC = 0.906 (95% CI [0.903, 0.909]), average correlation = 0.914 (range: 0.855–0.957, quartiles: 0.897, 0.913, 0.935); low dominance group ICC = 0.947 (95% CI [0.945, 0.949]), average correlation = 0.967 (range: 0.918–0.989, quartiles: 0.959, 0.969, 0.978).

All four trait type mental representations showed ICC > 0.90, providing assurance for subsequent analysis. Following the same RCIC procedure as Study 1, 512 × 512 pixel RGB mental representation face images for high trustworthiness, low trustworthiness, high dominance, and low dominance were generated as materials for pixel regression analysis in Study 2a (Figure 5 [FIGURE:5]).

3.3.2 Pixel Regression Analysis of Facial Traits Predicting Social Exclusion Mental Representations Method. Independent variables were the four trait-type mental representations generated in the pilot study (high trustworthiness, low trustworthiness, high dominance, low dominance). The dependent variable was the social exclusion mental representation obtained in Study 1. Using pixel brightness values as the unit of analysis, a pixel regression model was established to predict social exclusion mental representations from basic facial trait dimensions.

Results. Pixel regression results showed that in predicting social exclusion mental representations, low trustworthiness significantly and positively predicted social exclusion ($\beta = 0.78$, 95% CI = [0.77, 0.78]), with the largest effect among all facial trait mental representation types, indicating that low trustworthiness is the key factor in social exclusion mental representations. Additionally, high trustworthiness showed significant negative prediction ($\beta = -0.11$, 95% CI = [-0.12, -0.11]), meaning high trustworthiness attenuates social exclusion. These opposite predictive effects demonstrate that the trustworthiness dimension has a discriminative effect on social exclusion mental representations.

In contrast, the dominance dimension showed different predictive patterns. Results indicated that both high and low dominance positively predicted social exclusion mental representations, with similar magnitudes ($\beta = 0.19$, 95% CI = [0.18, 0.19]; $\beta = 0.15$, 95% CI = [0.15, 0.16]), differing from the trustworthiness dimension's pattern. These results suggest that social exclusion elicited by facial mental representations is influenced by parallel effects of trustworthiness and dominance traits, triggered by two trait integration patterns: cold-incompetent (low trustworthiness and low dominance) and cold-competent (low trustworthiness and high dominance). Both traits—particularly (un)trustworthiness—play key predictive roles. These results can be reasonably explained by both the trustworthiness primacy hypothesis and the negative halo effect hypothesis.

Residual results showed that the residual CI image was mostly filled with uniform dark pixels, with only the base image remaining, indicating that trait predictions explained most of the variance in social exclusion mental representations (as shown in Figure 6). The R^2 was 0.62, indicating good model fit.

A supplementary pixel regression model for social inclusion showed that both high trustworthiness ($\beta = 0.50$, 95% CI = [0.49, 0.50]) and high dominance ($\beta = 0.37$, 95% CI = [0.37, 0.38]) had significant positive predictive effects, with trustworthiness showing the strongest prediction. Meanwhile, low trustworthiness ($\beta = -0.17$, 95% CI = [-0.17, -0.16]) and low dominance ($\beta = -0.07$, 95% CI = [-0.07, -0.06]) both had negative predictive effects. These results indicate that trustworthiness and dominance operate in parallel on social inclusion mental representations, with both showing discriminative effects. The warm-competent (high trustworthiness and high dominance) trait integration pattern is most readily accepted, consistent with previous ingroup favoritism stereotype findings where people attribute more positive traits to ingroup members, perceiving them as "warm and competent"(Fiske et al., 2007). These results support the trustworthiness primacy hypothesis: trustworthiness plays a decisive role in trait inference, and regardless of dominance level, trait combinations with high trustworthiness are acceptable. The negative halo effect hypothesis cannot explain this integration pattern, as research has shown that warm-incompetent (high trustworthiness, low dominance) does not trigger social exclusion (Rudert et al., 2017). This pixel regression model showed moderate explanatory power ($R^2 = 0.34$), detailed in Figure 6 [FIGURE:6].

In summary, results consistently showed that both social exclusion and social inclusion mental representations are influenced by parallel effects of trustworthiness and dominance, but trustworthiness plays a key predictive role with greater weight than dominance. Hypothesis 3 was confirmed.

3.4 Study 2b: Subjective Evaluation of Facial Trait Effects on Social Exclusion Mental Representations

Study 2b used trait vocabulary assessment to transform the mental representation images from Study 2a into questionnaire content. By constructing linear

regression models of four single traits (trustworthiness: high vs. low; dominance: high vs. low) predicting social exclusion mental representations, it explored each trait' s predictive effect on social exclusion ratings to further validate Hypothesis 3.

3.4.1 Method Participants. This study recruited 153 college students (N_{female} = 105) with a mean age of 19.75 ± 0.97 years. G*Power 3.1 was used to calculate sample size (Faul et al., 2007). With $\alpha = 0.05$ and statistical power $1-\beta = 0.95$, 89 participants were sufficient to detect a medium effect size (Cohen' s f = 0.15) in linear multiple regression analysis. Thus, the current sample size met sampling requirements.

Stimuli and Equipment. Six facial mental representation images served as assessment stimuli: the two images of excluded and included persons generated in Study 1, plus the four trait-type mental representation images from Study 2a' s pilot study (low trustworthiness, high trustworthiness, low dominance, high dominance).

Stimuli were presented using PowerPoint software on an interactive smart 平板 (model: Seewo F86EA) with a 60Hz refresh rate and resolution set to 1920 × 1080 pixels.

Measures.
(1) **Social Exclusion Assessment.** A 7-point scale (-3 = excluded by others, +3 = accepted by others) assessed social exclusion levels for social exclusion and social inclusion mental representation images. The intraclass correlation coefficient was ICC = 0.93, 95% CI [0.73, 1.00], indicating excellent consistency.

(2) **Trait Assessment Vocabulary.** Trait assessment items came from domestic scholars' facial mental representation research (Liu, 2021). Trustworthiness trait assessment vocabulary included "trustworthy, friendly, honest," while dominance trait assessment vocabulary included "intelligent, skilled, capable." All items used a 5-point Likert scale.

Internal consistency reliability (Cronbach's α) for vocabulary assessments across conditions was: trustworthiness assessment $\alpha = 0.67$, 0.51, 0.63, and 0.52 for low trustworthiness, high trustworthiness, social exclusion, and social inclusion conditions, respectively; dominance assessment $\alpha = 0.54$, 0.52, 0.57, and 0.56 for low dominance, high dominance, social exclusion, and social inclusion conditions, respectively. All reliability coefficients exceeded 0.50, meeting moderate acceptability standards (Sutherland et al., 2020), though falling below the conventional threshold of 0.70 (Hussey et al., 2025). Therefore, Study 2b used composite scores generated from principal component analysis for subsequent linear regression analysis.

Procedure. Before questionnaire administration, participants signed informed consent forms. The experimenter then introduced knowledge about the two basic trait dimensions (trustworthiness and dominance) and clarified the mean-

ings of six trait vocabulary words (“trustworthy,” “friendly,” “honest,” “intelligent,” “skilled,” “capable”) to ensure understanding. The questionnaire test consisted of two phases: trait vocabulary assessment and social exclusion assessment.

(1) **Trait Vocabulary Assessment Phase.** To avoid mutual influence between the two trait dimensions, participants rated each face on only one dimension (trustworthiness or dominance) using three vocabulary items on a 5-point scale (1 = strongly disagree, 5 = strongly agree). The six stimulus images required eight assessment tasks: four single-trait mental representation images were assessed only on their corresponding dimension, while the two social exclusion and social inclusion mental representation images required assessment on both dimensions (two assessments each). In terms of presentation order, the experimenter pre-matched face lists with trait assessment questions and randomized face order to ensure consistency between on-screen faces and paper trait assessment content.

(2) **Social Exclusion Assessment Phase.** After trait vocabulary assessment, the social inclusion and social exclusion mental representation images were presented sequentially, and participants completed social exclusion assessments using the 7-point scale (-3 = excluded by others, +3 = accepted by others).

The entire test took approximately 15 minutes. All 153 participants completed the survey in three batches at the same location. Participants received a gift as compensation regardless of whether they completed all assessment tasks.

3.4.2 Results Assessment Validation. Paired samples t-test results showed that participants could significantly distinguish between social exclusion (M = -1.33, SD = 0.71) and social inclusion (M = 2.76, SD = 0.86) mental representation images, with a significant difference, $t(152) = 46.16$, $p < 0.001$, Cohen’ s d = 3.73. These results indicate that RCIC-generated social exclusion and social inclusion mental representation images were ideal and consistent with people’ s expected (stereotypical) impressions.

Linear Regression Analysis of Facial Traits Predicting Social Exclusion Mental Representations. Linear regression was used with composite scores of four single-trait types (high trustworthiness, low trustworthiness, high dominance, low dominance) as predictor variables to construct linear regression prediction models for social exclusion and social inclusion mental representation assessments. Correlation analysis results are shown in Table 1 .

Linear regression results showed that for social exclusion mental representations, low trustworthiness had a significant and largest positive predictive effect ($\beta = 0.21$, $t = 7.41$, $p < 0.001$, 95% CI = [0.15, 0.27]). Low dominance also had a significant positive predictive effect ($\beta = 0.16$, $t = 4.75$, $p < 0.001$, 95% CI = [0.09, 0.22]). This indicates that low trustworthiness and low dominance operate in parallel to influence facial social exclusion mental representations,

with low trustworthiness carrying greater weight. This integration pattern (cold-incompetent) can elicit facial social exclusion, replicating Study 2a' s findings.

For social inclusion mental representations, only high trustworthiness showed a significant positive predictive effect ($\beta = 0.44$, t = 28.13, $p < 0.001$, 95% CI = [0.41, 0.47]), while other trait types showed non-significant predictive effects (ps > 0.05). This indicates that high trustworthiness plays a more critical role in social inclusion mental representations (see Table 2). The predictive effect of low dominance was nearly zero, meaning the warm-incompetent (high trustworthiness, low dominance) integration pattern is basically unaffected by negative halo effects, a result more consistent with the trustworthiness primacy hypothesis. Both linear regression models showed good fit, with explained variance exceeding 70%.

In summary, whether for social exclusion or social inclusion mental representations, trustworthiness plays a key predictive role with greater weight than dominance. Hypothesis 3 was again confirmed.

3.5 Supplementary Meta-Analysis

If meta-analysis results support research assumptions, their reliability will be higher than exact replication of a single study. Based on this, this study used Pearson correlation coefficient r as the effect size indicator and employed meta-analysis to test the reliability of results regarding relationships between trait dimensions and social exclusion mental representations.

Comprehensive Meta-Analysis Version 3.3 (CMA 3.3) software was used for random-effects model analysis. Results showed a total effect size of r = 0.51, 95% CI [0.023, 0.800] between trait dimensions and social exclusion mental representations, not containing zero, indicating a strong correlation. Heterogeneity effects were Q = 471249, $I^2 = 99.999$, $p < 0.001$, exceeding Higgins et al.' s (2003) 75% threshold, indicating high heterogeneity and justifying use of a random-effects model. This high heterogeneity also suggests that other factors may cause differences in estimates across studies, warranting exploration of moderating variables affecting the relationship.

Based on this, this study further conducted subgroup analyses with trustworthiness and dominance as trait dimension subgroup variables. Results showed that the random-effects model for the trustworthiness dimension had a total effect size of r = 0.61, 95% CI [0.003, 0.891], k = 4, not containing zero; z = 1.97, p = 0.049, indicating a strong correlation between trustworthiness dimension and social exclusion mental representations. In contrast, the random-effects model for the dominance dimension had a total effect size of r = 0.39, 95% CI [-0.534, 0.887], k = 4, containing zero; z = 0.80, p = 0.43, indicating no significant relationship between dominance and social exclusion mental representations.

Meta-analysis results again confirmed that the trustworthiness dimension is the key factor eliciting facial social exclusion mental representations, validating Hy-

pothesis 3 once more. Unfortunately, although the high correlation between trustworthiness dimension and social exclusion came entirely from low trustworthiness (r_{study1} = 0.88; r_{study2} = 0.82), limited data (fewer than 3 studies each) prevented further subgroup analysis to determine the specific role of low trustworthiness.

4 General Discussion

Social exclusion is a ubiquitous phenomenon caused by multiple factors, among which mental representation may be one. This study examined the possibility that mental representations based on facial information can lead to excluding others. Faces serve as an indispensable information source in social interaction, containing rich trait information about individuals' psychological states (such as cooperative or harmful intentions) (Todorov et al., 2008). Humans have evolved to be highly skilled at reading these trait cues and may have formed mental representations that are used to enact prejudice or discrimination (Dotsch et al., 2008).

4.1 Trait Characteristics of Facial Social Exclusion Mental Representations

The current research shows that in social exclusion contexts, even without presenting actual faces, people can form social consensus about the expected appearance of exclusion targets. Using RCIC technology, this study revealed the intuitive facial image (or stereotype) of this mental representation. Within the trait attribution theory framework, mental representations of social exclusion were interpreted as containing rich facial information about trustworthiness and dominance, involving key diagnostic information regions for both traits. This may reflect people' s cautious attitude toward social exclusion, such that only when facial cues sufficiently match mental representation trait information are people inclined to exclude others. Furthermore, mental representation images of social exclusion and social inclusion differed markedly, with differences primarily arising from the eyebrow, nose, and mouth regions in the former. Among these, the mouth is the largest and most distinctive facial region. The mouth region conveys key trustworthiness information in mental representations, helping humans decide whether to trust others (Mo et al., 2022). This suggests that people focus more on trustworthiness dimension performance (particularly untrustworthiness) in social exclusion. This may explain why, after experiencing social exclusion, exclusion targets exhibit prosocial behavior—because they sense others' negation of their trustworthiness (exclusion) and display such behavior to rebuild social relationships (Chen et al., 2024). However, the mouth region in mental representations may be either a result of trait inference or a consequence of the visual system' s limited cognitive capacity prioritizing processing of the most information-rich region.

4.2 Trait Inference Patterns in Facial Social Exclusion Mental Representations

To further explore how facial trait information functions in the social exclusion process, this study examined trait information sources in social exclusion mental representations through both objective measurement and subjective evaluation. Given that facial trait dimensions are correlated, studying single dimensions separately has limited explanatory power (Oosterhof & Todorov, 2008), and trustworthiness and dominance are jointly encoded in facial mental representations (Gunaydin & DeLong, 2015), this study explored trait inference patterns in social exclusion mental representations using trait integration approaches. Both methods consistently found that untrustworthiness played the most critical role in predicting social exclusion, with low trustworthiness-low dominance (i.e., cold-incompetent) mental representations being the typical profile of social exclusion, consistent with stereotypical cognition about exclusion targets. People readily exclude individuals with such facial features, unaffected by their own moral inferences (Fiske, 2015; Rudert et al., 2017). Slight differences emerged: objective measurement also showed that low trustworthiness-high dominance fit social exclusion mental representation characteristics, while high trustworthiness-high dominance and high trustworthiness-low dominance characterized social inclusion mental representations, providing evidence supporting previous findings on social exclusion and inclusion (Fiske et al., 2007; Rudert et al., 2017) and helping expand theoretical explanations of social cognition content.

4.3 Theoretical Contributions

The question of how facial trait inference triggers social exclusion has been hypothesized to involve socially consensual facial stereotypes about exclusion targets that people hold in their minds (Rudert et al., 2020). Advancing this issue empirically requires effectively capturing and quantifying mental representations of facial traits. Our research advances the application of the trait attribution model (trustworthiness and dominance) in the domain of facial exclusion and promotes the development of mental representation research in social exclusion studies.

Two typical theoretical explanations currently exist for how trait inference triggers social exclusion: the trustworthiness primacy hypothesis and the negative halo effect hypothesis. Evidence from this study suggests that the trustworthiness primacy hypothesis may provide a more appropriate explanatory pathway, excluding the negative halo effect hypothesis' s theoretical explanatory power. The two traits operate in parallel on specific facial content in an integrated manner, following a principle of unequal weighting, with trustworthiness carrying greater weight than dominance—trustworthiness primacy. Although both trustworthiness and dominance can be spontaneously encoded in mental representations, trustworthiness is the core dimension and crucial discriminative evidence for determining social exclusion (Oosterhof & Todorov, 2008). This priority ranking is reflected in social contexts where individuals judge trustworthiness

earlier and extract trustworthiness cues more rapidly and accurately than dominance cues (Todorov et al., 2009). This means that when deciding whether we can trust others, we match strangers' faces with trustworthiness mental representations to infer whether we should approach or avoid them. Low trustworthiness inference results lead to insurmountable suspicion, making low-trustworthiness individuals more susceptible to exclusion (Vandewouw et al., 2020).

In summary, this study' s theoretical contribution lies in revealing that facial mental representations eliciting social exclusion use (un)trustworthiness as a key cue, thereby providing experimental evidence for the "trustworthiness primacy hypothesis" and demonstrating the limited explanatory power of the negative halo effect hypothesis. This research extends the trustworthiness primacy mechanism from verbal materials to nonverbal facial stimuli, offering new theoretical foundations for understanding exclusion processes at the level of higher-order social cognition.

4.4 Methodological Contributions

This study introduced random-noise-based RCIC technology to improve research methods for facial social exclusion mental representations. This technology, through its purely data-driven principles, combines interdisciplinary advantages from engineering, neurophysiological system identification (e.g., fMRI), psychophysics, experimental psychology, and computer science, overcoming limitations of traditional psychophysical methods in presenting cognitive content (Jack & Schyns, 2017). Random-noise-based RCIC technology enhances ecological validity, enables direct visualization of facial social exclusion mental representations, and calculates diagnostic information regions for psychological inference through pixel data (Oliveira et al., 2019). This technology advances understanding of the relationship between facial traits and social exclusion.

This study provided good convergent validity for findings through both objective and subjective methodological measures. The objective measurement approach used RCIC technology, unaffected by researchers' a priori assumptions, to obtain visualized images of social exclusion and single-trait mental representations, directly predicting the effects of different trait dimensions on social exclusion through pixel regression relationships among these images, without influence from explicit factors in the experiment. The subjective evaluation method provided beneficial corroboration for these findings. The consistent discovery was that untrustworthiness is the key trait dimension affecting social exclusion mental representations, and individuals with "cold-incompetent" profiles are highly vulnerable to exclusion.

4.5 Practical Implications

Mental representations are not fixed; they are influenced by multiple factors and can be shaped and adjusted through specific means. Facial trait inference is also affected by cultural environments. Therefore, through positive cultural

dissemination and education, we can not only optimize the construction process of mental representations but also directly improve stereotypes themselves, thereby promoting social inclusion and harmony.

Furthermore, mental representations can be shaped and adjusted through specific training methods. For example, Soto et al. (2020) found that categorization training can improve facial recognition or trait performance in individuals with autism spectrum disorder. This training changes individuals' mental representations, making them more stable amid changes in irrelevant dimensions and enhancing their social cognitive abilities (Soto et al., 2020). This suggests that targeted training can alter how individuals mentally represent social information, thereby weakening or eliminating certain social exclusions.

4.6 Limitations and Future Directions

First, the cross-cultural generalizability of conclusions remains to be verified. Although individuals extract specific cues about social exclusion from situational aspects and actively construct their interpretations of social exclusion based on these cues, trustworthiness information itself has cultural differences. When making explicit trust judgments, people tend to prefer faces with their own cultural characteristics (Sofer et al., 2017), and Chinese people's trustworthiness mental representations differ significantly from Westerners' (Mo et al., 2022). Therefore, future research requires evidence from cross-cultural studies.

Second, although this study demonstrated how people define social entities (images of excluded persons) in their minds as mental representations, it failed to fully reveal the formation process of social exclusion mental representations. This is because RCIC reflects a "top-down" processing form that, while providing diagnostic information results for the entire visual stimulus area, lacks information about "bottom-up"visual processing. Future research could attempt to use eye-tracking technology to obtain face visual fixation region processing results; the overlap or separation between these and mental representation diagnostic regions would help understand the mechanisms of social exclusion mental representations. This would undoubtedly be beneficial for comprehensively understanding the causes of social exclusion and developing effective interventions.

Additionally, current research' s equation of dominance with the SCM competence dimension remains controversial. However, the lack of specialized dominance assessment tools means the field still uses competence vocabulary for measurement, which may introduce conceptual bias. Future research should develop measurement schemes more aligned with dominance connotations to replicate these findings.

5 Conclusions

(1) Mental representation images of facial social exclusion contain trait information about trustworthiness and dominance, with the nose and mouth

(trustworthiness) and eyebrows (dominance) serving as core diagnostic information regions.

(2) Low-trustworthiness and low-dominance mental representations constitute the typical trait profile of social exclusion, with both playing crucial roles in the trait inference process underlying social exclusion. Low trustworthiness carries greater weight, supporting the trustworthiness primacy hypothesis.

References

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia, 50*(12), 2830–2838.

Casini, E., Glemser, C., Premoli, M., Preti, E., & Richetin, J. (2022). The mediating role of emotion regulation strategies on the association between rejection sensitivity, aggression, withdrawal, and prosociality. *Emotion, 22*(7), 1505–1516.

Chattalas, M. J. (2005). *The effects of national stereotypes on country of origin-based product evaluations*. New York: City University of New York.

Chen, F., Guo, T., & Wang, J. (2024). Divergent effects of warmth and competence social rejection: An explanation based on the need-threat model. *Journal of Personality and Social Psychology, 126*(3), 461–476.

Chen, L., Zeng, S., & Su, Y. (2023). The influence of social exclusion on adolescents' social withdrawal behavior: The moderating role of connectedness to nature. *Journal of Environmental Psychology, 87*, 101951.

Chen, Z., Du, J., Xiang, M., Zhang, Y., & Zhang, S. (2017). Social exclusion leads to attentional bias to emotional social information: Evidence from eye movement. *PLOS ONE, 12*(10), e0186313.

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., ···& Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1–33.

Dickerson, K. L., & Quas, J. A. (2024). Compensatory prosocial behavior in high-risk adolescents observing social exclusion: The effects of emotion feedback. *Journal of Experimental Child Psychology, 241*, 105840.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571.

Dotsch, R., Wigboldus, D., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19*(10), 978–980.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 39*(2), 175–191.

Fiske, S. T. (2015). Intergroup biases: a focus on stereotype content. *Current Opinion in Behavioral Sciences, 3*, 45–50.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83.

Giacomin, M., Tskhay, K. O., & Rule, N. O. (2022). Gender stereotypes explain different mental prototypes of male and female leaders. *The Leadership Quarterly, 33*(6), 101578.

Gunaydin, G., & DeLong, J. E. (2015). Reverse correlating love: Highly passionate women idealize their partner' s facial appearance. *PLOS ONE, 10*(3), e0121094.

Hales, A. H., Wesselmann, E. D., & Williams, K. D. (2016). Prayer, self-affirmation, and distraction improve recovery from short-term ostracism. *Journal of Experimental Social Psychology, 64*, 8–20.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560.

Hou, C. N. (2017). *Faces: The Evolutionary Code of Intergroup Trust.* Beijing: Science Press.

Hou, C. N., & Liu, Z. J. (2019). Visualization of mental representation: Noise-based reverse correlation image classification technology. *Advances in Psychological Science, 27*(3), 465–474.

Hussey, I., Alsalti, T., Bosco, F., Elson, M., & Arslan, R. (2025). An aberrant abundance of Cronbach's alpha values at .70. *Advances in Methods and Practices in Psychological Science, 8*(1), 25152459241287123.

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology, 68*(1), 269–297.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Liu, Z. J. (2021). *A study on the facial compensatory effect of stereotype in intergroup cognition.* Beijing: Social Sciences Academic Press.

Ma, F., Xu, F., & Luo, X. (2015). Children' s and adults' judgments of facial trustworthiness: the relationship to facial attractiveness. *Perceptual and Motor Skills, 121*(1), 179–198.

Mo, C., Cristofori, I., Lio, G., Gomez, A., Duhamel, J. R., Qu, C., & Sirigu, A. (2022). Culture-free perceptual invariant for trustworthiness. *PLOS ONE,*

17(2), e0263348.

Okazawa, G., Sha, L., Purcell, B. A., & Kiani, R. (2018). Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nature Communications, 9*(1), 3479.

Oliveira, M., Garcia-Marques, T., & Dotsch, R. (2019). Combining traits into a face: A reverse correlation approach. *Social Cognition, 37*(5), 516–545.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092.

Over, H., & Uskul, A. K. (2016). Culture moderates children' s responses to ostracism situations. *Journal of Personality and Social Psychology, 110*(5), 710–724.

Park, J., & Baumeister, R. F. (2015). Social exclusion causes a shift toward prevention motivation. *Journal of Experimental Social Psychology, 56*, 153–159.

Rudert, S. C., Keller, M. D., Hales, A. H., Walker, M., & Greifeneder, R. (2020). Who gets ostracized? A personality perspective on risk and protective factors of ostracism. *Journal of Personality and Social Psychology, 118*(6), 1194–1215.

Rudert, S. C., Reutner, L., Greifeneder, R., & Walker, M. (2017). Faced with exclusion: Perceived facial warmth and competence influence moral judgments of social exclusion. *Journal of Experimental Social Psychology, 68*, 101–112.

Schmitz, M., Vanbeneden, A., & Yzerbyt, V. (2024). The many faces of compensation: The similarities and differences between social and facial models of perception. *PLOS ONE, 19*(2), e0297887.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*(5), 391–396.

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception, 46*(8), 914–928.

Soto, F. A., Escobar, K., & Salan, J. (2020). Adaptation aftereffects reveal how categorization training changes the encoding of face identity. *Journal of Vision, 20*(10), 18,1–24.

Sutherland, C. A., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition, 157*, 257–267.

Sutherland, C. A., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality?. *British Journal of Psychology, 111*(2), 215–232.

Syrjämäki, A. H., & Hietanen, J. K. (2020). Social inclusion, but not exclusion, delays attentional disengagement from direct gaze. *Psychological Research, 84*(4), 1126–1138.

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*(1), 519–545.

Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine, 28*(2), 117–122.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460.

Vandewouw, M. M., Choi, E., Hammill, C., Arnold, P., Schachar, R., Lerch, J. P., ···& Taylor, M. J. (2020). Emotional face processing across neurodevelopmental disorders: a dynamic faces study in children with autism spectrum disorder, attention deficit hyperactivity disorder and obsessive-compulsive disorder. *Translational Psychiatry, 10*(1), 375.

Wang, Q., Hoi, S. P., Wang, Y., Song, C., Li, T., Lam, C. M., ···& Yi, L. (2020). Out of mind, out of sight? Investigating abnormal face scanning in autism spectrum disorder using gaze-contingent paradigm. *Developmental Science, 23*(1), e12856.

Wesselmann, E. D., Bagg, D., & Williams, K. D. (2009). "I Feel Your Pain" : The effects of observing ostracism on the ostracism detection system. *Journal of Experimental Social Psychology, 45*(6), 1308–1311.

Wirth, J. H., & Wesselmann, E. D. (2018). Investigating how ostracizing others affects one' s self-concept. *Self and Identity, 17*(4), 394–406.

Wyer, N. A., & Schenke, K. C. (2016). Just you and i: the role of social exclusion in the formation of interpersonal relationships. *Journal of Experimental Social Psychology, 65*, 20–25.

Source: ChinaXiv —Machine translation. Verify with original.