

# ***Faces, Spaces and Face Spaces: Studies on Facial Appearance Evaluation***

**Zhenghao Ju**

*The Department of Psychological and Brain Science, Colgate University, Hamilton, USA  
harj9102@gmail.com*

**Abstract.** In today's society, perceptions of facial appearance have become an integral part of social life, influencing a wide range of decisions, from trivial everyday choices to important and seemingly well-considered judgments. However, to what extent can "face" affect people is yet to be decided. Past research proposed two branches of interpretation on perception of facial images, as one argued the potency of faces while another stressed the importance of environment. This article is intended to evaluate the validity of studies from both perspectives by examining the level of intervention (how much the environmental factors were controlled) and racial identities of the selected facial image. In summary, the studies that supported "facial appearance" has more rigid control and used more racially ambiguous image and thus yield higher validity. In future research, it is recommended that researchers choose pictures from pools with higher racial diversity and include more rigid protocol in research design; incorporating 3D modelling can also improve the external validity because of more preciseness in controlling of facial features.

**Keywords:** Facial Appearance Evaluation, Methodology, Proposed Reasoning, Race

## **1. Introduction**

Imagine a scenario where a man was on his way to see a friend on Saturday. He stayed up really late last night—probably because he was invited to a social by friends, or because he could not help to watch a sport game, or because the sadness of breaking-up from a relationship kept him up—the man was fatigued at the moment. At the same time, the man needed to get to the appointed location as soon as possible; he was not very passionate about the date. At the peak of his weariness, he ran into someone who claimed to be a friend of his from his elementary school— "the best friend", according to his word. The man was intrigued but unfortunately failed to recall that such person existed, and he really needed to go. Now, the man faces the choice between staying a little longer to catch-up with the person and leaving immediately for another existing date that he was not very into. What would he choose, and, more importantly, what would be the strongest factor to influence the decision?

There could be so many answers. He might appeal to his emotion at that instant, the environment of the conversation, his experience last night, even the weather of the day. Amongst all the possible answers, some scholars might argue "facial appearance" to be one of the biggest influential factors to sway his choice. Although people are exposed to different faces under various social contexts and

despite how rapid the processing of the visual information is, human beings are able to make inference far from just faces even since ancient times. Not only inferences can be made by "reading faces", such perception reversely shaped people's cognition, too [1]. Moreover, increasing number of research on facial appearance indicated that human decisions ranging from reading emotions to political choices are highly dependent on facial appearance. Studies showed that facial appearances are able to inform the negative and positive valence of an individual even when the person who perceives is affected by depression; furthermore, the sensitivity for facial stimuli was found to have implication on political bipolarization between support for left- and right-centered policies [2, 3].

Todorov et al. conducted a study to support the same notion: even when the constituents have no idea about the identity of the candidate, they are able to successfully predict the results of the election [4]. However, the paper did not identify or specify any particular environmental factor that could contribute to the prediction— "is facial appearance the sole predictor of the results?" Inspired by the curiosity, this review paper will examine the past studies and how they impose control on environmental factors and racial identities of facial images to evaluate the validity of the theory, and testify how much can "faces" decide individuals' perceptions.

## 2. Past and current effort

How people get information from the others' look has been a topic studied by many past scholars, with various ways of interpretation. There are two mainstream theories. Previous studies largely focused on evaluating the accuracy of judgments based on facial appearance and investigating how much socially meaningful information individuals are able to infer from facial cues. Two perspectives look at the matter in distinct ways: from an ecological perspective, supported by Zebrowitz and Collins that people are able to, often in a consistent manner, accurately identify the social roles and sometimes key characteristics of the person based on basic physical characteristics, even with zero acquaintance [5]. Yet, as an opposing contention, supporters of stereotypical perspective suggested otherwise: the information obtained through perception of facial features often comes "very rich"—perhaps too rich. To be more specific, as Blair had noted in her writing, decision-makers (such as judges in criminal justice system) might make inference about the information pertaining to the person's categorized group too far, imposing traits that might not be true and therefore giving harsher conviction [6]. Both perspectives might diverge in many perspectives, e.g., how they assume about evolutionary functions in facial appearance perception, they aligned on the idea that, even on a rudimentary level, looking at a person's face influences inferences about the identity of the corresponding entity, although sometimes those inferences fall short in accuracy [6].

Before discussing these evaluations, it is important to establish how face-perception research defines and conceptualizes facial appearance. Within face-perception research, facial appearance is commonly described as the arrangement of facial features within a multidimensional "face space." Each individual's face occupies a unique location based on how its structural characteristics deviate from the average face of the population, and these distinctions are believed to guide judgments of identity and personality traits. The construction of an "average face" is therefore important because it provides a reference point for measuring facial distinctiveness. Researchers may then create caricatured faces by exaggerating the distance between an individual face and this average, allowing them to examine how deviations from the norm influence recognition and social judgment; the majority of relevant studies focus on the caricature of faces for the encoding of the baseline facial appearance. That is, past researchers define the so-called "average face" by observing on how people physiologically react to facial structures differently when they interact with various facial

appearance--each of which is quantified through the distance of deviation of facial features from a baseline position. The encoding procedure often involves multi electrode intracranial recordings, measured in neuron firing in inferotemporal cortex and other cortical responses [7, 8]. The non-linear model is the widely supported model for distinctiveness of facial appearance, which indicates that certain level of deviation from the baseline positions creates distinctiveness for the facial appearance, from which people are able make inference about the observed individual's identity [9]. To summarize, facial appearance means the ways that facial features are arrayed so particular responses from the perceiver are stimulated; responses include relating the appearance with a social role such as friend or family [10].

The modelling of facial appearance furthers, with scholars constructing categories that signify particular traits inferred from facial appearance. Oosterhof and Todorov used Principal Component Analysis (PCA, a statistical technique to reduce data dimensionality) narrowed down two dimensions that can be used to evaluate faces based on facial features: valence and dominance, each of which respectively signals whether to approach/avoid the target and whether the target possesses physical strengths [11]. Valence is further specified to trustworthiness for higher PCA score, and the results, surprisingly map onto past discoveries—ambition and agreeableness (aka circumplex model) warmth and competence [12, 13]. In short, past studies seem to agree that a model that takes inferred amiableness and inferred physical capacity into account to be the most reliable, by which facial appearances are standardized using these categories.

However, some counter-arguments have been proposed regarding the perception of facial appearance—some scholars critically pointed out the process that is hardly isolated and is highly relevant to the *within-individual variability*. In other words, even with the photograph staying the same, if the perception takes place under different circumstances, the results are likely to vary; sometimes the variance even exceeds the difference caused by variability in photo or image itself [14, 15]. The influence of image variability becomes especially evident when observers are unfamiliar with the individuals being depicted. In these situations, natural differences in lighting, viewpoint, facial expression, and image quality may cause people to incorrectly perceive separate photographs of the same individual as images of different people. Jenkins et al. found in their study that participants who are unfamiliar with the faces are more likely to identify the same faces as different people, and this is notably accentuated by *ambient photos*—photos that are taken under various environments, but for the same person [14]. Similarly, Todorov and Porter discovered that, consistent with the former study, variability within the same individual is matchable with that with different images; more importantly, the social context of the perception matters too: when asked to choose an image for political campaign, many participants' choices converge and showed the same preference for particular images [15]. Many other studies also pointed out the same, observed variance, but caused by other environmental factors such as lightings and contextual affective [16, 17]. Overall, the evidence indicates that first impressions of faces cannot be attributed solely to stable facial structures. Instead, these impressions arise through the interaction of the facial stimulus itself, the observer's familiarity with the individual, the nature of the judgment being made, and the broader social environment in which the evaluation occurs.

## 2.1. Levels of intervention

In 2022, Kurdi et al. conducted a study examining how implicit attitude is changed by reinterpretation of information [18]. Results of the study showed that attitudes toward trophy hunting of rhino in Namibia—an occupation that majority Americans found aversive to—can be easily altered after an 8-minute Podcast that explained the mechanism and significance of trophy

hunting in Africa. Significantly positive increase in both explicit and implicit attitudes toward the occupation showed that the update of information is powerful in changing participants' evaluation to a matter, even in the face of probed revision of their opinion. That is to say, the study not only looked at *how much* the Podcast was able to change the attitudes of the participants but also tested the extent to which the change *persisted* by including conditions where participants were guided to make decision in favor of their initial (i.e., negative) attitudes. Although the results maintained consistent with the first part of the study, it is important to note that experimental intervention can be crucial in manipulation of participants' attitudes.

The idea that attitudinal development could be shaped by reinforcement was established by Rhine and Silun in 1958 and was fundamentally extended by Kerpelman and Himmelfarb in 1971, where the both literatures notably found that the polarization of counterconditioning attitudes (i.e., attitudes acquired through reinterpretation) would vary directly with the percentage of reinforcement—although, surprisingly, the resistance to counterconditioning were the lowest with the most consistent reinforcement [19, 20]. Aligned with the findings, Rydell and McConnell also confirmed that explicit attitudes changed when participants were offered a goal that guides conscious information processing [21]. Therefore, how researchers shape their questions and prompts might implicitly probe or steer the participants to make different decisions. Such concerns were equally discussed in more recent studies, in which challenges including potential self-presentation bias, social desirability bias, recall bias and omission of implicit attitudes were specially noted [22].

It should be noted, however, that in many studies, experimenters have different extent of intervention to the perception of faces. Experimental manipulations can be categorized according to the aspect of perception they target. Observer-focused interventions modify factors such as attention, expectations, memory, or emotional interpretation, whereas stimulus- and context-focused manipulations alter the facial images themselves, the surrounding environment, or the contextual information accompanying the faces. These different types of intervention, hence, could lead to different results. For example, while both Jenkins et al. and Edwards et al. both looked at the difference in perception of the same facial appearance, the experimenter manipulated the variable using different procedures [14, 17].

Jenkins et al. were very intentional with minimalizing the intervention and deliberately allowed participants to judge based on their *within-individual* differences. The researchers intentionally maintained natural variability among the photographs to capture the range of appearance changes within the same individual [14]. Because the images originated from online sources rather than a controlled photography procedure, factors such as lighting, facial expression, viewpoint, and image quality were allowed to vary naturally. Although this methodology improved ecological validity, it also reduced the ability to isolate the effects of specific image features. Moreover, the experimenters did not intentionally *create* difference in environment (i.e., the environment that shaped the difference in facial appearance); instead, the difference between facial figures were close to random, as they were collected from Internet *without* specific protocol. In comparison, the method in Edwards et al. involves more substantial intervention. The study began with encoding session using provided pictures, with participants being "encoded" with the emotional context. Participants then were instructed to recall the faces along with the corresponding context. Note that for both pictures and emotional contexts (which is imposed using prompts) that participants are primed with in Edwards et al., the effects were standardized using past studies and pilot studies [23]. Eventually, Edwards et al. provided instruction for the participants to remember the given faces, namely prompts such as:

"Please look at each face and sentence carefully and try to remember both, as you will be tested on your memory for each when you come in next week [17]."

In short, Edwards et al. seemed to intervene in the opposite manner from Jenkin et al. Resembling difference is observable likewise in the study by Ramanathan et al.: the researchers in this particular study provides specific instruction entailing the levels of valence that could be possibly encountered in the subsequential visual stimuli, attempting to "conceptualize" the "bipolar characteristic of the emotional state of an individual that can be conveyed by their facial expression" [16]. The researchers encouraged the participants to consider the difference in expressed emotions, but ultimately it fell on the participants to interpret the facial appearance.

Taken together, these studies represent different levels of experimental intervention that imposes different extents of propositional reasoning. Jenkins et al. relied on naturally occurring image variation with minimal researcher guidance, Ramanathan et al. directed participants' attention toward emotional valence while preserving flexibility in interpretation, and Edwards et al. adopted a more controlled design by systematically manipulating the relationships among faces, emotional context, encoding, and memory retrieval. As a result, the differences observed across these studies may reflect not only the influence of environmental factors on face perception but also the varying degree to which perceptual processes were either naturally elicited or shaped by experimental control. A trend that was observed across the studies, however, was that when the researchers appeal to the contextual or environmental effect the procedure often involves encoding and guidance to aid the participants in recalling faces.

## 2.2. Racial identities

Social identities of an individual could comprise numerous aspects, with many of which being proven to be influential to individuals' judgement based on their perception of facial appearance. These identities include gender, racial identities, as well as perceived familiarities—the effect of celebrities and familiarities are demonstrated in several studies [14, 19]. Nevertheless, it is noteworthy that many of these potentially influential parameters were not systematically reckoned in much of the literature.

The way that racial characteristics becomes impactful to people's decision making—primarily rational and explicitly unbiased—was highlighted in the study by Blair et al., in which the authors have critically pointed out how the particular racial features could play an important role even at a scenario that requires extensive rationality, i.e., when undergraduate participants need to give sentences based on their assessment of the faces [20]. The study took 216 inmates ( $n = 216$ ) from Florida Department of Correction as samples. The central finding is that the more Afrocentric facial features the inmates have, the more likely they were to be convicted with longer sentence, and this is not only significant amongst the African American inmates, but also, surprisingly the White inmates. What is also interesting is how the same difference is not observed using binary models—that is, when the participants are asked to decide whether the inmates are Black or White. The results thus were especially powerful in demonstrating the implicit stereotyping under rational decisions like determining the level of offences of the individual and implied about potential affective influence from the racial identity that changes how people perceive the facial image.

Many studies, however, failed to include a balanced sample of different races. This is apparent in most of studies this review examines. A clear example of this is the study by Edwards et al., in which the researchers measured the ratio of faces correctly recalled changes when the facial images were are paired with certain emotional context [17]. However, the study only included one race (namely "Caucasians") in the study. The same could be seen in the study by Anderson and

Livingstone, a study that looked at the effect of caricaturing on people's report of facial proportion, with the sampled images of celebrities only comprising the ones that were "apparently European" [24]. In short, it appeared to be conventional for the researchers to not consider the racial variation in material facial images. It is a logistically efficient methodology indeed, but the impact of neglecting race as potential control variable in the design could be profound, as it was iterated before [25]; researchers ought to bear in mind that facing a particular race is able to elicit certain attitudes. Although statistically collecting samples from only one ethnicity might result in less variance in results, the external validity of the study might be lacking because of that.

Eventually, another point worth noting is the familiarity with certain race based on participants' reported identity. It is marked that familiarity could cause certain extent of difference in stimulation imposed by facial images [14]. Researchers should consider if that also map onto difference between provided facial images and participants—when there is disparity between sources of the images and that of participants, the mismatch could create conspicuous noises in the findings. However, this inference is chiefly hypothetical and needs more support from realistic data.

By contrast, studies that focus on or relate to encoding of facial spaces seemed to incorporate better impartiality in racial identities. That is to say, study by Alreja et al., and study by Cao et al. used a much broader sample and utilization of AI for standardization of facial appearance [10, 26]. More dated studies such as Todorov et al. also presented modelling of facial appearance using 3D models that focused on proportion of facial features whilst not considering the racial characteristics of the faces [27]. Common place in these studies is the lack of distinguishment of ethnicities when appealing to facial images; rather, the said researchers used ethnically ambiguous models and computational images for standardization across samples. Notwithstanding the lack of current effort that examine whether these models are leaning toward Eurocentric facial features, the encoding studies should inspire more usage of racially neutral facial model rather than racially unbalanced samples from real world.

### 3. Conclusion

As the aggregate of the studies showed, researchers often operationalized the studies without considering the level of intervention and racial identities of the provides images or participants, albeit it was already found that particular race might elicit different affective outcomes. However, when looking at specific facial features, the studies largely utilized racially ambiguous models and technologies that standardize distribution in "face spaces", by which the researchers achieve better validity using these techniques. Back to the study by Todorov et al. in 2005, as a groundbreaking study that seemed to overlook the environmental factors that are potentially affecting the voters' decision, the gross results of this review successfully validate the conclusion made by Todorov et al., supporting that facial features are somewhat more powerful factors indeed. In future studies, researchers should appeal to more literature in deciding if they are to add prompting and encoding procedure that reinforces the remembrance of facial information, as past studies have demonstrated that adding or removing the reinforcement can make substantial difference in thought process. Only when the levels of intervention are more consistent can the comparison between environment and facial appearance be valid. More importantly, the researchers should consider cooperation with 3D modelling for the involvement of racially ambiguous images to lower the potential influence from racial characteristics. Finally, it would be greatly helpful if more researchers include data of racial identities of the participants for exploratory data, so more evidence could be gathered to verify or challenge the existing theory about facial familiarity.

## References

- [1] Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78(5), 837–852. <https://doi.org/10.1037/0022-3514.78.5.837>.
- [2] Kan, Y., Mimura, M., Kamijima, K., & Kawamura, M. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(12), 1667–1671. <https://doi.org/10.1136/jnnp.2004.036079>
- [3] Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., Hatemi, P. K., & Hibbing, J. R. (2008). Political attitudes vary with physiological traits. *Science*, 321(5896), 1667–1670. <https://doi.org/10.1126/science.1157627>
- [4] Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623–1626. <https://doi.org/10.1126/science.1110589>
- [5] Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc.*, 1(3), 204–223. [https://doi.org/10.1207/s15327957pspr0103\\_2](https://doi.org/10.1207/s15327957pspr0103_2)
- [6] Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, 83(1), 5–25. <https://doi.org/10.1037/0022-3514.83.1.5>
- [7] Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102), 572–575. <https://doi.org/10.1038/nature04951>
- [8] Hoffman, K. L., & Logothetis, N. K. (2009). Cortical mechanisms of sensory learning and object recognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1515), 321–329. <https://doi.org/10.1098/rstb.2008.0271>
- [9] Dakin, S. C., & Omigie, D. (2009). Psychophysical evidence for a non-linear representation of facial identity. *Vision Research*, 49(18), 2285–2296. <https://doi.org/10.1016/j.visres.2009.06.016>
- [10] Alreja, A., Ward, M. J., Parker, L. S., Richardson, R. M., Morency, L.-P., Abel, T. J., & Ghuman, A. S. (2025). Neural encoding of real world face perception. ArXiv: arXiv:2505.08831v1.
- [11] Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>.
- [12] Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3), 395–412. <https://doi.org/10.1037/0022-3514.37.3.395>
- [13] Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>.
- [14] Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313–323. <https://doi.org/10.1016/j.cognition.2011.08.001>
- [15] Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, 25(7), 1404–1417. <https://doi.org/10.1177/0956797614532474>
- [16] Ramanathan, M., Peterson, L. M., & Palmer, C. J. (2025). Illusory facial expressions caused by lighting direction. *Journal of Experimental Psychology: General*, 154(1), 1–17. <https://doi.org/10.1037/xge0001697>
- [17] Edwards, B. H., Walden, D., Bloom, P. A., & Steinmetz, K. R. M. (2026). Investigating memory for faces based on emotional contextual information. *Memory & Cognition*, 54, 1650–1666. <https://doi.org/10.3758/s13421-025-01830-w>
- [18] Kurdi, B., Mann, T. C., & Ferguson, M. J. (2022). Persuading the implicit mind: Changing negative implicit evaluations with an 8-minute podcast. *Social Psychological and Personality Science*, 13(3), 688–697. <https://doi.org/10.1177/19485506211037140>.
- [19] Rhine, R. J., & Silun, B. A. (1958). Acquisition and change of a concept attitude as a function of consistency of reinforcement. *Journal of Experimental Psychology*, 55(6), 524–529. <https://doi.org/10.1037/h0048991>.
- [20] Kerpelman, J. P., & Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, 19(3), 301–305. <https://doi.org/10.1037/h0031447>.
- [21] Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, 91(6), 995–1008. <https://doi.org/10.1037/0022-3514.91.6.995>.
- [22] Phipps, D. J., Hagger, M. S., & Hamilton, K. (2025). Relative effects of implicit and explicit attitudes on behavior: A meta-analytic review and test of key moderators. *Psychological Bulletin*, 151(12), 1395–1447.

<https://doi.org/10.1037/bul0000506>

- [23] Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5), 295–306. [https://doi.org/10.1016/S0262-8856\(97\)00070-X](https://doi.org/10.1016/S0262-8856(97)00070-X)
- [24] Anderson, A. J., & Livingstone, M. S. (2024). The effect of caricaturing on the esthetic appeal of familiar faces, and its relation to simple proportion judgments. *i-Perception*, 15(6), 20416695241300099. <https://doi.org/10.1177/20416695241300099>
- [25] Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15(10), 674–679. <https://doi.org/10.1111/j.0956-7976.2004.00739.x>
- [26] Cao, R., Wang, J., Lin, C., De Falco, E., Peter, A., Rey, H. G., Brunner, P., Willie, J. T., DiCarlo, J. J., Todorov, A., Rutishauser, U., Li, X., Brandmeir, N. J., & Wang, S. (2025). Feature-based encoding of face identity by single neurons in the human amygdala and hippocampus. *Nature Human Behavior*, 9(9), 1959–1974. <https://doi.org/10.1038/s41562-025-02218-1>
- [27] 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. <https://doi.org/10.1016/j.tics.2008.10.001>