An analysis of the generalizability and stability of the Halo Effect during the COVID-19 pandemic outbreak

Giulio Gabrieli[1], Albert Lee[1], Peipei Setoh[1], Gianluca Esposito[2, 3*]

[1]Nanyang Technological University, Singapore, [2]Brain Science Institute, Nanyang Technological University, Singapore, [3]University of Trento, Italy

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Personality and Social Psychology

Article type:
Original Research Article

Manuscript ID:
631871

Received on:
21 Nov 2020

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

G.G. and G.E. conceptualized, designed and conducted the study. A.L. and P.S. revised the analytical method. G.G. drafted the manuscript, while all the authors contributed to the final version of the manuscript. G.E. supervised the project.

Keywords

halo effect, Aesthectics, trustworthiness, SARS-nCoV-2, ethnicity

Abstract

Word count: 273

The influence on the global evaluation of a person based on the perception of a single trait is a phenomenon widely investigated in social psychology. Mostly defined as Halo Effect, this phenomenon has been deeply studied for more than a hundred years now, and findings such as the relationship between aesthetic perception and other personality traits —like competence and trustworthiness— have since been uncovered. The latter plays an especially crucial role in individuals' social interactions. Despite the large body of literature published on the Halo Effect, and especially on the relationship between aesthetic appearance and perceived trustworthiness, little is known about the overall generalizability of the effect, as almost all of the studies have been conducted on adult participants from western countries. Moreover, little is known about the stability of the effect over time, in the event of major destabilization, such as the outbreak of a pandemic. In this work, the cross-cultural generalizability of the Halo Effect is investigated before and during the first few months of the COVID-19 pandemic. An analysis of the generalizability and stability over time of the Halo Effect is presented. Participants (N = 380, N = 145 Asians, N = 235 Caucasians) have been asked to rate the aesthetic and perceived trustworthiness of a set of human faces of different ages, gender, and ethnicity. Result of our analysis demonstrated that the Halo Effect (Aesthetic x Trustworthiness is influenced by the age of presented faces, but not by their gender or ethnicity. Moreover, our results show that the strength of the effect can be affected by external events and that the volatility is higher for adults than children's faces.

Contribution to the field

The manuscript presents a research work on the generalizability of the Halo Effect over different cultural groups. To the best of our knowledge, this is the very first study that consider age, gender, and ethnicity of presented faces simultaneously to investigate how they concur in shaping the Halo Effect (Aeshtetics x Trustworthiness). Moreover, this study is among the limited number of works that compares different ethnic groups within the study of the Halo Effect. Additionally, here we investigate the stability of the Halo Effect before and after the pandemic outbreak, providing a first investigation on how the Halo Effect is influenced by global events.

Funding statement

G.E. was supported by NAP SUG 2015, Singapore Ministry of Education ACR Tier 1 (RG149/16 and RT10/19).

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: The studies involving human participants were reviewed and approved by Institutional Review Board - Nanyang Technological University. The patients/participants provided their written informed consent to participate in this study.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.21979/N9/5IIVOM.

frontiers

An analysis of the generalizability and stability of the Halo Effect during the COVID-19 pandemic outbreak

Giulio Gabrieli [1], **Albert Lee** [1], **Peipei Setoh** [1] **and Gianluca Esposito** [1,2,3,*]

[1] *Psychology Program, School of Social Sciences, Nanyang Technological University, Singapore*
[2] *LKC School of Medicine, School of Social Sciences, Nanyang Technological University, Singapore*
[3] *Department of Psychology and Cognitive Science, University of Trento, Italy*

Correspondence*:
Gianluca Esposito
gianluca.esposito@ntu.edu.sg

ABSTRACT

The influence on the global evaluation of a person based on the perception of a single trait is a phenomenon widely investigated in social psychology. Mostly defined as *Halo Effect*, this phenomenon has been deeply studied for more than a hundred years now, and findings such as the relationship between aesthetic perception and other personality traits —like competence and trustworthiness— have since been uncovered. The latter plays an especially crucial role in individuals' social interactions. Despite the large body of literature published on the Halo Effect, and especially on the relationship between aesthetic appearance and perceived trustworthiness, little is known about the overall generalizability of the effect, as almost all of the studies have been conducted on adult participants from western countries. Moreover, little is known about the stability of the effect over time, in the event of major destabilization, such as the outbreak of a pandemic. In this work, the cross-cultural generalizability of the *Halo Effect* is investigated before and during the first few months of the COVID-19 pandemic. An analysis of the generalizability and stability over time of the *Halo Effect* is presented. Participants (N = 380, N = 145 Asians, N = 235 Caucasians) have been asked to rate the aesthetic and perceived trustworthiness of a set of human faces of different ages, gender, and ethnicity. Result of our analysis demonstrated that the *Halo Effect* (Aesthetic $\times$ Trustworthiness) is influenced by the age of presented faces, but not by their gender or ethnicity. Moreover, our results show that the strength of the effect can be affected by external events and that the volatility is higher for adults than children's faces.

Keywords: Halo Effect, Aesthetic, Trustworthiness, Ethnicity, SARC-CoV-2

1 INTRODUCTION

The *Halo Effect* (HE) is defined as the influence on the general evaluation of individuals' attributes or on the estimation of other traits based on the aesthetic appearance (Nisbett and Wilson, 1977). When applied to social perception, the Halo Effect Is observed when a single attribute is used as a basis for overall impression. For example, a stranger who looks good is also perceived as having good qualities (Todorov et al., 2009). As a subclass of the confirmation bias (Nickerson, 1998), the Halo Effect is known to be

inevitable, pervasive, constant, and ubiquitous (Cooper, 1981; Kozlowski et al., 1986; Feldman, 1986; Feeley, 2002). The Halo Effect is a widely investigated psychological phenomena, with impact on different academic fields such as Social Psychology, Computer Science, and Empirical Aesthetics (Todorov et al., 2009; Ferrari et al., 2017; Hartmann et al., 2008; Tuch et al., 2012).

1.1 Aesthetics and Trustworthiness

The impact of aesthetic appearance on perceived trustworthiness, also known as Halo Effect Aesthetics $\times$ Trustworthiness (Todorov et al., 2009), has been studied since the early years of the twentieth century. Unlike aesthetic appearance, trustworthiness is a global or "umbrella" trait that is fundamental to social perception (Fiske et al., 2007), with diverse implications in numerous life domains, such as in estimating another person's good or ill intentions.

An empirical investigation of the neurological pathways associated with trustworthiness judgments demonstrated a positive correlation between the brain activity of the area associated with implicit trustworthiness decision making and face perception. These results suggested the crucial role of the amygdala in retrieving information from the aesthetic appearance of faces (Winston et al., 2002; Adolphs et al., 1998).

Other works have replicated the impacts of aesthetic appearance on perceived trustworthiness, with more aesthetically/physically attractive individuals being perceived as more trustworthy. For example, in a study conducted by Carter (1978) on the appearance of counselors —a replication of a previous study conducted by Cash et al. (1975), revealed that attractive counselors are also perceived as more intelligent, warm, competent, and trustworthy. The strength of the effect was further confirmed in a review (Eagly et al., 1991), across 76 studies, aesthetic attractiveness was found to be positively linked with perceived social competence.

1.2 But is it always the case?

Despite the large body of literature on the relationship between aesthetic appearance and trustworthiness, several questions remain unanswered. Almost all of the available literature focused, in fact, on adult individuals sampled from the WEIRD population, rendering generalizability an issue (Jones, 2010; Henrich et al., 2010). Moreover, even though some studies have been conducted on children's faces, demonstrating that the effect exists in children (Dion 1972), there are limited comparisons on the impact of adults vs. children's targets. As child's faces are known to be special stimuli that automatically capture adults' visual attention and elicit parental care (Venturoso et al., 2019; Brosch et al., 2007; Proverbio and De Gabriele, 2019), the Halo Effect may be present with different strengths between adult and child faces. Finally, controversial results have been found for what concerns the importance of the rated individuals' gender. Significant differences between the scores given to males and females have been found in the works of Carter (1978), but not in others (Wetzel et al., 1981). One possibility for this is that in (Carter, 1978), there was an additional stereotype playing a part in the interaction, which is in people's mental representation of the stereotypical counselor (Chambers, 1983). To overcome the limitations of previous studies, this study aims to verify how the (a) ethnicity (ingroup vs outgroup), (b) age (adult vs baby), (c) gender (male vs female), and (d) aesthetic attractiveness concur in shaping trust perception. More specifically, in this work, we investigate the aesthetics and trustworthiness perception of Asians and Caucasians adults raters of both adults' and children's faces, both males and females, of Asians and Caucasians ethnicities.

The data collection stage of the project, with the methods described in the Study Design section, started in August 2019 and continued through April 2020. The data collection phase overlapped with the Covid-19 pandemic outbreak. Serendipitously, the data collected for this project allowed us to investigate the stability

over time of the halo effect as subjected to influence by COVID-19. One additional hypothesis —H_2— was therefore added to study such effects.

1.3 Aim & Hypothesis

We formulated two hypotheses. The first hypothesis, analytic plan, and method were pre-registered on the Open Science Framework; the second hypothesis was formulated after beginning the data collection.

H_1: *"Aesthetic attractiveness is positively correlated with perceived trust (Halo Effect). We predict the age of presented face to have an effect on the strength of the relationship, with the strength of the correlation higher for adults than for children' targets, but not the ethnicity or the gender of presented face."*

Rationale: Child faces capture greater attention compared to adults faces and to elicit parental care regardless of kinship (Brosch et al., 2007; Venturoso et al., 2019; Glocker et al., 2009; Parsons et al., 2011). Additionally, a recent study conducted by Collova et al. (2019) —based on a two dimensional model (trustworthiness and dominance) from Oosterhof and Todorov (2008)— investigated whether children's faces elicit the same signal threat responses of adults' faces. Results of Collova's studies revealed that adults rate children's faces on different dimensions then adults' faces. More specifically, when rating children's faces, the evaluation is not based on trustworthiness. These results suggest that children's faces should have little to do with in the perception of trust, regardless of how aesthetically attractive they are. If so, one should expect the relationship between aesthetic appearance and trustworthiness to be stronger for adults' ratings of adults', as compared to adults' ratings of children's faces. Therefore, we can expect the relationship between aesthetic appearance and trustworthiness to be stronger for adults' ratings of adults', as compared to adults' ratings of children's faces. From prior work, we know that gender (Wetzel et al., 1981) and ethnicity (Xu et al., 2012) do not seem to moderate the Halo Effect. But for the sake of completion, we decided to investigate these two demographic variables, with the expectation that neither gender nor ethnicity will have significant impact on our observed results. In line with previous studies, we do not expect to find a significant impact of gender on the strength of the effect. With regard to ethnicity, although differences may be present in the aesthetic ratings given to individuals of the ingroup or of the outgroup, as the implicit judgment of trustworthiness is based on the elaboration of facial cues in the amygdala, no differences between the strength of the effect for the ingroup and outgroup is expected.

H_2: *"When individuals are asked to rate the aesthetic and trustworthiness of others' faces, we expect to see changes in the variability of the ratings after the diffusion of news about COVID-19 in trustworthiness but not aesthetic judgments toward adults but not children's faces."*

Rationale: In a study conducted by Xu et al. (2012), it was reported that when making inferences about the trustworthiness of others from their aesthetic appearance, Chinese and Caucasians both adopt the same strategies. Despite that, Asians and Caucasians faces have been reported to be perceived as distinct categories (Zhou et al., 2020). However, Koopmans and Veit (2014) reported that negative interethnic contacts result in reduced trust toward members of the outgroups. In light of the COVID-19 pandemic global threat, following the diffusion of news about the spreading of the novel coronavirus in China, and with politicians targeting a specific ethnic group (Zheng et al., 2020), we can expect the situation to act as negative interethnic contact for Caucasian individuals, who may therefore reduce their estimation of trustworthiness, but not of aesthetics, toward Asians adults' faces. For what concerns Asian individuals, previous research work by Fincher et al. (2008) highlighted that regions with a stronger history of contagious diseases are more likely to adopt collectivistic behaviors, including outgroup hostilities. It is therefore possible that, with the development of the outbreak in western countries, together with the adoption of specific measures to counter the diffusion of the virus in eastern countries, collectivist beliefs induced a

reduction in the perceived trustworthiness, but not aesthetics, of Asians participants toward Caucasians. These findings suggest that salient threats of contagion, such as the case during the COVID-19 pandemic, may elicit the tendency to bind strongly with familliar ingroups and reject unfamiliar outgroups. This tendency, given its strong evolutionary undertone, should be present in most people regardless of their cultures. Applying this assumption to this hypothesis, one should expect a global reduction of trust in the perception of adult faces, regardless of the cultural backgrounds of these adult faces. Such global reduction, however, should not be observed in the aesthetic perception, which has nothing to do with threat of contagion. Taken together, these findings suggest that we should see a generalized reduction of trust, but not aesthetics, toward both Asians and Caucasians adults' faces. For what concerns children's faces, a different situation is expected. In an ERP a study conducted by Proverbio and De Gabriele (2019), it is reported that the other-race effect does not apply to infants' faces, supporting the specificity of the age of a face over its ethnicity for younger's faces. Differences in adults' perception of adults' and children's faces in Other-Race Effects studies were also reported by Kuefner et al. (2008), in a series of three experimental studies. These findings suggest that the salience of infants and children faces should limit the impact of race on the estimation of other traits. Building on the work from Collova et al. (2019) reported above (H_1) we can expect an early evaluation of infants' faces not to have an influence on perceived trustworthiness. Taken together, findings on the specificity of infants' and children's faces suggest that the age dimension plays a prominent role, more than the possible perceived threat dimension, in the early evaluation of children's faces. It is therefore possible that, when presented with faces of children, adults' trustworthiness judgments are less likely to be influenced by the aesthetics traits of a child's faces, as compared to when they are rating an adult's face. From a biological point of view, this behavior would reflect mammals, and especially humans, altruistic responses toward infants (Preston, 2013). In fact, the protection of the offspring of the species may come at a cost for the self, which can be trusting children even though they may be possible carriers of the pathogen. Consequently, we do not expect any difference in the judgment of both the aesthetic and trustworthiness of children's faces before and during the initial stages of the COVID-19 pandemic outbreak.

2 METHODS

2.1 Participants

The study was approved by the Internal Review Board of Nanyang Technological University (PSY-IRB-2019-008 and IRB-2019-10-019) and conducted according to the declaration of Helsinki. Informed consent was obtained from all the participants before the study. Participants (N = 380, M age = 25.0 ± 8.49) voluntarily participated and were recruited through the Nanyang Technological University School of Social Sciences Research Participation System or through different social media. The gender and ethnicity of participants are reported in Table 1.

Table 1. Participants' demographic information

Ethnicity	Gender	N	Age
Asian	Male	75	22.5 ± 1.83
	Female	70	21.0 ± 3.06
Caucasian	Male	80	29.0 ± 11.81
	Female	155	26.0 ± 8.99

2.2 Study Design

2.2.1 Stimuli

Participants were presented with 64 faces of two different age groups (32 adults, 32 children), genders (32 males, females), and ethnicities (Asians, Caucasians). This structure allowed for the presentation of 8 faces per combination of age, gender, and ethnicity (e.g. 8 adult Asian male faces). Front-facing images of faces (N = 64) were selected from the FFHQ Dataset(Karras et al., 2019), a dataset containing 70,000 high-quality (1024×1024) images published on Flickr2, an online photo management, and sharing tool, under different creative commons and public domain licenses (Creative Commons BY 2.0, Creative Commons BY-NC 2.0, Public Domain Mark 1.0, Public Domain CC0 1.0, or U.S. Government Works license). The dataset itself is released under the Creative Commons BY-NC-SA 4.0 license by NVIDIA Corporation and has been successfully used in previous publications (Karras et al., 2019; Wang et al., 2019; Kynkäänniemi et al., 2019; Zhao et al., 2020). Selected faces were presented in random order, with no time constraints.

2.2.2 Procedure

After having signed the informed consent, participants were instructed about the scope and procedure of the experiment, as well as the taxonomy employed in the study. Participants rated each face for aesthetic pleasantness (*"How much do you like this person?"*) and trustworthiness (*"How much do you trust this person?"*) on a 100-point sliding scale, anchored from 1 being "not at all" not to 100 = "extremely".

2.3 Analytic plan

The analytic plan was pre-registered on the Open Science Framework. Additional information can be found online on the Open Science Framework (https://osf.io/5cge3). A power analysis was conducted to estimate the number of participants required for this study (H_1). Given that previous works have found the effect size for the Halo Effect of human faces to be of medium strength, to take into account a possible bias in published works (Camerer et al., 2018; Collaboration, 2015), we assumed a very weak effect size to estimate the required number of participants. Assuming six groups (children/adult, male/female, Asian/Caucasian), a very weak effect size (Cohen's d = 0.1), and to achieve a power of 0.95 at a 0.05 alpha value, a power analysis conducted in G*Power (Faul et al., 2007, 2009) revealed that three hundred and thirty (N = 330) participants are required to perform an analysis of variance. The strength of the Halo Effect is measured as the Pearson's correlation between aesthetic and trustworthiness ratings. To test H_1, a 3×2 Analysis of Variance was employed to control for the existence of significant effects of gender, age, and ethnicity on the strength of the Halo Effect. A z-test is employed as a posthoc test to test whether the Halo Effect is stronger for adults than children faces. Additionally, a confirmatory analysis is conducted by means of a multiple linear regression analysis.

For what concerns the second hypothesis (H_2), four Levene's tests for equality of Variance have been conducted on Aesthetics and Trustworthiness, comparing the variance of data collected before and after the diffusion of news about the novel coronavirus, once for adults' and one for children's faces. As a threshold, we used February 1st, 2020, which is, according to *Google Trend*2, the moment in which people started to show interest toward the SARS-nCOV-2. In order for H_2 to be verified, we expected significant differences in the variance of trustworthiness ratings towards adults' faces before and after our threshold date, but not for adults' faces aesthetics ratings, nor for both aesthetics and trustworthiness ratings toward children's faces. To take into account the multiple numbers of tests conducted, a correction for multiple tests using the Benjamini–Hochberg procedure, with a False Discovery Rate of 0.10, is employed.

3 RESULTS

3.1 Effect of ethnicity, age, and gender on the strength of the Halo Effect

To evaluate the effects of ethnicity, age, and gender on the strength of the Halo Effect, an Analysis of Variance has been conducted. Results of the Analysis of Variance revealed only a main effect of age (F-value = 9.753, p-value = 0.00194) but no main effect of Gender or Ethnicity, as well as no significant effects of the interaction between the factors on the strength of the Halo Effect (Aesthetics $\times$ Trustworthiness). These results suggest that the strength of the relationship between Aesthetics and Trustworthiness (Pearson's r = 0.676, p < 0.001) is influenced by the age of presented faces, which is whether it is a child or an adult face, but not by its gender or ethnicity. Taken together, the findings suggest that, at a general level, when adults rater makes inferences about others' aesthetic and trustworthiness, do not differentiate between people of different gender or ethnicity, but do adopt different strategies for adults and children.

More specifically, the strength of the relationship between Aesthetics and Trustworthiness is significantly higher (z-test t = 3.626, p-value = 0.000287, Figure 1 and 2) for adult (M = 0.53 ± 0.41) than for children faces (M = 0.47 ± 0.46). These results indicate that adults are more likely to estimate the trustworthiness of other adults from their aesthetic appearance, while the estimation is less consistent when it comes to predicting the trustworthiness of children from their appearance.

Additionally, the strength of the relationship between the two variables has been further confirmed using a multiple linear regression analysis, with the formula reported in Equation 1. Results are reported in Table 2.

$$Trustw. = Int. + Aesthetics \times X_1 + Age \times X_2 + Gender \times X_3 + Ethnicity \times X_4 \quad (1)$$

A subsequent exploratory analysis revealed that the effect is significantly stronger for Asian participants, as compared to Caucasian Participants (t=13.2, uncorrected p-value=$9.68 \cdot 10^{-39}$). Further exploring the difference between Asian and Caucasian participants, both groups show no significant differences in the *Halo Effect* elicited by younger faces of their same ingroup and outgroup (Asian participants: t = -0.67, uncorrected p-value = 0.503; Caucasian participants: t = -0.935, uncorrected p-value = 0.351). However, when comparing the strength of the *Halo Effect* in Asian and Caucasian participants rating adults of their ingroup and outgroup, while no differences have been found in the strength of the *Halo Effect* in Asian participants (t = -1.551, uncorrected p-value = 0.122). On the other hand, the *Halo Effect* of Caucasian participants was significantly higher for adult faces of their ingroup as compared to their outgroup (t = 4.026, uncorrected p=value = $6.697 \cdot 10^{-05}$).

Table 2. Results of the Multiple Linear Regression used to investigate the strength of the Halo Effect and the influence of Age, Gender, Ethnicity, and Aesthetic on Trustworthiness.

	Coeff.	**std. err**	**t**	**P $>$ |t|**	**C.I.**
Intercept	7.4930	1.007	7.445	0.000*	[5.519, 9.467]
Ethnicity	-0.2078	0.533	-0.390	0.697	[-1.253, 0.838]
Gender	0.4762	0.534	0.839	0.372	[-0.570, 1.522]
Age	5.1196	0.554	9.243	0.000*	[4.034, 6.206]
Aesthetic	0.7797	0.014	55.726	0.000*	[0.752, 0.807]

3.2 Effect of SARS-CoV-2 on the strength of the Halo Effect over time

Results of the comparison between the variability in Aesthetics and Trustworthiness judgments toward both Adults' and children's faces are reported in Table 3. Results (q-values) highlight significant changes in the variability of Trustworthiness ratings towards adults' faces before and after the beginning of the COVID-19 pandemic outbreak, but not in Aesthetics ratings given to adults' faces, nor to Aesthetics or Trustworthiness ratings given to children's faces.

Table 3. Results of Levene's test of Variance for Aesthetics and Trustworthiness judgments toward adults' and children' faces (q-values are evaluated using Benjamini–Hochberg procedure at a .10 False discovery rate).

Age	Variable	Statistic	Uncorrected p-value	q-value
Adult	Aesthetics	4.633	0.034	0.05
	Trustworthiness	5.557	0.021	0.025
Children	Aesthetics	2.077	0.105	0.1
	Trustworthiness	3.861	0.053	0.075

4 DISCUSSION

Based on previous works within the field of the *Halo Effect*, we hypothesized that the impact of perceived aesthetic on trustworthiness judgments would depend on the age of presented faces, but not on their gender or ethnicity (H_1). Results of the Analysis of Variance show the main effect of the Age of presented faces but not of Gender or Ethnicity, nor of any interaction effect, confirming H_1. Moreover, our posthoc z-test confirmed that the relationship between Aesthetics and Trustworthiness is stronger for adults' as compared to children's faces. In tight of the results here presented, our analysis supports the specificity of children's faces. In fact, only the age of the presented face but not the gender or age influence in our sample the strength of the Halo Effect, measured as the Pearson correlation between individuals' aesthetic appearance and perceived trustworthiness. As reported in previous works on the Baby Schema effect (Venturoso et al., 2019), younger faces elicit specific responses in adult viewers. A possible explanation for this may be drawn from the evolutionary perspective. In fact, the cure of the offspring plays a central role in the survival of the species, and therefore adult individuals may be more prone to trust a younger individual even though the perceived aesthetic appearance is low. On the other hand, when looking at adult faces, the evaluation of someone's trustworthiness is solely made on the basis of the appearance.

Our exploratory analysis further confirmed the specificity of children faces. In fact, both Caucasian and Asian participants revealed no significant differences in the strength of the *Halo Effect* when exposed to either children of their same ingroup or of their outgroup. While the same can be said for what concerns Asian adults looking at Asian and Caucasian adult faces, the same can not be said for the Caucasians in our pool of participants, who indeed showed significant differences in the strength of the *Halo Effect* when exposed to faces of other Caucasians (higher Halo) as compared to adult Asians (lower halo). This confirms previously published results on both the specificity of children faces, and significant differences in adults' activation Esposito et al. (2014). While this goes beyond the initial plan of this work, and has been in fact not treated as hypothesis confirmation but as exploratory analysis, take together the findings here reported about the *Halo Effect* are in line with previous works that investigated cross-cultural differences across Asians and Caucasians with different methodologies. Future works should investigate significant differences between the strength of the Halo in Asian and Caucasian participants by properly defining one

or more hypothesis and by recruiting an adequate number of participants to verify novel hypotheses with adequate power.

For what concerns the stability of the Halo Effect over time (H_2), the analysis of the variance conducted in this study, revealed that adults' Trustworthiness ratings, but not Aesthetics ratings, are influenced by the diffusion of news about the novel coronavirus. Differently, no changes are found in the judgments of children's' Aesthetics and Trustworthiness. These results are in line with our predictions on the specificity of children's' faces. While our results confirm the possibility of modulating the strength of the Halo, the current dataset does not allow the study of the qualitative impact of an external event. Future studies should address this problem by empirically presenting the external events, using a priming procedure, and measuring the impact of it over time with a longitudinal approach

Despite the strength of the results here presented, there are several limitations worth highlighting. As reported earlier in this work, the data collection stage started before and continued during the novel coronavirus pandemic outbreak. As reported here, significant differences were found in the trustworthiness ratings given to adults faces before and during the pandemic outbreak. Therefore, while our first hypothesis (H_1) has been empirically verified accordingly to our preregistered plan, we can not exclude that the overall world's situation played a role in shaping our results. future works should investigate the stability of the effect under a controlled condition, such as by using a prime. Moreover, while we targeted Asian and Caucasian participants, we have not investigated the influence of participants' ethnicity at a more specific level (e.g. Chinese, Japanese, Korean, etc.). Future studies should focus on single ethnic group, to verify the consistency and generalizability of the results here presented. Additionally, while participants were informed of the scope of the experiment, including the fact that we were specifically interested in Aesthetic Appearance, participants' whose first language is not English may not have a specific counterpart for this concept. Future works should provide participants' whit questions posed in their native langauge.

5 CONCLUSION

In this work, we investigated the generalizability and stability over time of the *Halo Effect* (Aesthetic $\times$ Trustworthiness). Our results show that the strength of the correlation between the perceived Aesthetic and Trustworthiness of strangers' faces is affected by the age of presented faces, but not by their ethnicity or gender. These results support the specificity of children faces, as well as the consistency of the *Halo Effect* across cultures. Additionally, our results show that when a major event that disrupts people's perception of others is presented, such as the SARS-CoV-2 pandemic outbreak, the strength of the association between perceived Aesthetics and Trustworthiness is less stable for adults' as compared to children's faces. Taken together, our results of our work confirm the generalizability of the Halo Effect across cultures, as well as the specificity of children's faces. Future studies should investigate the effect on smaller subgroups (e.g. Japanese versus Chinese) and in period of times where there is a limited influence of external events on others' judgment.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

G.G. and G.E. conceptualized, designed and conducted the study. A.L. and P.S. revised the analytical method. G.G. drafted the manuscript, while all the authors contributed to the final version of the manuscript. G.E. supervised the project.

FUNDING

G.E. was supported by NAP SUG 2015, Singapore Ministry of Education ACR Tier 1 (RG149/16 and RT10/19).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Data Repository of the Nanyang Technological University (DR-NTU Data) at the following link: https://doi.org/10.21979/N9/5IIVOM.

REFERENCES

Adolphs, R., Tranel, D., and Damasio, A. R. (1998). The human amygdala in social judgment. *Nature* 393, 470–474

Brosch, T., Sander, D., and Scherer, K. R. (2007). That baby caught my eye... attention capture by infant faces. *Emotions*

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour* 2, 637

Carter, J. A. (1978). Impressions of counselors as a function of counselor physical attractiveness. *Journal of Counseling Psychology* 25, 28

Cash, T. F., Begley, P. J., McCown, D. A., and Weise, B. C. (1975). When counselors are heard but not seen: Initial impact of physical attractiveness. *Journal of Counseling Psychology* 22, 273

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science education* 67, 255–265

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science* 349, aac4716

Collova, J. R., Sutherland, C. A., and Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of personality and social psychology*

Cooper, W. H. (1981). Ubiquitous halo. *Psychological bulletin* 90, 218

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., and Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin* 110, 109

Esposito, G., Nakazawa, J., Ogawa, S., Stival, R., Kawashima, A., Putnick, D. L., et al. (2014). Baby, you light-up my face: culture-general physiological responses to infants and culture-specific cognitive judgements of adults. *PloS one* 9, e106705

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 1149–1160

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

Feeley, T. H. (2002). Comment on halo effects in rating and evaluation research. *Human Communication Research* 28, 578–586

Feldman, J. M. (1986). A note on the statistical correction of halo error. *Journal of Applied Psychology* 71, 173

Ferrari, C., Nadal, M., Schiavi, S., Vecchi, T., Cela-Conde, C. J., and Cattaneo, Z. (2017). The dorsomedial prefrontal cortex mediates the interaction between moral and aesthetic valuation: a TMS study on the

beauty-is-good stereotype. *Social Cognitive and Affective Neuroscience* 12, 707–717. doi:10.1093/scan/nsx002

Fincher, C. L., Thornhill, R., Murray, D. R., and Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences* 275, 1279–1285

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

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Valdez, J. N., Griffin, M. D., et al. (2009). Baby schema modulates the brain reward system in nulliparous women. *Proceedings of the National Academy of Sciences* 106, 9115–9119

Hartmann, J., Sutcliffe, A., and Angeli, A. D. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 1–30

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Beyond weird: Towards a broad-based behavioral science. *Behavioral and Brain Sciences* 33, 111

Jones, D. (2010). *A WEIRD view of human nature skews psychologists' studies* (American Association for the Advancement of Science)

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4401–4410

Koopmans, R. and Veit, S. (2014). Ethnic diversity, trust, and the mediating role of positive and negative interethnic contact: A priming experiment. *Social science research* 47, 91–107

Kozlowski, S. W., Kirsch, M. P., and Chao, G. T. (1986). Job knowledge, ratee familiarity, conceptual similarity and halo error: An exploration. *Journal of Applied Psychology* 71, 45

Kuefner, D., Macchi Cassia, V., Picozzi, M., and Bricolo, E. (2008). Do all kids look alike? evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance* 34, 811

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. (2019). Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*. 3927–3936

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 175–220

Nisbett, R. E. and Wilson, T. D. (1977). The halo effect: evidence for unconscious alteration of judgments. *Journal of personality and social psychology* 35, 250

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

Parsons, C. E., Young, K. S., Kumari, N., Stein, A., and Kringelbach, M. L. (2011). The motivational salience of infant faces is similar for men and women. *PloS one* 6, e20632

Preston, S. D. (2013). The origins of altruism in offspring care. *Psychological bulletin* 139, 1305

Proverbio, A. M. and De Gabriele, V. (2019). The other-race effect does not apply to infant faces: An erp attentional study. *Neuropsychologia* 126, 36–45

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

Tuch, A. N., Roth, S. P., HornbæK, K., Opwis, K., and Bargas-Avila, J. A. (2012). Is beautiful really usable? toward understanding the relation between usability, aesthetics, and affect in hci. *Computers in Human Behavior* 28, 1596–1607

Venturoso, L., Gabrieli, G., Truzzi, A., Azhari, A., Setoh, P., Bornstein, M. H., et al. (2019). Effects of baby schema and mere exposure on explicit and implicit face processing. *Frontiers in Psychology* 10, 2649

Wang, R., Ma, L., Juefei-Xu, F., Xie, X., Wang, J., and Liu, Y. (2019). Fakespotter: A simple baseline for spotting ai-synthesized fake faces. *arXiv preprint arXiv:1909.06122*

Wetzel, C. G., Wilson, T. D., and Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology* 17, 427–439

Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature neuroscience* 5, 277–283

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., and Lee, K. (2012). Similarities and differences in chinese and caucasian adults' use of facial cues for trustworthiness judgments. *PLoS One* 7, e34859

Zhao, S., Liu, Z., Lin, J., Zhu, J.-Y., and Han, S. (2020). Differentiable augmentation for data-efficient gan training. *Advances in Neural Information Processing Systems* 33

Zheng, Y., Goh, E., and Wen, J. (2020). The effects of misleading media reports about covid-19 on chinese tourists' mental health: a perspective article. *Anatolia* 31, 337–340

Zhou, Y., Gao, T., Zhang, T., Li, W., Wu, T., Han, X., et al. (2020). Neural dynamics of racial categorization predicts racial bias in face recognition and altruism. *Nature human behaviour* 4, 69–87

FIGURE CAPTIONS

Figure 1. Strength of the Halo Effect (*pearson-r*) by Age

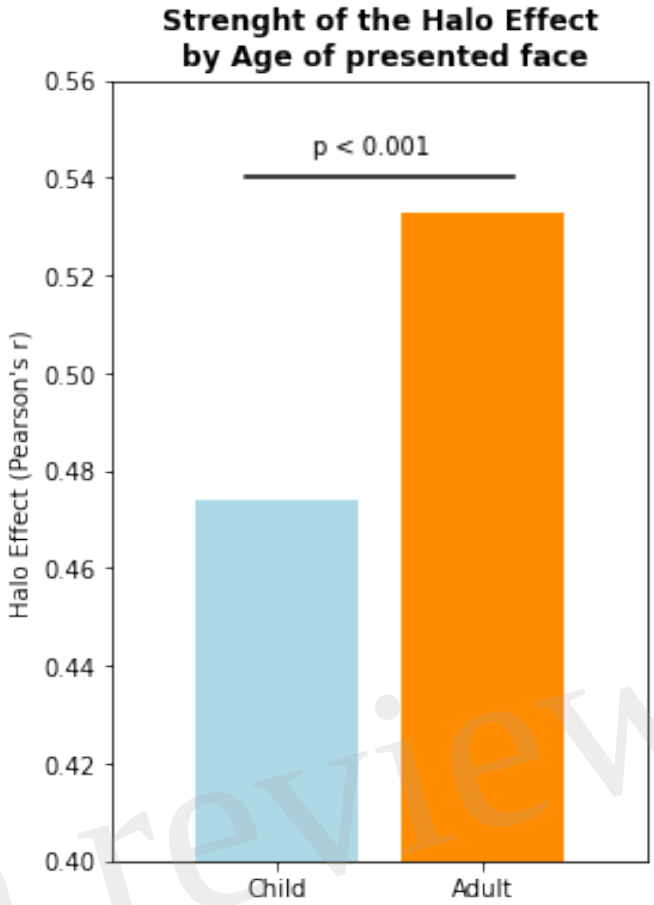

Figure 2. Distribution of Aesthetic and Trustworthiness judgments by Age

Figure 1.JPEG

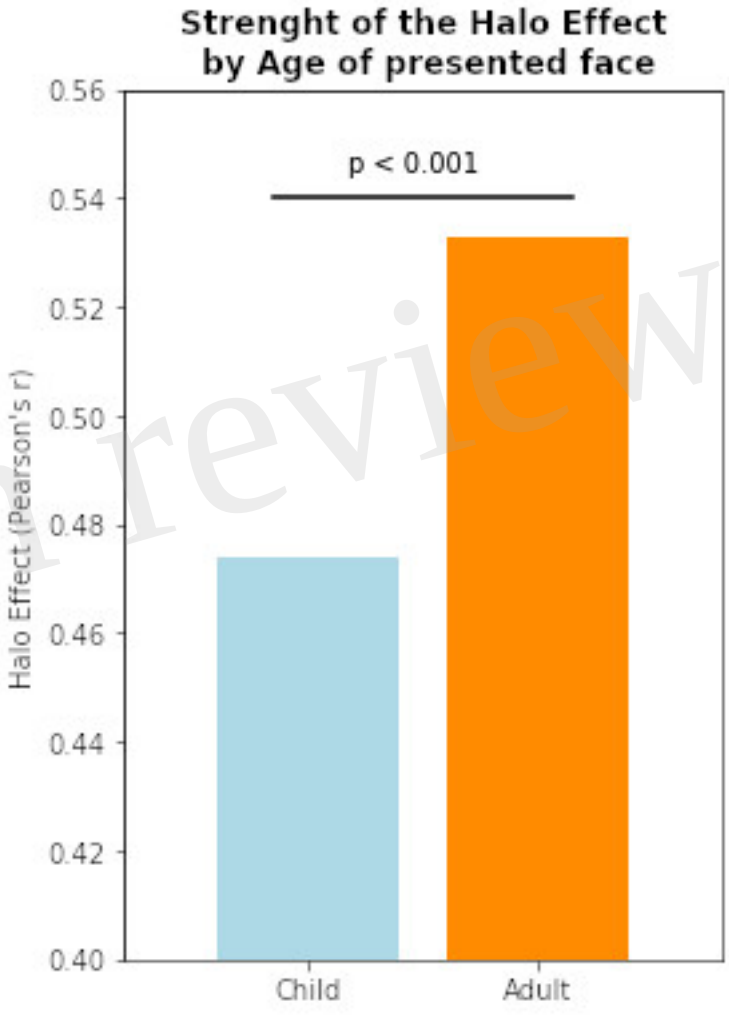

Figure 2.JPEG