Updating Stereotypical Attributions in Light of New Information: The Attractiveness Halo Effect Changes When Attractiveness Changes

Marine Rougier and Jan De Houwer

Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

Author Note

This research was made possible by the Methusalem grant BOF22/MET_V/002 of Ghent University to Jan De Houwer and by the LEARVUL H2020 Twinning Project No. 952464/2020. Correspondence should be addressed to Marine Rougier, Department of Experimental-Clinical and Health Psychology, Ghent University, H.Dunantlaan 2, 9000 Ghent, Belgium. Email: Marine.Rougier@UGent.be.

The pre-registration files, materials (stimuli and JsPsych code), data, and analytic (R) scripts for all experiments are made publicly available at https://osf.io/9wcmf/?view_only=fd1c6274fdf242059c34bd7ec6554220. Authors report having no conflict of interest in publishing this work. Studies received approval (number 2021/39) from the ethical committee of the Faculty of Psychology and Educational Sciences at Ghent University. We would like to thank Morgan Steurs for the comments and help she provided on Studies 1 and 2 in the context of her master thesis.

Abstract

In the attractiveness halo effect, a single known piece of information about a target stimulus (attractiveness of a person) influences assumptions about a host of other attributes about that target (e.g., this person is socially competent or vain). We examined for the first time whether this effect can be updated, that is, whether new information about physical attractiveness (e.g., that someone is not as attractive as initially thought) can undo the effects of earlier information. Across three preregistered experiments (n = 1131), we obtained evidence of a halo-update effect and showed that updating depended on the extent to which personality traits are stereotypically related to attractiveness (i.e., updating was larger for the traits that are typically influenced by attractiveness information). We also explored potential mediators of the halo-update effect. By shedding new light on the malleability of stereotypical attributions, our work has both theoretical and practical implications.

Keywords: Impression formation; Person perception; Conditioning; Halo effect; Updating; Stereotypes.

Updating Stereotypical Attributions in Light of New Information: The Attractiveness Halo Affect Changes When Attractiveness Changes

People can form impressions very quickly and easily on the basis of minimal information. For instance, research on the attractiveness halo effect shows that variation in the attractiveness of target persons typically leads to stereotypical assumptions on a myriad of positive and negative personality traits (e.g., Bassili, 1981; Dion et al., 1972; Eagly et al., 1991).

Impressions, however, sometimes need to be revised in light of new information. For instance, imagine chatting with a person online while only seeing a photograph of that person. In line with the attractiveness halo effect, you may form an impression and make stereotypical assumptions based on the fact that the person looks attractive (e.g., that s/he is sociable and vain). Imagine that later on you meet the same person in real life and realize that s/he is much less attractive than on the photograph. Will you change the stereotypical assumptions that you initially made (e.g., no longer assume that the person is sociable or vain)? In three preregistered studies, we examine for the first time whether the attractiveness halo effect can be updated when new information about attractiveness is provided.

We first discuss past research on updating effects in both conditioning and impression formation research. We then argue that the (attractiveness) halo-update effect goes beyond existing evidence in several interesting ways. Most notably, our research enriches current models of stereotyping by demonstrating new ways in which stereotyping can be malleable.

Updating in Conditioning and Impression Formation

Although conditioning and impression formation phenomena emerged from distinct literatures, both relate to how assumptions about unknown or ambiguous features of stimuli in our environment are shaped by known aspects of the environment (De Houwer et al., 2019). In evaluative conditioning (EC), the evaluation of an initially neutral stimulus (CS;

e.g., a neutral face) changes as the result of pairing it with a liked or disliked stimulus (US; e.g., a liked person; De Houwer, 2007; see also Förderer & Unkelbach, 2011, 2015). Likewise, in impression formation, assumptions about a target person can emerge from very limited information such as the person's visual appearance (e.g., Dion et al., 1972; Oosterhof & Todorov, 2008; Willis & Todorov, 2006), behavior (e.g., Carlston et al., 1995; Cone & Ferguson, 2015), or group membership (e.g., Fiske et al., 2002). [1]

In both research areas, when new information is provided that could lead to a reassessment of the initial information, effects are typically attenuated or even reversed (e.g., Cone & Ferguson, 2015; Walther et al., 2009). In EC, this is illustrated by the US-revaluation effect: the conditioned change in CS evaluation is reduced or reversed when the likeability of the US is revaluated (Baeyens et al., 1992; Sweldens et al., 2010; Walther et al., 2009). For instance, in research by Baeyens et al. (1992), neutral face pictures (CSs) were paired with (dis)liked face pictures (USs) in an EC procedure. After the CS-US pairings, the (dis)liked faces were paired with new information of opposite valence (e.g., a liked face was paired with the trait "egoistic") or with information of consistent valence (e.g., trait "friendly"). The observed EC effect was reduced in the opposite (vs. consistent) valence condition (see also Förderer & Unkelbach, 2016).

In impression formation, similar updating effects have been studied in the context of judgment and attitude formation. Early research on belief updating investigated the effect of mixed pieces of information (e.g., different scenarios on the efficiency of a coaching program) on judgments related to the manipulated feature (e.g., whether it increased performance; Hogarth & Einhorn, 1992). More recent work focused on updating attitudes. For instance, Cone and Ferguson (2015) asked their participants to form an impression about Bob based on behavioral statements about him (e.g., "gave a hitchhiker a ride to a shelter"). After measuring a first evaluation of Bob, they provided additional information of opposite

valence that should lead to an "attitude update" (e.g., "was recently convicted of molesting children") or neutral information (e.g., "recently bought a soda"). A second evaluation revealed a change in evaluation in the direction of the new information (e.g., weakened positive evaluation after new negative behavioral information; for procedural variations, see Mann & Ferguson, 2015; Mann et al., 2020). In the same vein, face-based first impressions (e.g., untrustworthy face) can also be altered by providing additional information about the target's behavior (e.g., trustworthy or pleasant behavior; Shen & Ferguson, 2021; Shen et al., 2020; see also McConnell et al., 2008).

The Halo-Update Effect and Its Trait Selectivity

In the present paper, we focus on updating the halo effect. In the halo effect, a positive characteristic of a stimulus influences how the stimulus is perceived on other dimensions for which no information is available (Forgas & Laham, 2016). We relied on the specific case of the attractiveness halo effect, where the attractiveness of a person influences how the person is perceived on a series of different personality traits (e.g., social competence; Eagly et al., 1991). Although it was initially assumed that only positive traits are impacted by attractiveness (cf. "what is beautiful is good" idea), research showed that also assumptions about negative traits can be influenced (e.g., vanity; Bassili, 1981; Dermer & Thiel, 1975; Han & Laurent, 2023). Interestingly, the attractiveness halo effect is also trait-selective as some traits are influenced more strongly than other traits (Bassili, 1981; Eagly et al., 1991). For instance, Rougier et al. (2023) relied on a series of 42 personality traits to assess the attractiveness halo effect. In line with previous literature, they observed that sociability- and vanity-related traits produced a larger halo effect than integrity- and intelligence-related traits (i.e., the difference in ratings between high and low attractive faces was larger for the former set of traits; see also Han & Laurent, 2023).

The attractiveness halo effect thus differs in several ways from the previously discussed conditioning and first impression effects in that: 1) the manipulated information (i.e., attractiveness) and the outcome measure(s) (e.g., sociability) differ, 2) the effect cannot be reduced to mere change in valence (i.e., attractiveness influences both positive and negative traits), and 3) it is trait-selective (e.g., attractiveness influences more sociability than intelligence). In sum, the attractiveness halo effect reveals a selective, non-attitudinal pattern of attributions. In this work, we examined for the first time the halo-update effect, that is, whether this selective pattern of change can be weakened or reversed when the initial information about attractiveness is updated. Thus, we tested whether updating attractiveness (e.g., learning that a person is not as attractive as initially thought) results in updating both positive and negative attributions (e.g., sociability and vanity attributions) and whether this updating effect is larger for the traits that are usually the most influenced by attractiveness (e.g., larger for sociability than intelligence).

Investigation of selectivity in the halo-update effect was based on the data collected by Rougier et al. (2023). Relying on 42 target personality traits, they gauged the extent to which each trait is likely to reveal a halo effect, that is, the extent to which it is "relevant" for attractiveness. They also measured the extent to which those traits are generally perceived as positive (vs. negative). From these two measurements, they computed a "trait halo relevance" score (i.e., average difference of rating between high vs. low attractive individuals) and a "trait valence" score (i.e., average positivity rating) for each trait. An increase in the halo effect (i.e., average difference in ratings between high and low attractive faces) as a function of increases in trait halo relevance thus reflects the typical trait selectivity of the halo effect. However, an increase in the halo effect as a function of increases in trait valence reflects judgments based on mere valence (i.e., the more positive the traits, the more it differentiates between high and low attractive faces).

We thus tested whether the halo and trait selectivity effects (i.e., the moderation of the halo effect by the trait halo relevance score) are influenced by updating. Alternatively, updating attractiveness might not change the typical pattern of assumptions but merely the overall positive vs. negative perception of the person, in line with previous updating effects in EC (Baeyens et al., 1992) and impression formation (Cone & Ferguson, 2015). Specifically, after learning that a high attractive face is in fact less attractive, people might form a negative impression of that person without necessarily making trait-specific assumptions. Changes in positive vs. negative perception would then translate in a particular, valence-based, trait selectivity: traits that are extremely valenced (i.e., extremely positive vs. negative) would be the most influenced by the updating information as compared to more neutral traits (i.e., we should observe a moderation of the halo-update effect by trait valence instead of trait halo relevance). Overall, considering the trait halo relevance and trait valence scores enabled us to test directly whether the pattern of attributions generally observed in the halo effect can be updated or whether the halo-update effect is merely determined by how positively/negatively the targets were perceived.

Implications for Stereotyping

In cognitive terms, the halo effect and its typical trait selectivity are often interpreted as an illustration of stereotyping. For instance, an attractive face would activate the group category of “high attractive people”, in turn leading to specific assumptions about (stereotyping of) the personality of this group member (e.g., this person is sociable because s/he belongs to the attractive people group; Ashmore, 1981). Surprisingly, research on the malleability of stereotyping effects is somewhat limited. On the one hand, it is well known that the way we categorize a target individual (e.g., as a woman vs. as teacher) determines the nature of stereotypical attributions (e.g., Macrae et al., 1995; van Knippenberg et al., 1994). Hence, re-categorizing (i.e., changing the category of a target) or individuating (i.e., focusing

on the individual beyond the category membership) a person should change stereotypical attributions toward that individual because the content of the activated stereotype changes (i.e., stereotype activation varies; Fiske & Neuberg, 1990; Kunda & Spencer, 2003). In line with this idea, research on subtyping or subgrouping shows that when individuals do not correspond to the stereotype about a social group (e.g., Black people), they are moved out of the category or placed in a sub-cluster within the category (e.g., Black lawyer; Maurer et al., 1995; Richards & Hewstone, 2001).

The halo-update effect, on the other hand, aims at changing the *value* (or level) within the same category (e.g., informing that a seemingly high attractive target is in fact low on attractiveness) rather than changing the category itself (i.e., individuals are not re-categorized, de-categorized, or categorized in a subgroup). In this context, a change of category value could decrease or reverse the pattern of stereotypical attributions (i.e., moderation of trait selectivity typical of halo). Thus, whereas past work showed malleability in terms of the *content* of the stereotypical attributions (i.e., which stereotypical traits are assumed), the halo-update effect would demonstrate malleability in the *degree or direction* of stereotyping (i.e., the extent to which stereotypical traits are assumed).

At the practical level, the halo-update effect relates to multiple real-life contexts in which initial information turns out to be incorrect (e.g., impression management, fake news, deepfakes, social media filters). Exploring new ways to change stereotypical attributions is an important mission, as some of them can lead to unfair treatment. In the case of the attractiveness halo effect, for instance, attractive people are judged less likely to be guilty and are sentenced to lighter prison terms (Efran, 1974) and, in the case of women, less likely to be hired for managerial positions (Heilman & Stopeck, 1985).

Overview

Across three studies, participants were presented with photographs of faces varying on attractiveness (e.g., low vs. high attractive faces). After forming a first impression, they were told that the faces were in fact less or more attractive than they appeared on the photographs (updating manipulation). Specifically, half of participants learned that that the high (low) attractive face was photoshopped to appear more (less) attractive than it is in reality (opposite condition). The other half learned that the high (low) attractive face was photoshopped to appear less (more) attractive than it is in reality (consistent condition). Participants' attributions were then measured on a series of attributes referring to personality traits (e.g., "sociable") or to personality outcomes (e.g., "likely to achieve career success"; Dion et al., 1972) known to be more or less influenced by attractiveness (e.g., Eagly et al., 1991; Rougier et al., 2023). Finally, we measured perceived targets' attractiveness. This allowed us 1) to check that updating attractiveness indeed resulted in changes in perceived attractiveness (i.e., manipulation check) and 2) to test whether changes in perceived attractiveness in turn predicted changes in personality ratings (i.e., moderated mediation analyses).

In Study 1, we failed to produce the expected halo-update effect on personality traits, most likely because of the ambiguity of the updating manipulation. Study 2 addressed the limitations of the first study. Finally, Study 3 included measures before and after updating to assess more directly the change in perception caused by the updating manipulation. Studies 2 and 3 provided evidence for a halo-update effect with trait selectivity similar to past studies.

Transparency, Openness, and Analytical Strategy

We preregistered our studies on the Open Science Framework (OSF). Preregistrations include a priori theoretical reasoning, hypotheses, power estimations, procedures, and statistical analyses. Data were analyzed using RStudio, version 1.4.1106 (RStudio Team, 2021). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. The preregistration files, materials (including the reference

list of face stimuli and JsPsych code), data, and analytic (R) scripts for all experiments are made publicly available at https://osf.io/9wcmf/?view_only=fd1c6274fdf242059c34bd7ec6554220. Deviations from the preregistrations are reported as Supplementary materials (section "Deviations from the preregistrations"). Studies received approval (number 2021/39) from the ethical committee of the Faculty of Psychology and Educational Sciences at Ghent University.

Study 1

Study 1 involved a preliminary attempt to test the moderating role of updating on the attractiveness halo effect and its trait selectivity. We used a typical attractiveness halo procedure (see Dion et al., 1972) while additionally manipulating the updating of attractiveness between participants with instructions.

Method

Participants and Design

In a previous replication study of the attractiveness halo effect, Rougier et al. (2023; Study 1) produced an averaged sized halo effect of $dz = 0.42$. We opted for 360 participants, which provided us with a power of 0.99 to detect a halo effect of similar size (two-tailed *t*-test for two paired samples) and of 0.80 to detect a minimal halo-update effect of $d = 0.26$ (two-tailed *t*-test for two independent samples) with a 5% false-positive rate. Participants (M_{age} = 24.18, SD_{age} = 3.75, 188 women, 164 men, and 8 participants responding "other") took part in exchange for £1.50 and were recruited via Prolific Academic (www.prolific.co). Participants spoke English as their first language, did not take part in any other study of our lab, and had an approval rate of at least 95% (this last criterion leads to improved data quality; Peer et al., 2014). In line with studies on the attractiveness halo effect that mainly rely on college undergraduates (Bassili, 1981; Dermer & Thiel, 1975; Dion et al., 1972), we recruited participants aged between 18 and 30 years. A 3 (*Attractiveness*: low vs. medium vs. high) x

continuous (*Trait Relevance*: from -0.74 to +1.43) x 2 (*Updating*: opposite vs. consistent) mixed design was used with the last variable manipulated between participants.

Materials

Materials were the same as in Rougier et al. (2023; Study 1). Specifically, we used colored photographs of six male and six female faces from the 10k US Adult Faces Database (Bainbridge et al., 2013), that is, two faces per level of physical attractiveness and per gender. Relying on normative ratings provided by Bainbridge et al., faces were selected to vary on attractiveness but to not significantly differ on a series of other facial features (i.e., emotional intensity, image quality, memorability, and quantity of teeth visible; for more information, see Rougier et al., 2023). Faces were compared using two orthogonal contrast codes: a linear contrast C1 opposing low with high attractiveness conditions (low = -1/2, medium = 0, high = 1/2) and a quadratic C2 opposing low and high conditions taken together with the medium condition (low = -1/3, medium = 2/3, high = -1/3). Low and high attractive faces significantly differed on attractiveness, $t(9) = 3.04$, $p = .014$, $d = 1.01$, 95% CI [0.20; 1.79], whereas the medium faces did not significantly differ from low and high attractiveness faces taken together, $t(9) = 0.59$, $p = .57$, $d = 0.07$, 95% CI [-0.58; 0.73] ($M_{low} = 4.57$, $SD_{low} = 0.43$, $M_{med} = 5.35$, $SD_{med} = 1.64$, $M_{high} = 6.92$, $SD_{high} = 0.83$). Faces were all white, unknown (i.e., not celebrities), and most of them belonged to the 20-30 years old category.

The 42 personality traits and outcomes varied along seven dimensions: social competence (sociable, fun-loving, likable, popular), vanity/materialistic orientation (elitist, snobbish, shallow, humble, materialistic, pompous, prudish, boastful, vain), adjustment (normal, well-adjusted, satisfied, happy, confident, having a positive self-regard, mature, healthy), potency (strong, self-assertive, dominant, leader), intellectual competence (intelligent, skillful, rational, scientific, ambitious, hard-working, likely to receive good grades, likely to achieve career success), concerns for others (sensitive, empathic,

compassionate, generous, modest, egoistic), and integrity (trustworthy, honest, likely to be faithful to the spouse).

We relied on Rougier et al.'s continuous scores of trait halo relevance (i.e., the difference in ratings for high and low attractive people) as well as for the other trait controls (pilot study). Specifically, they collected ratings on trait valence (i.e., to what extent a trait is positive or negative; from -3 [*extremely negative*] to +3 [*extremely positive*]), self- vs. other-relevance (i.e., to what extent a trait is consequential for the trait holder vs. for the individuals living nearby the trait holder; score of difference with both scales from 1 [*low consequences*] to 7 [*high consequences*]), and face-readability (i.e., to what extent is it easy to infer a personality trait on the basis of someone's face; from 1 [*not easy at all*] to 7 [*extremely easy*]). We used these trait scores in order to exclude their potential effects in statistical analyses. Because analyses for trait halo relevance and trait valence were central, we present them in the manuscript whereas analyses for self/other relevance and face readability are presented as Supplementary Materials (Table S7).

Procedure

Halo Instructions. The experiment was programmed using jsPsych (de Leeuw, 2015). After giving their consent, participants were informed that our aim was to determine the extent to which impressions are generally accurate (e.g., Dion et al., 1972; Rougier et al., 2023). Participants were told that their accuracy in person perception would be compared with other groups who had been trained in various interpersonal perception techniques (students in clinical psychology and professional clinical psychologists). We told them that certain individuals without training might be as accurate as some professionals in their first impression judgements. Finally, participants were told that the photographs they would encounter were part of a group of college students currently enrolled in a longitudinal study of personality development and that it would therefore be possible to assess their judgement

accuracy by comparing judgements with people's real behavior. Then, participants were presented with three photographs (displayed side-by-side). They were asked to take a few moments to carefully look at the persons on the photographs and form an impression on them. Photographs were randomly selected from the pool of faces so that, for a given participant, all of them showed male or female faces, and one picture was included for each of the three levels of attractiveness (low vs. medium vs. high).

Updating Manipulation. Participants were then randomly assigned to the "opposite" or "consistent" updating condition. They were all informed that we sometimes used the Adobe Photoshop® software to modify the physical attractiveness of the person on the photograph. As a result, the photographs they previously saw may not have been representative of how these people look like in reality. We then provided additional information.

In the "opposite" condition, the updating information was in the opposite direction to the initial attractiveness values. Specifically, we indicated next to the low attractive face that we decreased the physical attractiveness of this person (implying that, in reality, this person is more attractive than on the picture) and next to the high attractive face that we increased the physical attractiveness of this person (implying that, in reality, s/he is less attractive). In the "consistent" condition, the updating information was in the same direction as the initial attractiveness values. Specifically, we indicated next to the low attractive face that we increased the physical attractiveness of this person (implying that s/he is even less unattractive) and next to the high attractive face that we decreased physical attractiveness of this person (implying that s/he is even more attractive). In both conditions, we indicated next to the medium attractive face that we did not modify the physical attractiveness of this person.

Personality Ratings. Then, participants evaluated each face on 42 different personality traits and personality outcomes (all scales ranged from 0 = *not at all* to 5 = *totally*). Each face was presented on the top of the screen with a reminder of the updating information (e.g., "Reminder: we increased the physical attractiveness of the person on this photograph"). Once participants completed the 42 ratings for one face, they moved on to the next face. The order in which the personality traits and outcomes were presented was randomized for each face and each participant separately. Participants were encouraged to answer as honestly and as spontaneously as possible.

Attractiveness Ratings. After the personality rating phase, participants were asked to rate the faces on attractiveness (from -3 = *extremely unattractive* to +3 = *extremely attractive*). Each face was shown on the top of the screen but there was no reminder of the updating information (random face order).

Memory and Believability of the Updating Information. Then, participants reported the updating information that was provided for each face (response options: "we increased the physical attractiveness", "we decreased the physical attractiveness", "we did not modify the physical attractiveness"). We also asked them to indicate whether they believed that faces' attractiveness was indeed modified using Photoshop (response options: "yes", "no", "I don't know"). We included this last measure as it seems that participants' beliefs about the truth of the updating information influences the expected results (so that a larger updating effect is observed for believers; Cone et al., 2017; Ferguson et al., 2019; Gregg et al., 2006; Shen & Ferguson, 2021). Because these two variables were declared as exploratory (cf. preregistration), results for these questions can be found in the Supplementary Materials (Tables S5 and S6).

Participants also answered a series of questions related to perceived demand awareness, awareness of the influence of attractiveness, and demand compliance (for more

information, see Supplementary Materials section “Exploratory questions used in Studies 1-3”). Finally, they provided demographic information (age, gender, English fluency) and could leave optional comments about the study before the debriefing.

Results

We excluded one participant having zero variance in their ratings for at least one target face. We reversed the ratings for negatively valenced traits, except for vanity-related traits, so that all ratings were in the expected direction of the effect (the higher the attractiveness, the higher should be the rating). The halo effect score corresponded to the difference in rating for high as compared to low attractive faces and it was computed per participant and per trait (i.e., as both were random factors in the mixed-model analyses; see Judd et al., 2017). Internal reliabilities within each of the seven trait dimension were relatively high (from α = .74, 95% CI [.70; .78], for vanity to α = .89, 95% CI [.87; .91], for intellectual competence). The variable of attractiveness was coded via two orthogonal contrast codes (linear contrast C1: low = -1/2, medium = 0, high = 1/2; quadratic contrast C2: low = -1/3, medium = 2/3, high = -1/3). Given that we focused on the linear contrast code C1, results for the quadratic contrasts are presented as Supplementary Materials (Tables S2 and S3). The updating condition ($N_{opposite}$ = 191, $N_{cconsistent}$ = 168), trait halo relevance, and valence variables were centered on zero. Mixed-model analyses were performed using the lmerTest package version 3.1-0 (Kuznetsova et al., 2017).

Based on previous work, we expected that the halo effect score (difference in rating between high and low attractive faces) would increase as a function of the trait relevance, that is, to be larger for highly (vs. low) relevant traits (Rougier et al., 2023). Crucially, we also examined whether the updating information moderated the halo effect and its trait selectivity. First, the halo effect should be larger in the “consistent” than in the “opposite” condition. Second, the hypothesized trait selectivity of the halo effect – the interaction between the

linear contrast C1 and the trait halo relevance – should be larger in the "consistent" than in the "opposite" condition. Additional analyses tested the alternative prediction that the halo effect and its interaction with the updating condition would be moderated by the trait valence score (i.e., interaction between attractiveness, updating condition, and trait valence) rather than by the trait halo score. In this case, the interaction between the halo effect and the trait valence would be larger in the "consistent" than in the "opposite" condition.

Attractiveness Ratings (OLS Regression)

In line with our a priori categorization (based on Bainbridge et al., 2013), participants judged high attractive faces as more attractive than low attractive faces, $t(357) = 23.28$, $p < .001$, $dz = 1.23$ [2], 95% CI [1.09; 1.37] ($M_{low} = -0.89$, $SD_{low} = 1.54$, $M_{med} = 0.17$, $SD_{med} = 1.50$, $M_{high} = 1.58$, $SD_{high} = 1.26$). This effect, however, was not moderated by the updating condition, $t(357) = 1.28$, $p = .20$, $d = 0.14$, 95% CI [-0.07; 0.34] (Consistent condition: $M_{low} = -0.89$, $SD_{low} = 1.60$, $M_{med} = 0.03$, $SD_{med} = 1.50$, $M_{high} = 1.73$, $SD_{high} = 1.15$; Opposite condition: $M_{low} = -0.88$, $SD_{low} = 1.49$, $M_{med} = 0.29$, $SD_{med} = 1.48$, $M_{high} = 1.45$, $SD_{high} = 1.34$).

Halo Effect and Its Moderation by the Updating Condition (Mixed-Model)

We ran a mixed-model with attractiveness (with C1 and C2), updating condition, and their interaction as fixed effects and personality ratings as the outcome measure. We estimated the random slope of attractiveness for both random factors of participants as well as traits and the slopes for updating condition and the interaction for the random factor of traits. We observed an attractiveness halo effect, that is, a higher average rating (with negative traits reversed, except vanity-related traits) for high attractive faces compared to low attractive faces, $t(49.73) = 3.50$, $p < .001$. [3] This halo effect, however, did not significantly differ between the opposite and the consistent conditions, $t(263.52) = 1.29$, $p = .20$, that is, we did not observe a halo-update effect (see Table 1).

[insert Table 1 here]

Trait Selectivity and Its Moderation by Updating Condition (Mixed-Models)

When adding the continuous score of traits relevance in the model, the interaction between the attractiveness linear contrast (low vs. high) and trait relevance was significant, $t(227.42) = 11.45$, $p < .001$. In line with our predictions, the halo effect increased as a function of the trait relevance, thus indicating a trait-selectivity in line with past literature (see Figure 1, left panel; Rougier et al., 2023). However, this interaction effect was not moderated by the updating condition, $t(340.40) = 1.20$, $p = .23$.

When adding the trait valence score in the model, the interaction between the attractiveness linear contrast and trait relevance remained significant, $t(181.11) = 10.90$, $p < .001$, and was not moderated by trait valence, $t(38.02) = 0.61$, $p = .55$. Importantly, however, we observed an interaction between the attractiveness linear contrast, trait valence, and the updating condition, $t(50.66) = 3.67$, $p < .001$ (see Figure 1, right panel). Specifically, in the opposite condition, the more negative the trait, the larger the difference between the high and the low attractive face (so that the high attractive face was rated more negatively than the low attractive face), $t(38.72) = 2.09$, $p = .04$. This effect, however, was not observed for the consistent condition, $t(38.57) = 1.18$, $p = .24$. No other effect came across as significant.

[insert Figure 1 here]

Discussion

Study 1 replicated the classic attractiveness halo effect and its typical trait selectivity. However, neither the halo or its trait selectivity were influenced by the updating condition. Rather, the halo-update effect on the how the target faces were perceived depended on the

trait valence: in the opposite condition, the difference in evaluation between high and low attractive faces increased as a function of traits' negativity, whereas this effect did not emerge in the consistent condition.

A possible explanation for a lack of halo-update effect could be that participants struggled to imagine how exactly faces were changed on attractiveness (i.e., which facial features were manipulated and to what extent). This could also explain why attractiveness ratings were not influenced by the updating manipulation. Regarding the unexpected interaction between the halo effect, the trait valence, and the updating condition, a possibility is that participants interpreted opposite updating from high to low attractiveness (i.e., that the high attractive face was in reality less attractive) as more negative than the updating from low to high attractiveness (i.e., that the low attractive face was in reality more attractive). This could be the case if participants believed that the person on the picture – rather than the researcher – was responsible for the photoshopped picture. In this case, people who want to appear more attractive than they are could be viewed in a negative light. Even though we indicated in the instructions that the manipulation was implemented by the researchers, this information may have been insufficiently salient. Study 2 addresses these two issues.

Study 2

Study 2 was similar to Study 1 except for two main changes. First, participants were provided examples of what changes (i.e., increase vs. decrease) in attractiveness would look like using example faces that were not used during the experimental manipulation. This was meant to help participants visualize how photoshopping could be used to change faces. We hence hoped that this example would increase the strength of our manipulation. Second, we made it even more clear and salient in the updating manipulation that *we* (i.e., the researchers) modified the faces' attractiveness. This excluded the alternative interpretation

that the photoshopping reflected the pictured persons' intention to appear more or less attractive.

Method

Participants and Design

Given that our design was identical to the one in Study 1, we relied on a similar sample size (N = 401, M_{age} = 23.53, SD_{age} = 3.60, 240 women, 156 men, and 5 participants responding "other"). Retribution and selection criteria were also identical.

Materials and Procedure

Materials and procedure were similar to those in Study 1 except for the following changes. First, we only used female faces given technical constrains in creating examples for attractiveness changes. [4] Second, after the initial halo instructions, participants were informed that "sometimes *we intentionally* photoshopped (with the Adobe Photoshop software) the persons on the pictures so that they appear more or less attractive than they are in reality". Third, participants were presented with two pairs of face pictures that served as examples to illustrate the attractiveness change. The pairs consisted in an authentic and a photoshopped version of the same face picture. One face picture was used to represent changes in the case of an increase in attractiveness ("to make a person more attractive"), the other to represent changes in the case of a decrease in attractiveness ("to make a person less attractive"). Faces were modified using FaceApp ("Hollywood 2" filter) which is intended to increase the attractiveness of a face picture. Accordingly, we used the modified version as the photoshopped picture for one face (i.e., to represent increase in attractiveness) and as the authentic picture for the other face (i.e., to represent decrease in attractiveness). The example faces were not used as face stimuli in the experimental manipulation.

After providing examples of attractiveness changes, participants were again presented with the three faces, along with instructions explaining how the attractiveness of faces was

modified. Participants were asked to "try to picture how these persons look like in real life". As in Study 1, participants were randomly assigned to the opposite or the consistent condition, after which they went through the personality traits/outcomes ratings and attractiveness ratings. Participants were also instructed to consider the updating information for the attractiveness ratings (i.e., "keep in mind the information we gave you before regarding how we photoshopped the photographs") contrary to Study 1 in which this information was only provided before the personality ratings. Then, participants reported their memory and believability of the updating information, and answered the same exploratory questions and demographics as before.

Results

We excluded four participants having zero variance in their ratings. We used the same contrast codes and models as in Study 1. Reliabilities within each trait dimension were relatively high (from α = .71, 95% CI [.67; .76], for concerns for others to α = .87, 95% CI [.85; .89], for intellectual competence).

Attractiveness Ratings (OLS Regression)

Participants judged high attractive faces as more attractive than low attractive faces, $t(395) = 22.33$, $p < .001$, $dz = 1.12$, 95% CI [1.00; 1.25] (M_{low} = -0.58, SD_{low} = 1.61, M_{med} = 0.79, SD_{med} = 1.32, M_{high} = 1.75, SD_{high} = 1.20). This effect was moderated by the updating condition, $t(395) = 9.75$, $p < .001$, $d = 0.98$, 95% CI [0.77; 1.19], so that the difference between high and low attractive faces was larger in the consistent (M_{low} = -1.17, SD_{low} = 1.38, M_{med} = 0.67, SD_{med} = 1.29, M_{high} = 2.11, SD_{high} = 1.12) than in the opposite condition (M_{low} = 0.06, SD_{low} = 1.59, M_{med} = 0.90, SD_{med} = 1.33, M_{high} = 1.35, SD_{high} = 1.17).

Halo Effect and Its Moderation by the Updating Condition (Mixed-Model)

The average rating was higher for high attractive faces compared to low attractive faces, $t(48.81) = 5.01$, $p < .001$, indicating an attractiveness halo effect. Moreover, this halo

effect was moderated by the updating condition, $t(106.50) = 2.61$, $p = .01$, indicating a halo-update effect (i.e., larger halo effect in the consistent than in the opposite condition; see Table 1). Simple effect analyses revealed that the halo effect was significant in both the consistent, $t(49.57) = 4.63$, $p < .001$, and the opposite condition, $t(18.22) = 4.84$, $p < .001$. Additional analyses (OLS regression) showed that the halo-update effect also emerged when only considering ratings of vanity-related traits, that is, negatively valenced traits (for the halo-update effect per dimension see Supplementary Materials Table S4).

Trait Selectivity and Its Moderation by the Updating Condition (Mixed-Models)

The interaction between the attractiveness linear contrast (low vs. high) and trait relevance was significant, $t(118.40) = 10.21$, $p < .001$, so that the halo effect increased as a function of the trait relevance (see Figure 2, left panel). Moreover, this interaction effect was moderated by the updating condition, $t(333.50) = 4.84$, $p < .001$, so that the interaction between the halo effect and trait relevance was larger in the consistent than in the opposite condition.

Contrary to Study 1, when adding the trait valence in the model, there was no significant interaction between the attractiveness linear contrast, trait valence, and the updating condition, $t(38.11) = 0.10$, $p < .92$ (see Figure 2, right panel). No other effect came across as significant.

[insert Figure 2 here]

Moderated Mediation (Mixed-Models)

We tested a moderated mediation (non-preregistered) model where the manipulated attractiveness of the faces would lead to changes in perceived attractiveness as a function of the updating condition, in turn influencing ratings on the personality traits. We relied on the

joint-significant testing method (see Muller et al., 2005; Yzerbyt et al., 2018) and mixed-model analyses with only participants as a random factor. [5] A first model tested the updating moderation on the effect of face's attractiveness on personality ratings. The total effect of face's attractiveness on personality ratings was significant, $B = 0.51$, $B\ SE = 0.03$, $t(395.00) = 15.85$, $p < .001$, and this effect was moderated by the updating condition, $B = 0.24$, $B\ SE = 0.06$, $t(395.00) = 3.79$, $p < .001$. A second model tested the updating moderation on the effect of face's attractiveness on attractiveness ratings. The effect of face's attractiveness on attractiveness ratings emerged, $B = 2.28$, $B\ SE = 0.10$, $t(395.00) = 22.33$, $p < .001$, and it was moderated by the updating condition, $B = 1.99$, $B\ SE = 0.20$, $t(395.00) = 9.75$, $p < .001$. A final model tested the effect of face's attractiveness, updating condition, attractiveness ratings, and all their interaction effects on personality ratings. The effect of attractiveness ratings on personality ratings was also significant, $B = 0.08$, $B\ SE = 0.01$, $t(6297.00) = 11.90$, $p < .001$, and not significantly moderated by the updating condition, $B = -0.02$, $B\ SE = 0.01$, $t(6297.00) = 1.79$, $p = .07$. Face's attractiveness still had a significant effect on personality ratings, $B = 0.33$, $B\ SE = 0.03$, $t(601.50) = 9.62$, $p < .001$, and this effect was not significantly moderated by the updating condition, $B = 0.13$, $B\ SE = 0.07$, $t(601.60) = 1.86$, $p = .06$. This indicates that the updating condition moderated the mediation on the link between faces' attractiveness and attractiveness ratings, in turn influencing personality ratings (see Figure 3).

[insert Figure 3 here]

Discussion

Study 2 demonstrated the halo-update effect. Providing additional opposite (vs. consistent) information on the attractiveness of the faces decreased perceived differences between faces on attractiveness and personality traits and outcomes. This effect emerged on

both positive (e.g., sociable) and negative (e.g., vain) traits. Moreover, it decreased trait selectivity so that the typical pattern of larger halo effect for some personality traits than others was less extreme. Said otherwise, the halo-update effect was larger for personality traits that are stereotypically related to attractiveness.

Moreover, results of the (non-preregistered) moderated mediation analysis were in line with the idea that the observed halo-update effect (i.e., interaction between attractiveness and updating on the personality ratings) was mediated by changes in perceived attractiveness. In other words, the manipulated attractiveness influenced the perception of attractiveness as a function of the updating condition, in turn influencing personality ratings. Importantly, whereas the first step illustrates that updating was successful (i.e., updating of attractiveness indeed occurred), the second step shows that the extent to which the updating occurred (i.e., the extent to which attractiveness perception was changed) predicted the extent to which personality ratings changed.

Study 2, however, is limited in two important ways. First, participants were explicitly asked to take into account the Photoshop manipulation when evaluating targets on attractiveness and personality traits. As a result, participants may have feel constrained to use this information while under more natural conditions they would not have considered it. Second, we did not measure participants' stereotypical attributions before the updating manipulation but only after. Hence, we cannot conclude that participants' stereotypical attributions can be *changed*: it could be that once participants expressed a judgment, this judgment can no longer be changed by an updating information – or, alternatively, that such changes only materialize on attractiveness but not on personality ratings. We address these two limitations in Study 3.

Study 3

Study 3 implemented two main changes. First, we did not ask participants to consider the updating manipulation in any of their ratings. Second, we relied on two measurement times. After presenting the target faces, participants evaluated them on the personality traits/outcomes and attractiveness (Time 1). We expected to observe stereotypical attributions typical of the halo effect. Then, participants were presented with the (consistent vs. opposite) updating information as in Study 2 and had to evaluate again the faces on personality and attractiveness (Time 2). We expected a *change* on stereotypical attributions between Time 1 and Time 2 as a function of the updating condition. Again, we tested whether this change was explained by changes in perceived attractiveness.

Method

Participants and Design

Our design was similar to the one in Study 2 except for the within-participant variable of the measurement time (Time 1 vs. Time 2) and the attractiveness variable that only involved two conditions (low vs. high). Hence, we relied on a similar sample size (N = 400, M_{age} = 26.01, SD_{age} = 4.38, 152 women, 235 men, and 13 participants responding "other"). Retribution and selection criteria were identical.

Materials and Procedure

Materials and procedure were similar to those of Study 2 with the differences mentioned hereafter. First, we only used low and high attractive faces, that is, the most informative levels of attractiveness (i.e., we discarded the medium level). Second, we specified in the instructions that the individuals to which the faces belong to were named "Ann" or "Judy" (random assignment). Faces were always labelled with their first name in the instructions and participants were asked to memorize which face belonged to Ann and Judy. Hence, we did no longer present the faces during the evaluation phases (personality and attractiveness ratings) and referred only to the individual being evaluated (e.g., "Evaluate

Judy on the following personality traits and outcomes"). This allowed to create a more ambiguous situation and avoid that participants relied solely on the visible features of the faces (e.g., attractiveness). Third, we used two measurement times. After receiving the initial instructions and being presented with the target faces, participants were asked to rate the faces on personality traits/outcomes and attractiveness (Time 1). No mention was made of faces' attractiveness. Then, participants received the exact same updating information as in Study 2 and were asked again to evaluate the faces on personality traits/outcomes and attractiveness (Time 2). We did not ask participants to consider the updating for their judgment (i.e., instructions were identical between Time 1 and Time 2).

Finally, in addition to the exploratory and demographic questions of Study 2, we also asked participants to report their memory of the first name for each face ("What is the first name of this person?"; response options: "Judy", "Ann"), and to indicate whether their first name was Ann or Judy ("Is you name Ann or Judy?"; response options: "Yes (my name is Ann or Judy)", "No").

Results

We excluded eight participants having zero variance in their ratings, one participant who reported being named Ann or Judy, and sixteen participants who failed to remember correctly the first name of each face. This left us with a sample of 375 participants (M_{age} = 26.04, SD_{age} = 4.42, 146 women, 217 men, and 12 participants responding "other"). We used the same contrast codes except for attractiveness that was contrast coded (low: -0.5, high: +0.5). Reliabilities within each trait dimension at measurement Times 1 and 2 were relatively high (from α = .75, 95% CI [.70; .78], for concerns for others at Time 1 to α = .88, 95% CI [.87; .90], for intellectual competence at Time 2). The results section follows the same structure as before except that it focuses on Time 2 measurement for both attractiveness and personality ratings (results for Time 1 measurement are presented as Supplementary

Materials Table S8). We used the same models as before but additionally controlled for personality ratings of Time 1, that is, we used this variable as a covariate in mixed-models.

Attractiveness Ratings (OLS Regression)

Participants judged high attractive faces as more attractive than low attractive faces, $t(373) = 15.93$, $p < .001$, $dz = 0.82$, 95% CI [0.71; 0.94] ($M_{low} = -0.07$, $SD_{low} = 1.64$, $M_{high} = 1.52$, $SD_{high} = 1.34$). This effect was significantly moderated by the updating condition, $t(373) = 13.85$, $p < .001$, $d = 1.43$, 95% CI [1.21; 1.66], so that the difference between high and low attractive faces was larger in the consistent ($M_{low} = -0.97$, $SD_{low} = 1.54$, $M_{high} = 2.07$, $SD_{high} = 1.24$) than in the opposite condition ($M_{low} = 0.79$, $SD_{low} = 1.22$, $M_{high} = 1.00$, $SD_{high} = 1.21$). [6]

Halo Effect and Its Moderation by the Updating Condition (Mixed-Model)

The average rating was only descriptively higher for high attractive faces compared to low attractive faces, $t(76.40) = 1.88$, $p = .06$. The halo effect, however, was moderated by the updating condition, $t(62.88) = 5.24$, $p < .001$, indicating a halo-update effect (see Table 1). Simple effect analyses revealed that the halo effect emerged in both the consistent, $t(55.38) = 4.05$, $p < .001$, and the opposite conditions, $t(240.20) = 4.83$, $p < .001$, signaling a significant reversal of the halo effect (i.e., lower ratings for high [vs. low] attractive faces) in the latter. Again, the halo-update effect also emerged in OLS regression when only considering ratings of vanity-related traits (see Supplementary Materials, Table S4).

Trait Selectivity and Its Moderation by the Updating Condition (Mixed-Models)

The interaction between faces' attractiveness (low vs. high) and trait relevance was significant, $t(201.70) = 6.62$, $p < .001$, so that the attractiveness effect increased as a function of the trait relevance (see Figure 4, left panel). Moreover, this interaction effect was moderated by the updating condition, $t(156.70) = 7.27$, $p < .001$, so that the interaction between attractiveness and trait relevance was larger in the consistent than in the opposite condition.

When adding the trait valence in the model, there was no significant interaction between attractiveness, trait valence, and the updating condition, $t(38.01) = 1.53$, $p = .13$ (see Figure 4, right panel). No other effect came across as significant.

[insert Figure 4 here]

Moderated Mediation (Mixed-Models)

We relied on the same three mixed-models as in Study 2, except that we additionally controlled for the personality ratings at Time 1 (for model 1 and 3) and for the attractiveness ratings at Time 1 (for model 2). First, the total effect of face's attractiveness on personality ratings at Time 2 was significant, $B = 0.06$, $B\ SE = 0.03$, $t(376.00) = 2.02$, $p = .04$, and it was moderated by the updating condition, $B = 0.72$, $B\ SE = 0.06$, $t(373.00) = 3.79$, $p < .001$. Second, the effect of face's attractiveness on attractiveness ratings at Time 2 emerged, $B = -0.33$, $B\ SE = 0.02$, $t(112.60) = 18.88$, $p < .001$, and it was moderated by the updating condition, $B = -3.12$, $B\ SE = 0.03$, $t(90.10) = 100.38$, $p < .001$. Finally, the effect of attractiveness ratings at Time 2 on personality ratings at Time 2 was also significant, $B = 0.10$, $B\ SE = 0.10$, $t(739.00) = 9.74$, $p < .001$, and not significantly moderated by the updating condition, $B = 0.01$, $B\ SE = 0.02$, $t(735.00) = 0.39$, $p = .70$. Face's attractiveness still had an effect on personality ratings at Time 2, $B = -0.11$, $B\ SE = 0.04$, $t(504.40) = 2.92$, $p = .004$, and this effect was still moderated by the updating condition, $B = 0.45$, $B\ SE = 0.07$, $t(503.90) = 6.07$, $p < .001$. Overall, the updating condition moderated the mediation on the link between faces' attractiveness and attractiveness ratings, in turn influencing personality ratings (see Figure 5).

[insert Figure 5 here]

Discussion

We observed a halo-update effect that was also trait-selective even though 1) participants were not explicitly asked to consider the updating information for the ratings and 2) personality and attractiveness ratings were measured both before and after the updating. Findings were consistent with the idea that stereotypical attributions can be changed even after having been expressed. Moreover, the halo effect in the opposite condition was not only reduced but significantly reversed, suggesting a strong effect of the updating manipulation. As in Study 2, changes in perceived attractiveness emerged as a mediator explaining (at least partly) the halo-update effect.

General Discussion

In line with past EC research on the US-revaluation and past impression formation research on updating, we investigated the updating effect in the context of the attractiveness halo effect – a phenomenon where attractiveness information influences assumptions on multiple personality traits. This allowed us to test whether a full, non-attitudinal pattern of stereotypical assumptions that is based on a single attribute (here, attractiveness) can be changed by updating the information about that one attribute.

Study 1 failed to produce the expected halo-update effect. The updating information did influence how positively/negatively the faces were perceived. More specifically, in the opposite – but not the consistent – condition, faces that were initially highly attractive but were updated to lower attractiveness were perceived as more negative than faces that were initially low and were updated to higher attractiveness. We reasoned that participants may have attributed the attractiveness change to the intention of the target individuals (e.g., s/he wants to appear more attractive than s/he is).

Study 2 made explicit that the change in attractiveness was the result of interventions by the experimenters. We observed both the expected halo-update effect and its trait

selectivity (i.e., smaller halo and trait selectivity effects in the opposite condition). Additional analyses showed that the observed halo-update effect was mediated by changes in perceived attractiveness. Study 3 addressed two limitations of previous studies. First, participants were not explicitly asked to consider the updating information when rating the faces. Second, we assessed the ratings both before and after receiving the updating information. We replicated results of Study 2.

Taken together, these studies show for the first time that updating can also apply in the case of the halo effect: changing the attractiveness of faces modified the pattern of stereotypical attributions. In the following, we explain how the present work empirically contributes to literatures on updating and stereotyping.

Empirical Contribution to Updating and Stereotyping Literatures

As past research on updating effects focused on phenomena involving mere changes in evaluation (e.g., EC effect; Walther et al., 2009), it remained unknown whether updating could apply on non-attitudinal effects (i.e., not relying on mere change in valence). Crucially, even if sometimes researchers used an outcome measure that differed from the initial information, a mere valence effect could not be excluded (e.g., Förderer & Unkelbach, 2016; Hogarth & Einhorn, 1992; Shen et al., 2020). Hence, a first contribution of this work is to show for the first time that the updating effect can apply on a non-attitudinal phenomenon that involves a pattern of attributions (i.e., selective assumptions on a series of traits). Our research thus opens the possibility that the updating effect could be observed for other impression formation effects that typically go beyond mere changes in evaluation (e.g., Spontaneous Trait Transference effects; Skowronski et al., 1998).

Second, our work hints an important boundary condition of the halo-update effect. As suggested by a comparison of the results observed in Studies 1 and 2, the halo-update effect seems to depend on whether the attractiveness update can be attributed to the intention of the

target. When the target has the intention to manage her/his impression, additional inferences can be made (e.g., the person is insecure) that could prevent or counteract the effect of the attractiveness stereotype. Importantly, this idea is preliminary at best because we did not directly compare different updating manipulation in the same study.

Third, our work is consistent with previous literature on order effects (e.g., Hogarth & Einhorn, 1992). When providing different pieces of information, sometimes the first information has a greater impact on the judgment of a target (primacy effect; Anderson, 1965), whereas sometimes the last information has a greater impact (recency effect; Hogarth & Einhorn, 1992). Importantly, asking participants to emit a judgment only after integrating the different information together typically leads to a primacy effect, whereas asking participants to emit a judgment after each information leads to a recency effect (Hogarth & Einhorn, 1992). Consistently, it was only when participants judged the target two times (Study 3) that we observed a reversed halo effect in the opposite updating condition. Again, this conclusion is preliminary given that it is based on a between-study comparison.

As a final contribution, our results demonstrate that changing the value of the source of the stereotype constitutes one other possible way of changing stereotypical attributions. Interestingly, training procedures have been created to modify stereotypical attributions emanating from a category (e.g., Burns et al., 2017; Kawakami et al., 2000, 2007). In a counterstereotypic training, participants are asked to repeatedly select traits that are opposite to those culturally associated with the category (e.g., to select “weak” instead of “sloppy” for men), ultimately leading to changes in stereotypical attributions (as compared to the “no training” condition). One obvious advantage of the updating procedure, however, is that it only requires a short verbal instruction (here, on attractiveness) to observe a change in stereotyping (see also Mann et al., 2020; Shen & Ferguson, 2021; Shen et al., 2020).

Theoretical Contribution to Impression Formation and Conditioning

As outlined in the introduction, stereotype models perfectly account for the halo effect. These could also accommodate the influence of updating: when there are changes in the value of the category on which individuals rely for their judgment (e.g., attractiveness changes from low to high), the application of the stereotype changes accordingly (i.e., diminishes or reverses; Fiske & Neuberg, 1990; Kunda & Spencer, 2003).

Interestingly, the stereotyping perspective can also be linked to theories of EC. The updating effect has been investigated intensively in EC, notably because different associative accounts of EC lead to different predictions (for an overview see De Houwer & Hughes, 2020). Specifically, S-R (Stimulus-Response) models assume that the evaluative response (R) of the US becomes directly linked to the CS (S) after the conditioning procedure, whereas S-S (Stimulus-Stimulus) models assume that the US (S) and the CS (S) become associated in memory. Because the S-S account suggests that the conditioning effect depends on the knowledge about the US (i.e., the CS becomes bad only because it is related to a US that is bad), only this type of model is in line with the US revaluation effect (e.g., Gast & Rothermund, 2011; Walther et al., 2009). Although the available evidence on US revaluation supports S-S models, it has been argued that under some circumstances (i.e., after many pairings), EC and other conditioning effects could become habitual, that is, based on S-R associations (e.g., Dickinson, 2012; Gast & Rothermund, 2011). From this perspective, future studies could test whether stereotyping can become habitual (i.e., based on S-R associations).

Contrary to associative accounts of EC that postulate the formation of associations in memory, propositional accounts argue that EC results from the formation and activation of propositions about relations in the environment (e.g., the proposition “the CS and US go together”; e.g., De Houwer, 2009, 2018; Mitchell et al., 2009). Propositional models are in line with updating effects as they allow for inferential processes to revise previous propositions that were formed on the basis of the initial information. This is also in line with

recent research suggesting that propositional processes are responsible also for the updating of impressions. Indeed, they could easily explain why mere verbal statements can override previous repeated pairing information (Mann et al., 2020), and that highly diagnostic (Cone & Ferguson, 2015; Cone et al., 2017; Shen & Ferguson, 2021; Shen et al., 2020), and believable information (Cone et al., 2021; Shen & Ferguson, 2021) leads to larger updating effects. Propositional approaches are also in line with the present halo-update effect demonstrating that a verbal instruction can override contradictory visual-based impression (see also Shen & Ferguson, 2021; Shen et al., 2020). In line with previous work, we also found in Study 3 that the halo-update effect increased when belief in the updating information increased (see Supplementary Materials Table S6; but see Study 2).

However, current accounts of EC do not provide a straightforward explanation for the fact that the updating information selectively influenced assumptions on the personality traits. Still, these accounts could accommodate this finding based on the auxiliary assumption that the observed effects result from both the updating of attractiveness and pre-existing knowledge participants have about attractiveness (see Rougier et al., 2023). Overall, this interpretation underlines that theoretical explanations of impression formation and learning both have merits in explaining updating effects and suggests the possibility of models integrating elements of both impression formation and learning (see De Houwer et al., 2019).

Limitations and Future Work

Several limitations can be noted regarding our series of studies. First, it remains unknown whether the halo-update effect from both consistent and opposite conditions would differ from a control condition. It could be that only the opposite updating leads to a decrease of the halo effect, that only the consistent updating leads to an increase, or, alternatively, that the two conditions differ from the control group. For a more definitive answer, a condition

not including any attractiveness change (hence controlling for the mere passage of time between the two measurements) should be added in the design (e.g., Mann et al., 2020).

Second, we only tested the halo-update effect using a direct measure of self-reports. Yet, literature on updating suggests that this effect is more difficult to observe in indirect measures (e.g., Implicit Association Test; Gregg et al., 2006; McConnell et al., 2008; Rydell & McConnell, 2006) but that it can emerge when the additional information is extreme and diagnostic (Cone & Ferguson, 2015; Cone et al., 2021; Shen & Ferguson, 2021; Shen et al., 2020). One could argue that the updating information we used was diagnostic because it was specifically about the initial information of attractiveness (see also Mann et al., 2019; Mann & Ferguson, 2015). Future studies should test whether the current procedure also leads to a halo-update effect in indirect measures.

Third, when and how initial assumptions partly "stick" to the targets after updating attractiveness remains unknown. Moderated mediation analyses in Studies 2 and 3 suggested that some part of the initial stereotypical assumptions were maintained beyond the halo-update effect (i.e., there was still a direct effect of attractiveness on personality ratings after controlling for the indirect effect via perceived attractiveness). [7] Because of the obvious practical importance of lingering assumptions (e.g., fake news), future studies should investigate in which contexts initial assumptions are most likely to linger.

Fourth, it is well known that the attractiveness halo effect leads to problematic real-life consequences with, in general, a more positive treatment of attractive (vs. unattractive) individuals (for a meta-analysis, see Langlois et al., 2000). Hence, future research could focus on whether the observed halo-update effect can translate into meaningful real life behaviors or behavioral intentions (e.g., Cone & Ferguson, 2015; Exp. 5). As an intermediary step, one would need to make sure that the updating verbal instruction can compete with the visual information (facial attractiveness) because the latter often remains available in real-life

contexts (i.e., when individuals continue to be exposed to the face after receiving the updating information). A possible way around this issue might be to provide information about the person's behavior related to his/her attractiveness instead of information about the facial information *per se* (e.g., s/he won a beauty contest; see also Shen et al., 2020). [8]

Finally, it remains unknown whether the observed halo-update effect can endure over time. Indeed, after some time the consideration of the additional information provided on the target's attractiveness could vanish (e.g., Kurdi & Banaji, 2019). In the updating literature, the updating effect can maintain for a few days, or even up to a few weeks or months (e.g., Cone et al., 2021; Mann & Ferguson, 2015, 2017; Mann et al., 2020; Shen & Ferguson, 2021). Future studies should investigate whether similar results can be observed in the context of the halo effect. We predict that the halo-update effect is likely to maintain, as long as the updating information is highly diagnostic (Mann et al., 2020).

Conclusion

Relying on conditioning and impression formation literatures, we examined whether the typical updating effect observed in both domains also applies in the context of the attractiveness halo effect. The halo-update effect extends previous research in demonstrating that the initial pattern of stereotypical assumptions one typically derives from attractiveness can be changed when attractiveness is changed. By underlying conjoint moderators of updating, this work emphasizes empirical and theoretical commonalities between domains of research that are usually considered as distinct.

References

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology, 2*(1), 1–9. https://doi.org/10.1037/h0021966

Ashmore, R. D. (1981). Sexstereotypes and implicit personality theory. In D.L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 37-81). Hillsdale, NJ: Erlbaum.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US-revaluation. *Learning and Motivation*, *23*(2), 200–224. https://doi.org/10.1016/0023-9690(92)90018-H

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323c1334. https://doi.org/10.1037/a0033872

Bassili, J. N. (1981). The attractiveness stereotype: Goodness or glamour?. *Basic and Applied Social Psychology*, *2*(4), 235–252. https://doi.org/10.1207/s15324834basp0204_1

Burns, M. D., Monteith, M. J., & Parker, L. R. (2017). Training away bias: The differential effects of counterstereotype training and self-regulation on stereotype activation and application. *Journal of Experimental Social Psychology, 73*, 97–110. https://doi.org/10.1016/j.jesp.2017.06.003

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of personality and social psychology*, *69*(3), 420–436. https://doi.org/10.1037/0022-3514.69.3.429

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57. https://doi.org/10.1037/pspa0000014

Cone, J., Flaharty, K., & Ferguson, M. J. (2021). The long-term effects of new evidence on implicit impressions of other people. Psychological science, 32(2), 173-188. https://doi.org/10.1177/095679762096355

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology*, 131–199. https://doi.org/10.1016/bs.aesp.2017.03.001

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish journal of psychology, 10*(2), 230–241. https://doi.org/10.1017/s1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*(1), 1□–20. https://doi.org/10.3758/lb.37.1.1

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin, 13*(3), 1–21. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., & Hughes, S. (2020). *The psychology of learning: An introduction from a functional-cognitive perspective*. MIT Press.

De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2019). On the Assumptions That We Make About the World Around Us: A Conceptual Framework for Feature Transformation Effects. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.229

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dermer, M., & Thiel, D. L. (1975). When beauty may fail. *Journal of Personality and Social Psychology, 31*(6), 1168–1176. https://doi.org/10.1037/h0077085

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1603), 2733–2742. https://doi.org/10.1098/rstb.2012.0220

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285–290. https://doi.org/10.1037/h0033731

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin, 110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Efran, M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality, 8*(1), 45–54. https://doi.org/10.1016/0092-6566(74)90044-0

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit first impressions can be updated. *Current Directions in Psychological Science, 28*(4), 331-336. https://doi.org/10.1177/0963721419835206

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology* (Vol. 23, pp. 1-74). Academic Press. https://doi.org/10.1016/s0065-2601(08)60317-2

Förderer, S., & Unkelbach, C. (2011). Beyond evaluative conditioning! Evidence for transfer of non-evaluative attributes. *Social Psychological and Personality Science, 2*(5), 479c486. https://doi.org/10.1177/1948550611398413

Förderer, S., & Unkelbach, C. (2015). Attribute conditioning: Changing attribute-assessments through mere pairings. *Quarterly Journal of Experimental Psychology*, *68*(1), 144–164. https://doi.org/10.1080/17470218.2014.939667

Förderer, S., & Unkelbach, C. (2016). Changing US attributes after CS-US pairings changes CS-attribute-assessments: Evidence for CS-US associations in attribute conditioning. *Personality and Social Psychology Bulletin*, *42*(3), 350–365. https://doi.org/10.1177/0146167215626705

Forgas, J. P., & Laham, S. M. (2016). Halo effects. In *Cognitive Illusions* (pp. 286–300). Psychology Press. https://doi.org/10.4324/9781315696935-22

Gast, A., & Rothermund, K. (2011). I like it because I said that I like it: Evaluative conditioning effects can be based on stimulus-response learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(4), 466–476. https://doi.org/10.1037/a0023077

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1–20. https://doi.org/10.1037/0022-3514.90.1.1

Han, D. E., & Laurent, S. M. (2023). Beautiful seems good, but perhaps not in every way: Linking attractiveness to moral evaluation through perceived vanity. *Journal of Personality and Social Psychology*, *124*(2), 264–286. https://doi.org/10.1037/pspa0000317

Heilman, M. E., & Stopeck, M. H. (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. *Organizational Behavior and Human Decision Processes*, *35*(2), 202–215. https://doi.org/10.1016/0749-5978(85)90035-4

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*(1), 1–55. https://doi.org/10.1016/0010-0285(92)90002-j

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology, 68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*(5), 871–888. https://doi.org/10.1037/0022-3514.78.5.871

Kawakami, K., Dovidio, J. F., & Van Kamp, S. (2007). The impact of counterstereotypic training and related correction processes on the application of stereotypes. *Group Processes & Intergroup Relations, 10*(2), 139-156. https://doi.org/10.1177/1368430207074725

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*(4), 522–544. https://doi.org/10.1037/0033-2909.129.4.522

Kurdi, B., & Banaji, M. R. (2019). Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *Journal of Personality and Social Psychology, 116*(5), 681–703. https://doi.org/10.1037/pspa0000151

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390–423. https://doi.org/10.1037/0033-2909.126.3.390

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*(3), 397–407. https://doi.org/10.1037/0022-3514.69.3.397

Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology, 116*(3), 349–374. https://doi.org/10.1037/pspa0000146

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108*(6), 823–849. https://doi.org/10.1037/pspa0000021

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. Journal of Experimental Social Psychology, 68, 122-127. https://doi.org/10.1016/j.jesp.2016.06.004

Mann, T. C., Kurdi, B., & Banaji, M. R. (2020). How effectively can implicit evaluations be updated? Using evaluative statements after aversive repeated evaluative pairings. *Journal of Experimental Psychology: General, 149*(6), 1169–1192. https://doi.org/10.1037/xge0000701

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology, 69*(5), 812–824. https://doi.org/10.1037/0022-3514.69.5.812

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Forming implicit and explicit attitudes toward individuals: Social group association cues. *Journal of Personality and Social Psychology, 94*(5), 792–807. https://doi.org/10.1037/0022-3514.94.5.792

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183-198. https://doi.org/10.1017/S0140525X09000855

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*(6), 852–863. https://doi.org/10.1037/0022-3514.89.6.852

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

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*(4), 1023–1031. https://doi.org/10.3758/s13428 -013-0434-y

Richards, Z., & Hewstone, M. (2001). Subtyping and Subgrouping: Processes for the Prevention and Promotion of Stereotype Change. *Personality and Social Psychology Review*, *5*(1), 52–73. https://doi.org/10.1207/s15327957pspr0501_4

Rougier, M., De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2023). From halo to conditioning and back again: Exploring the links between impression formation and learning. *Collabra: Psychology*, *9*(1), https://doi.org/10.1525/collabra.84560

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

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

Shen, X., & Ferguson, M. J. (2021). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, *97*, 104219. https://doi.org/10.1016/j.jesp.2021.104219

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face?: Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, *86*, 103888. https://doi.org/10.1016/j.jesp.2019.103888

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of personality and social psychology, 74*(4), 837–848. https://doi.org/10.1037/0022-3514.74.4.837

Sweldens, S., Van Osselaer, S. M., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of consumer Research*, *37*(3), 473–489. https://doi.org/10.1086/653656

van Knippenberg, A., van Twuyver, M., & Pepels, J. (1994). Factors affecting social categorization processes in memory. *British Journal of Social Psychology*, *33*(4), 419-431. https://doi.org/10.1111/j.2044-8309.1994.tb01038.x

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition and Emotion*, *23*(5), 889–917. https://doi.org/10.1080/02699930802212423

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, *115*(6), 929–943. https://doi.org/10.1037/pspa0000132

Footnotes

[1] We wish to be transparent about the fact that the idea for this series of studies emerged from our analysis of EC and impression formation effects in terms of "feature transformation effects" (De Houwer et al., 2019). The conceptual framework of Feature Transformation effects aims at creating bridges between impression formation and learning research by describing effects in both domains using the same generic, theory-free concepts (i.e., source/target features, source/target objects). The re-description of EC and impression formation effects helped us to identify similarities and differences between EC and impression formation effects, such as the way both are sensitive to updating and how updating can be described (i.e., source feature updating). We then applied the same reasoning to the attractiveness halo effect and tested whether a similar effect could also emerge in this context.

[2] We computed the effect sizes (Cohen's *d*) for the classic OLS regression analyses ('by-participants' analyses), however, we did not compute the effect sizes for the mixed-model analyses given that there is no clear consensus on this matter (Judd et al., 2017).

[3] The attractiveness linear contrast was moderated by the target gender, $t(351.13) = 8.27$, $p < .001$, so that the attractiveness effect was larger for female than for male targets. Importantly, the attractiveness linear contrast remained significant on average when controlling for the target gender, $t(127.94) = 8.56$, $p = .02$.

[4] We did not manage to create convincing authentic and photoshopped examples for male faces using FaceApp. Therefore, we chose to rely only on female faces for examples and stimuli.

[5] The models could not be run with OLS regression because of the within-participants IV (attractiveness). Moreover, we only used participants as a random factor because having traits as a random factor was impossible for one of the regression model.

[6] To test more directly whether attractiveness judgments *changed* as a function of the updating condition (i.e., whether attractiveness judgment at Time 1 predicted attractiveness judgment at Time 2 differently as a function of the updating condition) we conducted an additional mixed-model analysis. Specifically, we tested a model having the attractiveness ratings at Time 2 as the outcome measure and the attractiveness ratings at Time 1 and the updating condition as predictors. We used only participants as a random factor and estimated its random intercept. We observed that attractiveness ratings at Time 1 significantly predicted the ratings at Time 2, $t(746.00) = 26.47$, $p < .001$, so that the more attractive the face was perceived at Time 1, the more attractive it was perceived at Time 2. Importantly, this effect was moderated by the updating condition, $t(746.00) = 10.18$, $p < .001$, so that this relationship was larger in the consistent than in the opposite condition.

[7] This effect, however, was inconsistent as it emerged in opposite directions in Studies 2 and 3, suggesting that yet unidentified moderators could influence the direction of this effect.

[8] Note, however, that participants should not assume additional features from the behavioral information – for instance, the assumption that individuals taking part to beauty contest are especially vain or unpleasant in general (e.g., see Study 1).

Tables

Table 1

Mean (and standard deviation) values of personality ratings per attractiveness and updating condition for all studies

		Attractiveness condition		
Study	**Updating condition**	Low	Medium	High
1	/	2.82 (1.34)	2.78 (1.35)	3.15 (1.24)
	Opposite	2.87 (1.33)	2.82 (1.33)	3.15 (1.26)
	Consistent	2.77 (1.34)	2.74 (1.36)	3.14 (1.23)
2	/	2.70 (1.39)	3.00 (1.28)	3.22 (1.27)
	Opposite	2.77 (1.36)	3.01 (1.31)	3.16 (1.27)
	Consistent	2.64 (1.42)	2.99 (1.24)	3.27 (1.27)
3	/	2.92 (1.32)	*NA*	3.20 (1.20)
	Opposite	3.12 (1.16)	*NA*	3.09 (1.13)
	Consistent	2.70 (1.45)	*NA*	3.32 (1.25)

Note. Standard deviation values are presented in parentheses. The symbol "/" indicates that the reported mean values are independent of the updating condition and *NA* means "Non Applicable". The updating effect is represented by a larger halo effect – that is, a larger difference in ratings between high and low attractiveness faces – in the consistent than in the opposite condition.

Figure Legends

Figure 1

Halo effect (High-Low scores) at the trait level, as a function of the updating condition (opposite vs. consistent), trait halo relevance (left panel), and trait valence (right panel)

Note. Grey areas represent the 95% confidence intervals.

Figure 2

Halo effect (High-Low scores) at the trait level, as a function of the updating condition (opposite vs. consistent), trait halo relevance (left panel), and trait valence (right panel)

Note. Grey areas represent the 95% confidence intervals.

Figure 3

Moderated mediation model in Study 2

Note. The unstandardized regression coefficient representing the total effect is in parentheses. * $p < .05$, ** $p < .001$, *** $p < .001$.

Figure 4

Halo effect (High-Low scores) at the trait level, as a function of the updating condition (opposite vs. consistent), trait halo relevance (left panel), and trait valence (right panel)

Note. Grey areas represent the 95% confidence intervals.

Figure 5

Moderated mediation model in Study 3

Note. The unstandardized regression coefficient representing the total effect is in parentheses. * $p < .05$, ** $p < .001$, *** $p < .001$.