

# A Group's Physical Attractiveness Is Greater Than the Average Attractiveness of Its Members: The Group Attractiveness Effect

Personality and Social  
Psychology Bulletin  
1–16

© 2015 by the Society for Personality  
and Social Psychology, Inc  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0146167215572799  
pspb.sagepub.com



Yvette van Osch<sup>1</sup>, Irene Blanken<sup>1</sup>, Maartje H. J. Meijs<sup>1</sup>, and Job van Wolferen<sup>1</sup>

## Abstract

We tested whether the perceived physical attractiveness of a group is greater than the average attractiveness of its members. In nine studies, we find evidence for the so-called group attractiveness effect (GA-effect), using female, male, and mixed-gender groups, indicating that group impressions of physical attractiveness are more positive than the average ratings of the group members. A meta-analysis on 33 comparisons reveals that the effect is medium to large (Cohen's  $d = 0.60$ ) and moderated by group size. We explored two explanations for the GA-effect: (a) selective attention to attractive group members, and (b) the Gestalt principle of similarity. The results of our studies are in favor of the selective attention account: People selectively attend to the most attractive members of a group and their attractiveness has a greater influence on the evaluation of the group.

## Keywords

group attractiveness effect, group evaluation, selective processing, social cognition, physical attractiveness, cheerleader effect, perceptual averaging

Received May 8, 2014; revision accepted January 18, 2015

Identifying how we perceive and judge groups of people is important for understanding stereotypes (e.g., Fiske, Cuddy, Glick, & Xu, 2002). Differences in individual versus group perceptions may have implications in the formation or consolidation of stereotypes (Ford & Stangor, 1992; Willis, 1960). But how do people perceive and subsequently judge groups in terms of a particular trait? On one hand, we could assume that when judging a group, people take into account the individual variation within the group and thus come to a correct estimation of the average trait possessed by the group. On the other hand, we could also argue that groups are judged differently than a mere amalgamation of individuals, potentially resulting in a different judgment than the true average of all individuals in the group. Research mainly supports the claim that group impressions follow an *averaging rule*, such that the evaluation of a group reflects the average evaluations of its individual members (Anderson, 1965). For instance, both sequential and simultaneous group judgments of faces on good–bad and tense–relaxed dimensions mirrored the average judgments of those faces (Levy & Richter, 1963). Likewise, the likeability of a group of three faces was equivalent to the average ratings of the individual faces (Anderson,

Lindner, & Lopes, 1973). Similar results were obtained for the perceived likeability of attractive groups in a study by Miller and Felicio (1990), who presented subjects sequentially with either individual portrait photos or sets of photos. However, there are also indications that evaluations of groups and their group members may differ. For example, research has shown that a similar set of behavioral traits results in different impressions of groups versus individuals, when these traits are descriptive of one person compared to when each trait is descriptive of one person in the group (Hamilton & Sherman, 1996).

We set out to examine how people evaluate groups in terms of physical attractiveness. Physical attractiveness is an important trait (Gangestad, 1993) because many first impressions, and thus selection and subsequent evaluation, can be based on this obvious visual feature (e.g., Dion, Berscheid,

<sup>1</sup>Tilburg University, The Netherlands

## Corresponding Author:

Yvette van Osch, Tilburg University, Warandelaan, 2, Tilburg, 5037AB, The Netherlands.

Email: y.m.j.vanosch@tilburguniversity.edu

& Walster, 1972; Landy & Sigall, 1974; Watkins & Johnston, 2000). However, currently, very little is known about how people evaluate the physical attractiveness of groups. So far, the majority of research on physical attractiveness has studied the variability of judgments of physical attractiveness in isolation (Berry, 2000), meaning that individuals were evaluated in terms of physical attractiveness without a social context (e.g., Cunningham, Roberts, Barbee, Druen, & Wu, 1995). Recent efforts have revealed that context does matter. For instance, individuals presented in groups are rated differently in terms of attractiveness than when presented alone (Walker & Vul, 2014). In this article, we report nine studies that examine what happens when people evaluate the physical attractiveness of groups compared with the physical attractiveness of its individual members. This is theoretically relevant as we can test whether the *averaging rule* for group impressions also applies to judgments of physical attractiveness. Based on the averaging rule, one would expect the evaluation of a group's physical attractiveness to be based on the average attractiveness of its members. In our studies, however, people judge the physical attractiveness of a group and we observe that they find the group more attractive than the average of its members. We coin this effect the group attractiveness effect (GA-effect),<sup>1</sup> and it implies that the physical attractiveness of a group as a whole *does not* reflect the average attractiveness of its members.

To our knowledge, only one study did not support the averaging rule in group judgment (Willis, 1960). Interestingly, this single exception studied the physical attractiveness of groups, whereas the previously mentioned studies assessed other traits, such as likeability. Willis (1960) found that the evaluation of groups is more extreme than the average of its members. In his experiments, participants saw sets of two or three portrait photos of relatively attractive and unattractive faces and judged the *average attractiveness* of the group. Relatively attractive groups were rated as more attractive than the mean attractiveness of the group members,<sup>2</sup> which is in line with the GA-effect.

Thus, to our best knowledge there is only one study on the difference between group and individual evaluations of physical attractiveness, and it is the only one to challenge the applicability of the averaging rule. We believe further examination of group judgments of physical attractiveness is necessary for two reasons. First, the only study that finds evidence in line with the GA-effect is based on evaluations of sets of portrait pictures that were composed specifically for the purpose of the experiment. The portraits were also pretested in terms of relative attractiveness and combined to create specific group compositions. Methodologically, these are strong features of the experiment, but none of the materials showed target individuals in a way that resembles a natural encounter with a group (this holds for all other previously mentioned studies). We report tests of the GA-effect using more natural stimuli and larger groups of people. Second, the literature is silent on why such a GA-effect may occur and

we aim to uncover the underlying psychological mechanism that drives this effect.

## Selective Attention

Several mechanisms could produce such a GA-effect. First, attractive individuals could capture attention more strongly than unattractive individuals in a group. This idea is based on the observation that people selectively attend to physically attractive female targets. Specifically, people spend proportionally more time looking at more attractive female faces than less attractive female faces (Maner et al., 2003, Study 4) and they overestimate the number of attractive women in a set of 15 simultaneously presented faces (Maner et al., 2003, Study 1). Selective attention could cause the most attractive members of a group to have a relatively large influence on the evaluation of the group. Put differently, if people selectively attend to the most attractive members in the group, it is unlikely that the *averaging rule* is descriptive of the evaluative process. Attending to attractive individuals in a group seems adaptive because physical attractiveness seems key in mating success (Gangestad, 1993) and might aid gene survival (Maner et al., 2003). In general, physical attractiveness is associated with success in life (Langlois et al., 2000). For example, physically attractive individuals are perceived to also be "good" in other traits (Dion et al., 1972) and have higher incomes (Judge, Hurst, & Simon, 2009). Thus, selectively attending to attractive others of the opposite sex may function as a heuristic for mate selection. Likewise, attending to attractive individuals of the same sex may give you an advantage too because it helps you scale potential rivalry and improve your own opportunities (Maner, Gailliot, Rouby, & Miller, 2007). Moreover, selective attention toward attractive individuals is in line with the notion that people selectively attend and seek out information that is in line with their attitudes (Eagly & Chaiken, 1993).

From this selective attention account, we can derive several testable hypotheses. First of all, if people pay selective attention to the more attractive group members, forcing attention on all members of the group should eliminate or attenuate the GA-effect. Second, if there is selective attention for the most attractive individuals in a group, we should see effects in memory: If people pay more attention to the more than to the less attractive group members, they should be better at remembering the more rather than the less attractive group members. Third, the attractiveness ratings of the most attractive group members should be most predictive of the group rating. Fourth, if group members do not differ much in the extent to which they are seen as physically attractive, a smaller GA-effect should occur than when there is great variation in attractiveness. To illustrate, if people judge the attractiveness of a group by looking at only a few members (i.e., the most attractive ones), the judgment will not differ much from the average of the individual ratings

when all members of a group are equally attractive. So, perceived heterogeneity in terms of attractiveness in a group should foster the occurrence of the GA-effect. Fifth, selective attention should be reflected in how long people look at the relatively attractive individuals in a group. In addition, the longer people look at the most attractive group member, the larger the GA-effect should be. Finally, if selective attention explains the GA-effect, larger groups should produce larger GA-effects. When judging larger groups, it is more likely that individuals are overlooked than when judging small groups.

We must note that the studies supporting the idea for selective attention to attractive faces only found these effects for female targets, not male targets. Men and women both selectively attended to attractive females and overestimated their frequency in a group. Women also selectively attended to male targets, but their selective attention did not lead to a biased estimate, whereas men did not selectively attend nor overestimate the frequency of attractive male targets presented in a group (Maner et al., 2003). These results are explained in terms of mating-related motives. Specifically, uncommitted men and women were more likely to focus on attractive members of the opposite sex than committed women (not men). This reasoning is in line with data revealing that men and women only differ in their preferences for attractive mates in the context of a long-term relationship (Meltzer, McNulty, Jackson, & Karney, 2014). To test whether selective attention is indeed different for men and women, we checked for gender effects in all studies.

## Group Gestalt

Another explanation relies on the basic tenet of Gestalt psychology in which the whole is different from the sum of its parts. In the studies we report, this means that the whole is different from the *mean* of its parts. Considering the group as the Gestalt and the individual group members as the individual constituents, the GA-effect could be caused by global precedence, meaning that the Gestalt is processed (and possibly judged) before the more local features are taken into account (i.e., the individual group members; Navon, 1977). Global precedence, however, seems to hold in some but not all perceptual processes. As such, Gestalt psychologists have suggested the concept of holistic properties, which are “emergent properties that cannot be predicted by considering only the individual component parts” and “they arise from the interrelations between the parts” (Wagemans, Feldman, et al., 2012, p. 1223). These holistic properties seem to have primacy in visual processing, which could theoretically explain why people first see the group and subsequently its members. However, this process of primacy or global precedence cannot explain why the attractiveness ratings of the group as a whole would be consistently higher, as opposed to lower, than the mean attractiveness of its members. Willis (1960) found that relatively attractive groups

(scoring 4.15 on a scale from 0 to 8 in pretests) were judged to be more attractive as a group (4.58), and that relatively unattractive groups (scoring 2.15 in pretests) were judged to be less attractive as a group (2.02; both these differences were statistically significant). It could thus be the case that when a group is relatively attractive, the group as a whole is judged to be even more attractive, whereas if a group is relatively unattractive, the group as a whole is judged to be even less attractive. To date, Gestalt psychology exclusively considers abstract, simple nonsocial stimuli and does not provide us with clues regarding the perception of complex social stimuli such as social groups (Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012), let alone the perception and judgment of such groups (for first ventures in that direction see Martin, Fowlkes, Tal, & Malik, 2001, and Pinna, 2012).

A basic principle in Gestalt psychology describing how elements are grouped to create a Gestalt, however, may give us some clues as to how to approach the Gestalt-like aspects of the GA-effect. The principle of similarity suggests that the most similar elements are grouped together (Wagemans, Elder, et al., 2012), suggesting that in more homogeneous groups the GA-effect is more likely to occur. This idea is corroborated by the fact that Willis (1960) found that more homogeneous groups received the highest group judgments in comparison with the groups where group members deviated more from the pretested mean scale value of the group. Interestingly, this prediction conflicts with the prediction we derived before from our argument that the GA-effect is caused by selective attention. The selective attention argument predicts that the GA-effect is smaller, not larger, for homogeneous groups.

In sum, we investigate whether physical attractiveness ratings of a group follow the *averaging rule*. We report nine experimental studies. All studies demonstrate the existence of the GA-effect in female, male, and mixed-gender groups for natural and constructed groups of females. A meta-analysis on the 33 comparisons between group versus individual evaluations of physical attractiveness reveals that the effect is medium to large. Finally, we explored whether the GA-effect is caused by the two mechanisms we proposed in this introduction.

## Method and Results

### Analytical Strategy

Statistical results of all analyses for all studies can be found in Figure 1. We report how we determined our sample size (see Online Appendix I for power calculations), all data exclusions (only one participant in Study 9b, see Footnote 10), all manipulations, and all measures in our studies (cf. Simmons, Nelson, & Simonsohn, 2012). All studies were programmed in Qualtrics online survey software or in Tobii Studio 3.1 software.

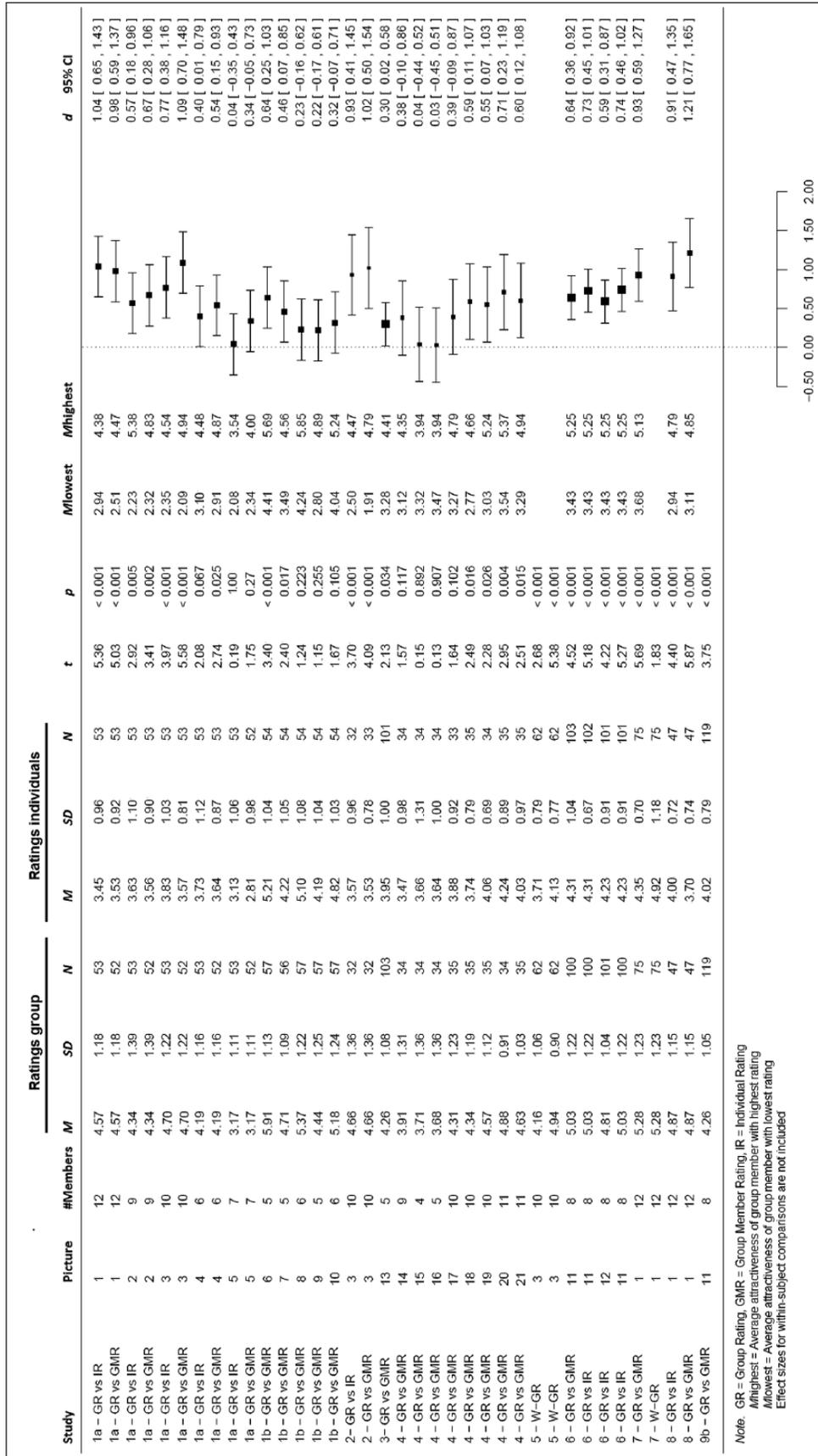


Figure 1. Overview of study results including the forest plot from the meta-analysis.

Note. GR = Group Rating, GMR = Group Member Rating, IR = Individual Rating  
 Mhighest = Average attractiveness of group member with highest rating  
 Mlowest = Average attractiveness of group member with lowest rating  
 Effect sizes for within-subject comparisons are not included

### Study 1a: Demonstration in Female Groups

In one condition, we asked people to look at and rate a female group as a whole (group-rating condition). In two other conditions, we asked people to look at and rate individual females. Research has shown that the perception of objects is affected by the ensemble in which they are presented (e.g., object size; Brady & Alvarez, 2011). More specifically, it has been suggested that people's perception of persons may also be influenced by the social context (e.g., Walker & Vul, 2014). To check for such context effects, we included two control conditions. In one condition, people saw and rated cropped portrait photographs of faces sequentially (the individual-rating condition); in another condition, they saw the whole group at once but rated each individual separately (the group-member condition; see Figure 2). The inclusion of these two individual conditions was based on two considerations: (a) Differences between the group condition and the individual-rating condition could be attributed to the mere lack or presence of context. Differences could also be attributed to the fact that in the individual-rating condition people evaluate people on different stimuli (only the face) than in the group condition (face and part of or the whole body). Including the group-member condition could rule out these alternative explanations. (b) Context effects may also explain differences between the group-rating condition and the group-member condition. Specifically, research has shown that the judgment of schematic faces is influenced by the differences in features (such as width of the nose) between the target face and other schematic faces surrounding the target face (e.g., Wedell & Pettibone, 1999). Perhaps, rating each of the individuals in the group-member condition elicits comparisons between the individuals in the group and thereby alters judgments compared with the group-rating condition that might not elicit interindividual comparisons. The individual-rating condition that does not have the (social) context allows us to check for the occurrence of such context effects.

A total of 158 Tilburg University students (number of female participants [ $\square$ ] = 134;  $M_{\text{age}} = 19.53$ , range<sub>age</sub> = 18-42) evaluated the physical attractiveness of five different groups of women.<sup>3</sup> The order of presentation was randomized and participants were randomly assigned to one of three conditions; in the group-rating condition ( $n = 53$ ) participants evaluated the attractiveness of the group as a whole: "How attractive do you find this group?" (1 = *not at all attractive*, 7 = *very attractive*). In the group-member condition ( $n = 52$ ), each member in the photo of the group-rating condition was assigned a number and participants evaluated the attractiveness of each group member individually: "How attractive do you find group member X?" In the individual-rating condition ( $n = 53$ ), participants also rated the attractiveness of the group members individually, but based on cropped images of the women, and thus never saw the group as a whole. For the latter two conditions, we averaged the

individual attractiveness ratings to obtain the groups' mean attractiveness rating. This study had a mixed 5 (group: 5 different groups)  $\times$  3 (rating: group vs. group member vs. individual) design with group as within-subject factor and rating as between-subject factor.

We found a GA-effect for four out of five groups, such that the attractiveness ratings of the group as a whole were significantly higher than the mean attractiveness rating of its members. The group for which we did not find a GA-effect was rated as the least attractive group, scoring below the midpoint of the scale (see Figure 1).

### Study 1b: Replication with Different Stimuli

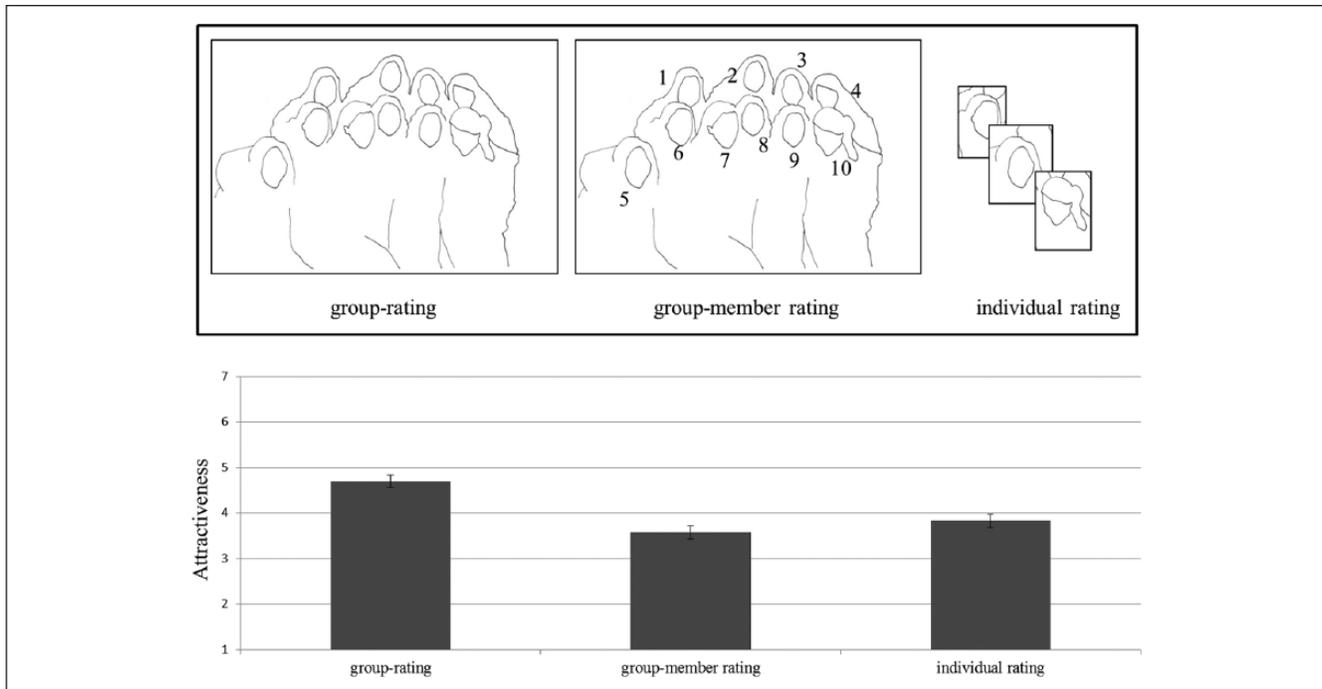
A total of 111 MTurk-workers ( $\square = 45$ ;  $M_{\text{age}} = 31.05$ , range<sub>age</sub> = 18-74) evaluated the physical attractiveness of five other groups of women than in Study 1a. We only administered the group-rating condition ( $n = 57$ ) and group-member condition ( $n = 54$ ; 1 = *very unattractive*, 7 = *very attractive*).<sup>4</sup> The GA-effect was found for two of the five groups (see Figure 1).

### Study 2: Ruling Out Social Attractiveness

To rule out the possibility that we found differences between conditions because participants rated different types of attractiveness in the group and the individual conditions, we replicated Study 1a. However, we only used Photograph 3 (displayed in Figure 2) and asked respondents on what type of attractiveness they had judged the women. We used Photograph 3 because it portrays a group of typical Dutch female university students. The participants in this study are Dutch university students and thus it seemed most natural to confront them with a group that they could actually encounter.

Ninety-three Tilburg University students ( $\square = 85$ ;  $M_{\text{age}} = 19.38$ , range<sub>age</sub> = 17-27) provided attractiveness ratings in the same three between-subject conditions as in Study 1a. Subsequently, they indicated whether they had based the attractiveness rating of the women on (a) physical attractiveness (how pretty the people in the photograph were), (b) friendliness (how kind the people in the photograph seemed), or (c) social attractiveness (to what extent you would want to belong to these people).

We replicated the GA-effect (see Figure 1) and found no differences between conditions on what the attractiveness rating was based,  $\chi^2(2, N = 97) = 4.24, p = .374$ . Overall, the majority of participants reported judging on the basis of physical attractiveness (70.1%). In addition, an ANOVA, with both condition and type of attractiveness rated as factors on the attractiveness ratings, revealed only a main effect of condition,  $F(2, 93) = 7.76, p = .001, \eta_p = .14$ , not a main effect of type of attractiveness rated,  $F(2, 93) = 0.44, p = .646, \eta_p = .01$ , nor an interaction effect between these two factors,  $F(4, 93) = 0.34, p = .852, \eta_p = .01$ . To be sure, though, we specifically asked to rate "physical attractiveness" in all subsequent studies.



**Figure 2.** Abstract version of Photograph 3 to illustrate the three between-subject conditions, including the mean attractiveness evaluations per condition for this specific photograph in Study 1a. Note. Error bars indicate standard error of the mean.

### Study 3: Male Group

The previous studies only tested the GA-effect in female groups. Here we test whether the effect occurs in a male group. A total of 204 MTurk-workers ( $n = 53$ , 1 gender missing;  $M_{\text{age}} = 28.11$ ,  $\text{range}_{\text{age}} = 18-63$ ) rated the physical attractiveness of our male author and four of his friends either as a group (group-rating condition;  $n = 103$ ) or each man separately (group-member condition;  $n = 101$ ; 1 = *very unattractive*, 7 = *very attractive*). For this male group, the GA-effect also emerged (see Figure 1).

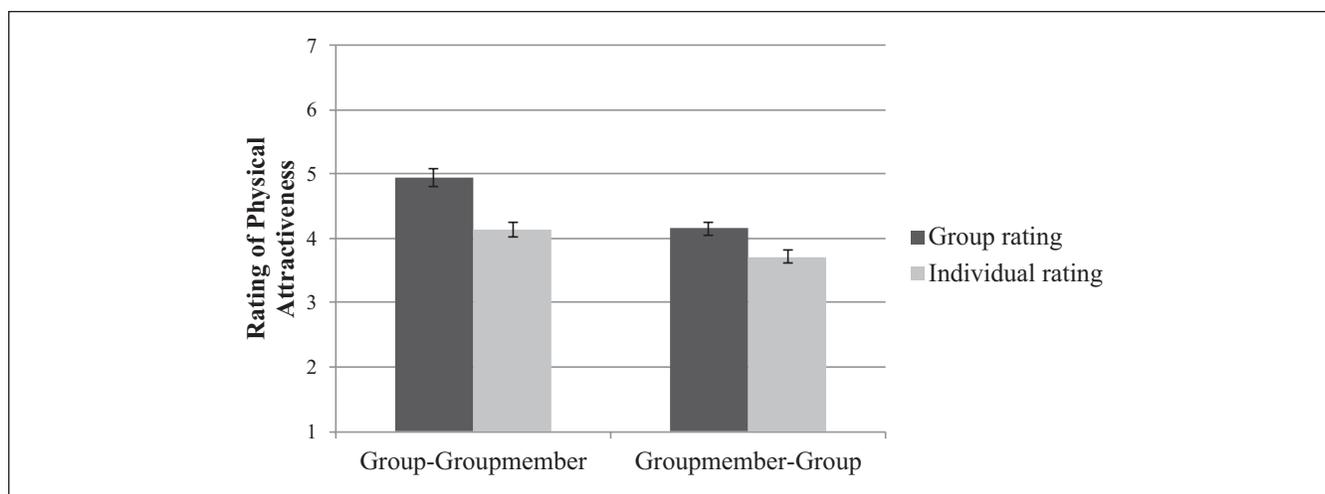
### Study 4: Mixed-Gender Groups

In this study, we tested for GA-effects in mixed-gender groups. A total of 550 MTurk-workers ( $n = 163$ , 1 gender missing;  $M_{\text{age}} = 28.24$ ,  $\text{range}_{\text{age}} = 18-75$ ) were randomly assigned to 1 of 16 conditions in an 8 (group: 8 different groups)  $\times$  2 (rating: group vs. group member) between-subjects design ( $n$  per condition was 34 or 35). Three photographs depicted groups of exclusively college-age men (the groups consisted of 4, 5, and 9 males). The other five photographs depicted mixed-gender college-age women and men (female:male ratios were 1:9, 2:8, 3:7, 4:7, 5:6). Subjects were asked, "How physically attractive do you find these men/women and women?" and rated this for either the group as whole or for each group member individually (1 = *very unattractive*, 7 = *very attractive*).

The GA-effect was established in four of the five mixed groups (see Figure 1). The data revealed a marginally significant effect for the mixed group with 9 males and 1 female, and for the group with 9 males. The GA-effect was absent for the smaller groups of 4 and 5 males.

### Study 5: Within-Subject Test

In this study, we aimed to replicate the GA-effect in a within-subject design, and additionally tested several hypotheses as suggested by the mechanisms of selective attention and group gestalt. The selective account suggests that if we force people to pay more attention to all members of the group, instead of naturally focusing on the most attractive ones, the GA-effect should disappear or be attenuated. To do so, we asked half of our participants to first rate the attractiveness of all group members individually, and subsequently the group as a whole. The other half first rated the group and then all individuals in the group. According to the selective attention account, the GA-effect should only appear, or be larger, in the second than in the first order. The proposed mechanism of selective attention also predicts that the rating of the most attractive group member has the biggest influence on the group rating. The within-design of Study 5 allowed us to test this prediction. We used each participant's ratings of all individual group members to predict his or her overall group judgment. In addition, the gestalt explanation predicts that the perceived homogeneity (similarity in terms of attractiveness) of the group interacts



**Figure 3.** Attractiveness ratings per condition for Study 5.  
Note. Error bars represent standard errors.

with the difference between the group and the individual ratings, such that greater homogeneity results in a greater GA-effect. In contrast, the selective attention account predicts that greater homogeneity results in a smaller GA-effect. In this study, we thus also controlled for homogeneity to see whether it influences the GA-effect. A total of 124 Tilburg University students ( $n = 101$ ;  $M_{\text{age}} = 20.46$ ,  $\text{range}_{\text{age}} = 18\text{-}29$ ) evaluated the physical attractiveness of the women in Photograph 3 from Study 1a. Again, we used this photo because it portrays a group of typical Dutch female university students and participants in this study are Dutch university students. This study had a mixed 2 (rating: group vs. individual rating)  $\times$  2 (order: rate group first then group members vs. rate group members first then group) design with rating as within-subject factor and order as between-subject factor. Participants either rated the group as a whole first (“How attractive do you find these women?” 1 = *not at all attractive*, 7 = *very attractive*) and then its group members (“How attractive do you find woman X?”  $n = 62$ ), or they rated all group members first, and then the group as a whole ( $n = 62$ ).

A mixed repeated-measures ANOVA with rating and order as factors revealed a main effect of rating, Wilks’s  $\lambda = .53$ ,  $F(1, 122) = 109.41$ ,  $p < .001$ ,  $\eta_p^2 = .47$ , a main effect of order,  $F(1, 122) = 16.16$ ,  $p < .001$ ,  $\eta_p^2 = .12$ , qualified by an interaction effect between these two factors, Wilks’s  $\lambda = .93$ ,  $F(1, 122) = 8.92$ ,  $p = .003$ ,  $\eta_p^2 = .07$ . Thus, the data revealed a GA-effect in both conditions. Regardless of order group, attractiveness ratings were higher than the mean of the individual ratings (see also Figure 1). When participants rated the group first and then its members individually, the effect size was twice as large as when participants first rated all members individually and then the group as a whole (see Figure 3). These data thus suggest that if people are more aware of the variation in the group in terms of attractiveness, the GA-effect is smaller. This could imply that if not made

aware of the individual variation, people generally focus on the more attractive group members and thus judge the group to be more attractive than the average attractiveness of its members. Please note that overall ratings were lower when participants first rated the individuals and then the group, than if the order was reversed. It could be the case that the initial rating of the group served as an anchor for the subsequent ratings. If participants experienced the GA-effect in the initial group rating, and thus found the group high on attractiveness, they might use their judgment as an anchor for their subsequent judgments of the individuals.

To test whether the ratings of the more attractive group members have a greater influence on the group rating, we ran simultaneous linear regressions, collapsed across order. The ratings of three of the six individuals who scored higher on attractiveness than the scale midpoint (4) significantly predicted the group rating (all  $\beta$ s  $> .18$ ,  $t$ s  $> 2.23$ ,  $p$ s  $< .029$ ), including the rating of the most attractive individual. None of the ratings of the four individuals who scored below the scale midpoint significantly predicted the group rating (all  $\beta$ s  $< .12$ ,  $t$ s  $< 1.38$ ,  $p$ s  $> .171$ ). In addition, the average attractiveness of each target individual was positively, though not significantly, associated with the beta of that individual’s score in the regression analysis. With only 10 observations, this correlation could suggest that the more attractive a target, the more predictive the target is for the group,  $r(8) = .43$ ,  $p = .22$ . It thus seems that the more attractive targets in the group are indeed more predictive of the group judgment, indicating that when judging the group people are more focused on the more than the less attractive group members.

To test whether the perceived homogeneity of the group influences the GA-effect, we calculated the standard deviation of each participant’s individual ratings and entered this variable as a covariate in the mixed repeated-measures ANOVA. We find an interaction such that lower homogeneity

(larger standard deviation) resulted in greater GA-effects, Wilks's  $\lambda = .96$ ,  $F(1, 121) = 8.92$ ,  $p = .031$ ,  $\eta_p^2 = .04$ . The main effect of rating was no longer significant, Wilks's  $\lambda = .99$ ,  $F(1, 122) = 1.52$ ,  $p = .22$ ,  $\eta_p^2 = .01$ , whereas the main effect of order,  $F(1, 121) = 15.93$ ,  $p < .001$ ,  $\eta_p^2 = .12$ , and the interaction between the two, Wilks's  $\lambda = .93$ ,  $F(1, 122) = 8.92$ ,  $p = .002$ ,  $\eta_p^2 = .08$ , still were. Thus, controlling for heterogeneity in the analysis (i.e., including *SD* as a covariate) eliminated the GA-effect. A follow-up regression analysis revealed that the more variation participants perceived in the attractiveness of the individuals (as captured in the *SD* of the individual attractiveness ratings per participant) predicted the size of the GA-effect. This effect was marginally significant,  $\beta = .31$ ,  $t = 1.96$ ,  $p = .052$ . The data thus suggest greater perceived heterogeneity in attractiveness, and not homogeneity, increases the GA-effect.

### Study 6: Ruling Out Methodological Artifacts and Testing for Similarity

In this study, we sought to replicate the GA-effect with photographs we construed ourselves and to rule out alternative explanations based on our methodology. The specific presentation of the individuals and groups in the previous studies might have interfered with the attractiveness ratings themselves. To test this proposition, we took several photographs of a female field hockey team consisting of eight group members and had them rated on physical attractiveness. We varied different aspects of the presentation of individuals and groups to test three things. First, we tested if the GA-effect is affected by whether the individuals are indicated by rectangles or numbers. Second, in all previous studies, the group-member conditions showed numbers, whereas the group condition did not. Therefore, we created one condition where we added numbers to the picture presented in the group-rating condition. Third, we tested whether the group ratings are different when the "group picture" is composed of portrait photos of each individual compared with more natural group pictures we used in the previous studies.

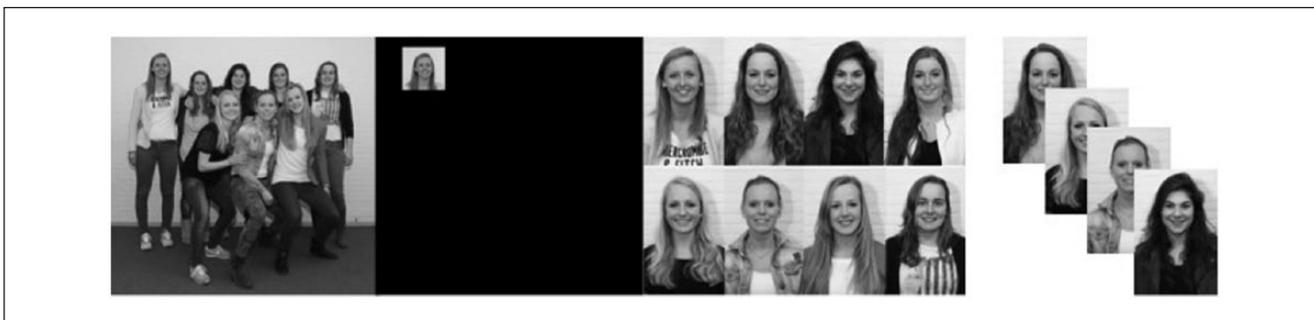
A total of 697 MTurk-workers ( $\bar{M}_{age} = 28.67$ , range<sub>age</sub> = 16-76) were randomly allocated to one of seven conditions (also see Figure 4); (1) In this condition, the women were dressed casually and participants rated the attractiveness of the group: "How physically attractive are these women?" (1 = *very unattractive*, 7 = *very attractive*). (2) Participants rated all individual group members on the same photograph ("How physically attractive is this woman?"), but instead of using numbers, as we had done in the previous studies, each member was subsequently indicated by a red rectangle around the face (order of presentation randomized). (3) This condition was similar to Condition 2, but instead of a red rectangle, the entire photo was covered up (made black), except for a square around the target

member. The average group ratings in Conditions 2 and 3 were compared with the group rating in Condition 1. (4) We also took individual photographs of the group members and presented these pictures to participants in a grid. Participants were asked to rate the group as a whole. (5) In this condition, participants rated the individual photographs of the group members serially (one by one). The ratings of Conditions 4 and 5 were compared. In addition, we compared the average rating of the individual photographs with the group rating in Condition 1. (6) In this condition, participants saw the same photograph as the participants in Condition 1 and they rated the group as a whole, but the photograph now contained numbers alongside the individual members. This allowed us to check whether the mere presence of numbers would interfere with the attractiveness rating. (7) Participants rated the attractiveness of the same group of women, but this time not in casual outfits but in their field hockey apparel. We included this condition to check whether the similarity of group members would enhance the attractiveness of the group. If members of a group look more similar, this could increase the perceived "groupness" of the group (Wagemans, Elder, et al., 2012) and thus enhance the GA-effect in line with the gestalt account. The latter two conditions were compared with the group ratings of Condition 1 (see Figure 4 for the photographs of Conditions 1, 3, 4, and 5).

A GA-effect was found for each comparison between group- and individual/group-member ratings (see Figure 1).<sup>5</sup> In addition, the group ratings in Condition 1 did not differ from the group ratings in Condition 6 (with numbers posted next to the women). They also did not differ from the ratings of the group wearing field hockey apparel (Condition 7), suggesting that if groups are displayed as more similar (wearing all the same outfit) they were not rated as more attractive than when they were presented as more dissimilar (wearing all different outfits). Also, comparable GA-effects were found if we compared the group photographs in Conditions 1 and 7 with the average of the individual ratings in Conditions 2, 3, and 5: In all cases, the effect size was medium to large (for Condition 1 comparisons,  $d = .67-.74$ , for Condition 7 comparisons,  $d = .50-.57$ ).

### Study 7: Take a Closer Look!

In this study, we forced people to pay more attention when making the group rating. We expected this would reduce the difference between the attractiveness rating of the group and the average attractiveness ratings of its members. A total of 150 MTurk-workers ( $\bar{M}_{age} = 31.55$ , range<sub>age</sub> = 18-67) evaluated the physical attractiveness of the group of women in Photograph 1 from Study 1a. In the group-rating condition ( $n = 75$ ), participants evaluated the attractiveness of the group as a whole: "How physically attractive do you find these women?" (1 = *very unattractive*, 7 = *very attractive*). These participants were subsequently told that the experimenters wanted their answer to be carefully considered, and



**Figure 4.** Photographs of Conditions 1, 3, 4, and 5 in Study 6.

were instructed to look at the photograph *again* while the option to proceed to the next screen was disabled for 20 seconds. Then they rated the attractiveness of the group of women for the second time on the same scale. In the group-member condition ( $n = 75$ ), participants rated the women individually.

First, the data revealed the GA-effect, so in the group-rating condition (first rating) the women received higher attractiveness scores than in the group-member condition (see Figure 1). In addition, within the group-rating condition the first time people took significantly fewer seconds ( $M = 12.12$ ,  $SD = 7.26$ ) and rated the group as more attractive ( $M = 5.28$ ,  $SD = 1.23$ ) compared with the second time, seconds:  $M = 32.52$ ,  $SD = 17.14$ ,  $t(74) = -9.80$ ,  $p < .001$ ; attractiveness:  $M = 4.92$ ,  $SD = 1.18$ ,  $t(74) = 4.65$ ,  $p < .001$ . On average, people adjusted their attractiveness rating of the group downward. In fact, 62.7% did not alter their rating of the group the second time, 26.7% decreased their attractiveness rating with 1 scale point, and 6.7% decreased their rating with 2 scale points. Only three participants (4%) increased their evaluation by 1 scale point. People who decreased their evaluation had higher first group ratings ( $M = 5.76$ ,  $SD = 0.83$ ) than people who did not decrease or increase their evaluation,  $M = 5.04$ ,  $SD = 1.32$ ;  $t(73) = 2.48$ ,  $p = .015$ ,  $d = 0.58$ . When we compare the second group rating with the individual-rating condition, we find an attenuated but still significant GA-effect,  $t(148) = 3.60$ ,  $p < .001$ ,  $d = 0.59$  (vs.  $d = 0.93$  when we compare the first group rating with the individual-rating condition). Thus, if people pay more attention while rating a group on attractiveness on average their ratings are lower.

### Study 8: Whom Do You Remember?

In this study, we tested how well the most versus least attractive group members were remembered. If indeed ratings of groups are based on the most attractive members because initial attention is drawn to them, one could expect participants to remember less attractive group members less than more attractive group members.

A total of 141 undergraduate students at Tilburg University ( $n = 141$ ,  $\bar{M}_{age} = 19.53$ ) rated the entire

group ( $n = 47$ ), each member in the group picture ( $n = 47$ ), or each group member individually ( $n = 47$ ), using Photograph 1 from Study 1a. Subsequently, they were shown 12 separate individual photographs in a randomized order. Six of these photographs were pictures of women they had not seen before, whereas the other six photographs were those of three members who were rated most attractive ( $M_1 = 3.96$ ,  $M_2 = 4.43$ ,  $M_3 = 4.47$ ) and three members who were rated least attractive ( $M_4 = 2.51$ ,  $M_5 = 2.96$ ,  $M_6 = 3.00$ ) in Study 1a. We replicated the GA-effect (see Figure 1). Participants were then asked whether they had seen the woman in the picture before or not. A nonparametric Kruskal–Wallis ANOVA<sup>6</sup> revealed no differences in memory for the most attractive group members between the group condition ( $M$  rank = 64.90), the group-member condition ( $M$  rank = 77.10), and the individual-rating condition ( $M$  rank = 71.00),  $H$  (corrected for ties) = 5.487,  $df = 2$ ,  $n = 141$ ,  $p = .064$ , Cohen's  $f = 0.202$ . Interestingly, there was a significant effect of condition on memory for the least attractive group members,  $H$  (corrected for ties) = 29.399,  $df = 2$ ,  $n = 141$ ,  $p < .001$ , Cohen's  $f = 0.516$ , with a smaller proportion remembered in the group condition ( $M$  rank = 48.11) than in the group-member condition ( $M$  rank = 83.78) and the individual-rating condition ( $M$  rank = 81.12). Online Appendix II provides an overview of the calculations that we performed to construct separate hit rates for the most and least attractive group members and false alarm rates for the new (unseen) photographs.

In sum, we find that people who judge a group of people on physical attractiveness are less likely to pay attention to the less attractive people in that group.

### Study 9a: Eye-Tracking Data

In this study, we wanted to find out whether more attractive group members receive more visual attention than less attractive group members using eye tracker methodology. As mentioned in the introduction, Maner and colleagues (2003, Study 4) established that observers “were biased toward attending selectively to physically attractive, as compared with less attractive, female targets” (p. 1116) using eye

tracking. Here, we tested whether this was also the case for our stimuli.

Students at Tilburg University ( $n = 50$ ,  $\text{♀} = 31$ ;  $M_{\text{age}} = 21.14$ ) participated voluntarily. In a soundproof cubicle, they were seated in front of the computer screen of a Tobii T60 eye-tracking system. After calibration, participants saw three photographs we used as stimuli in the previously reported studies.<sup>7</sup> When they felt ready to make a judgment, participants proceeded to the next screen by pressing the space bar and rated the attractiveness of the group, again on a 7-point scale.<sup>8</sup> The Tobii Studio 3.1 software allowed us to indicate each face as an equally sized area of interest (AOI) and measure the total amount of time (in seconds) participants fixated on a particular AOI. Based on the attractiveness ratings in Studies 1 and 6, we selected the most attractive and least attractive group member in each group, and calculated for each participant the proportion of time they had fixated on these AOIs (conform Maner et al., 2003).

A repeated-measure ANOVA, with photograph and target (most attractive vs. least attractive group member) as within-subject factors and proportion of time fixated on these targets as dependent variable, revealed a significant main effect of target, Wilks's  $\lambda = .78$ ,  $F(2, 44) = 12.15$ ,  $p = .001$ ,  $\eta^2 = .22$ . Proportionally, more time was spent fixating on the most attractive faces than on the least attractive faces across the three photographs. There was no main effect of photograph ( $p = .264$ ) and no interaction effect between target and photograph ( $p = .702$ ).

### Study 9b: Connecting the GA-Effect to Eye-Tracking Data

In Study 9a, we replicated the effect found by Maner and colleagues (2003, Study 4): Participants proportionally spent more time looking at the most than least attractive faces. This, however, does not directly reveal that the GA-effect is driven by selective attention. In Study 9a, we did not test for the GA-effect because we did not ask participants for both individual and group ratings. In this study, we set out to test whether there is a relationship between the GA-effect (the difference between a group rating and the average of individual ratings in that group) and the time spent looking at more and less attractive targets. Please note that attractiveness of targets in Study 9a was based on attractiveness ratings in our previous studies; in this study, we used the attractiveness of targets as defined by the participant himself or herself. The selective attention account suggests that people who experience the GA-effect do so because they proportionally attend more to the most than the least attractive individuals in the group. We thus expected that there is a positive correlation between the GA-effect (group rating—average individual ratings) and the time spent fixating on the most attractive target(s) for participants who experience the GA-effect (for whom the group rating was higher than average of individual ratings).

We ran this study for three consecutive weeks in our lab. Again, participants were Tilburg University students ( $n = 120$ ,  $\text{♀} = 88$ , 3 participants did not report gender;  $M_{\text{age}} = 20.78$ ). We ran the study on the same eye tracker, and employed the same procedure as in Study 9a with three alterations. First, participants only rated the photograph used in Study 6 (Figure 4, panel 1). Second, participants rated both the group as a whole (while they were in the cubicle, while collecting eye-tracking data), and they rated each individual in the group separately (while they were in a room adjacent to the cubicle). To collect the individual ratings, the same photograph was printed on paper with numbers displayed next to the individuals. The order of this procedure was randomized: 62 participants rated the individuals before they went into the cubicle where they saw the photograph again and rated the group while their eyes were being tracked, 58 participants first rated the group on the eye tracker and then left the cubicle to rate all individuals in the group. Third, all ratings were collected on paper to prevent data loss as we experienced in Study 9a. We employed the same 7-point scales as in our other studies.

For each participant, we calculated the GA-effect (group rating—average individual ratings) and the proportions they spent looking at what they indicated were the most and least attractive individuals. For many participants, we could not use one individual as the most attractive one because they, for instance, rated three individuals as most attractive (all of them received a 5 out of 7). Therefore, the proportion of time spent looking at the most or least attractive individuals was calculated as the mean proportion time spent looking at all of those rated highest or lowest on individual attractiveness.<sup>9</sup>

First, a mixed repeated-measures ANOVA with rating and order as factors revealed a main effect of rating, Wilks's  $\lambda = .90$ ,  $F(1, 117) = 13.70$ ,  $p < .001$ ,  $\eta_p^2 = .11$ , but no main effect of order,  $F(1, 117) = 0.96$ ,  $p = .329$ ,  $\eta_p^2 = .01$ , nor an interaction effect between these two factors, Wilks's  $\lambda = .97$ ,  $F(1, 117) = 3.23$ ,  $p = .075$ ,  $\eta_p^2 = .03$ . So, regardless of order, the data revealed a GA-effect. Second, a paired-sample  $t$  test revealed that participants spent more time fixating on the most ( $M$  proportion = .15) than the least attractive faces ( $M$  proportion = .11),  $t(113) = 2.01$ ,  $p = .046$ , replicating our findings from Study 9a.

Third, we correlated the effect size with the proportions spent looking at the most and least attractive group members for all participants. The proportion of time spent looking at the most attractive members was not correlated to the GA-effect,  $r(114) = .02$ ,  $p = .854$ , and neither was the proportion of time spent looking at the least attractive members,  $r(114) = -.02$ .

We only hypothesized a positive correlation for those people who experienced the GA-effect (a positive effect size), so we also ran the correlation analyses reported in the previous paragraph separately for this subgroup of participants. We found a nonsignificant but positive correlation

between the GA-effect and the proportion spent looking at the most attractive individuals,  $r(76) = .18$ ,  $p = .115$ , regardless whether the individual ratings were collected first,  $r(44) = .21$ ,  $p = .180$ , or last,  $r(32) = .15$ ,  $p = .400$ . We did not find similar correlations for the proportion spent looking at the least attractive individuals,  $r(76) = .02$ ,  $p = .883$ , individual ratings first:  $r(44) = .08$ ,  $p = .603$ , and individual ratings last:  $r(32) = -.02$ ,  $p = .916$ . This final analysis thus revealed a positive, though nonsignificant, relationship between attention for the most attractive targets and the size of the GA-effect, but only for those who perceived the group to be more attractive than the average attractiveness of its members.

### Gender Effects

We reran all our analyses and tested for main and interaction effects of participant gender. Based on previous studies (e.g., Maner et al., 2003), we expected that selective attention to attractive individuals could differ depending on gender of the participant. In all but one study, we found significant GA-effects after including participant gender. In Study 4, the effect of condition on attractiveness rating became marginally significant after including participant gender,  $p = .06$ , but for all separate comparisons the group ratings were higher than the average of the individual ratings. Further, a main effect of participant gender was found in Studies 2, 3, 4, and 6; in Studies 3, 4, and 6, female participants rated the attractiveness of the people in the pictures higher than male participants and in Study 2 the reverse was found. Study 5 was the only study in which we found an interaction effect between condition and participant gender; we found the GA-effect for both genders, but the GA-effect was stronger for female participants than for male participants. Statistical results of all analyses for all studies can be found in Online Appendix III. These analyses reveal that gender does not have a large effect on the occurrence of the GA-effect.

### Meta-Analysis

We examined the overall GA-effect size through conducting a meta-analysis.<sup>10</sup> We conducted the meta-analysis including all comparisons between group ratings and individual (group member) ratings (all these comparisons were between subjects). The analysis was conducted in the statistical software program R, using the metafor package (Viechtbauer, 2010).

The random effects meta-analysis ( $n = 33$ ) produced a mean GA-effect size of Cohen's  $d = 0.60$ , 95% confidence interval (CI) = 0.49 to 0.70. There was thus a significant GA-effect across all between-study comparisons ( $z = 11.45$ ,  $p < .001$ ). In addition, the number of group members in the picture moderated the effect, with larger numbers of group members resulting in larger overall GA-effects,  $\beta = .09$ ,  $t(31) = 7.11$ ,  $p < .001$ . We conducted two moderator analyses to test whether the attractiveness of the group moderates the

effect. Average group-member attractiveness did not moderate the effect,  $\beta = -0.03$ ,  $t(31) = -0.24$ ,  $p = 0.811$ . Average group attractiveness did moderate the effect, with more attractive groups resulting in larger overall GA-effects,  $\beta = 0.26$ ,  $t(31) = 3.19$ ,  $p = 0.003$ . This finding may reflect a true moderation of the GA-effect by average group attractiveness or it may be a statistical artifact: Average group ratings are measured less precisely (one measurement) than average group-member ratings (as many measurements as targets in the photo). Therefore, the average group ratings may have greater sampling error, which could directly influence the size of the GA-effect. If average group ratings are high due to sampling error, they also automatically produce a greater GA-effect. Our data do not allow us to disentangle these two explanations for the significant moderation when using average group attractiveness as a moderator. Two overall conclusions can be drawn from this meta-analysis. First, the GA-effect is medium to large, and second, group size moderates the GA-effect such that larger groups produce a larger GA-effect.

### Comparison With the Cheerleader Effect in Walker and Vul (2014)

Recently, a somewhat different interpretation of the GA-effect was examined. Walker and Vul (2014) investigated the so-called cheerleader effect, which they defined as "people seem more attractive in a group than in isolation" (p. 230).<sup>11</sup> They studied whether females who are presented in a picture with other females are seen as more attractive than females presented alone. Their study employs a within-subject design in which subjects rate each face alone as well as in a group. Individual faces were rated as more attractive in a group than alone. The authors hypothesize that this effect occurs because people mentally morph and average all the faces in the group, and, because people find average faces more attractive (Langlois & Roggman, 1990), the female in the group is judged more attractive. Walker and Vul employed a within-subjects design and found that, on average, individuals portrayed in a group are rated 5.5% standard deviations more attractive than individuals portrayed alone. Interestingly, the control conditions in Studies 1a, 2, and 6 allow us to test the cheerleader effect proposed by Walker and Vul. In contrast to their studies, we employed a between-subjects design in which each participant rates the attractiveness of a number of targets in only one condition. We submitted our data to a mixed repeated-measure ANOVA, with condition as between-subject factor and target person in the photographs as within-subject factor, to test whether we also find that individuals portrayed in a group are on average perceived to be more attractive than individuals portrayed alone. We did not find such differences in any of these studies: Study 1a,  $F(1, 103) = 0.36$ ,  $p = .549$ ,  $\eta^2 = .004$ ; Study 2,  $F(1, 66) = 0.35$ ,  $p = .558$ ,  $\eta^2 = .005$ ; or Study 6,  $F(1, 199) = 0.03$ ,  $p = .855$ ,  $\eta^2 < .001$ . We do, however, find significant variation in

the Walker and Vul cheerleader effect depending on the person who is rated, as indicated by significant interactions between target person and condition in Study 1a,  $F(43, 4429) = 4.34, p < .001, \eta^2 = .04$ , Study 2  $F(9, 594) = 3.29, p < .001, \eta^2 = .05$ , and in Study 6,  $F(7, 1393) = 3.93, p < .001, \eta^2 = .02$ .

The first set of ANOVAs may suggest that the Walker and Vul (2014) cheerleader effect is too small to be detected with our sample sizes and the number of stimuli we use. However, it may also indicate that there is a qualitative difference between within- and between-subjects elicitations of attractiveness judgments. The Condition  $\times$  Target interaction effects that we found suggest that there is considerable variation in the cheerleader effect as proposed by Walker and Vul, which might be due to assimilation and contrast effects (Wedell & Pettibone, 1999) in the conditions in which targets were part of a group. Perhaps, the composition of the group in terms of attractiveness or the contrast between the target and the context dictates whether and when individuals are rated differently alone compared with when they are in a group. This is beyond the scope of our investigation but remains an interesting question to be explored in future research.

## Discussion

We find substantial evidence for the existence of a GA-effect: People perceive groups of people as more attractive than the average attractiveness of its members. Across Studies 1 to 4 we repeatedly find the effect, and it proves quite robust across different stimuli and methods. Our meta-analysis reveals that across all between-subject studies the effect is medium to large in size. Ratings of groups in terms of physical attractiveness might thus be different from group ratings in terms of other social traits, such as friendliness, because they often do not follow the “averaging rule.” The question that remains, though, is why judging physical attractiveness is different from judging other traits? In most other studies we cite in the introduction, visual stimuli are also used, including scales similar to the ones we employed, so the explanation should not be sought in procedural differences. Perhaps the mechanism explaining this effect may give us some clue as to why judgments of groups for physical attractiveness do not follow the averaging rule. In the introduction, we presented two potential accounts that could explain the occurrence of the GA-effect. Across all studies, we tested several of the hypotheses that could be derived from the two accounts.

### Selective Attention

The selective attention account posits that the GA-effect is caused by people selectively attending to the most attractive group members, which causes the less attractive individuals in the group to be taken less into account when the group judgment is made. This account suggests that if we make

people more aware of the variation in the group the GA-effect should disappear or be attenuated. Study 5 provided support for this hypothesis; the GA-effect was attenuated if participants first rated all individuals in the group and then the group as a whole. Please note that Study 9b did not replicate this effect: We found the GA-effect, but this effect did not differ depending on the order that the ratings were made in. However, the procedure and instructions during the eye-tracking study might have interfered with the natural effect. Telling participants that their eyes are being followed can make them more aware of the experimental situation leading to demand effects. Finally, in Study 7, we explicitly instructed participants to pay better attention to the group and again found an attenuated GA-effect.

We also hypothesized that if group members do not differ much in the extent to which they are seen as physically attractive, a smaller GA-effect should occur than when there is great variation in terms of attractiveness. In Study 5, we found that the variation in perceived attractiveness positively predicted the size of the GA-effect. Moreover, controlling for perceived variation eliminated the difference between the group rating and the average of the individual group-member ratings. This suggests that if group members do not differ much in terms of attractiveness a GA-effect is less likely to occur because selective attention will yield almost identical ratings as paying attention to all members in a group; there will be no selective attention.

Studies 8 and 9 revealed that people pay less attention to the less than the more attractive individuals in a group. In Study 8, we found that after making a group rating, participants had problems remembering the less attractive individuals in the group but were very able to remember the more attractive group members. No such differences were found when people evaluated individuals in a group on attractiveness. In Study 9a, this differential attention account for more and less attractive group members was corroborated in an eye-tracking paradigm. Finally, Study 9b revealed that there was a positive, though nonsignificant, relationship between paying attention to the most attractive individuals and the size of the GA-effect for those who experienced the GA-effect. Study 5 also revealed that the more attractive group members' attractiveness ratings were predictive of the group rating, whereas those of the less attractive group members were not. This suggests that the group rating is mainly based on the ratings of the more than the least attractive group members, again indicating that selective attention to the more attractive group members could explain the differences between the group ratings and the average of the individual ratings.

Finally, our meta-analysis indicated that group size was a significant moderator of the GA-effect. We were more likely to find the GA-effect in larger groups (six or more group members) than in smaller groups. Again, this finding is in line with a selective attention account because selective attention should bring about greater deviations from the

average attractiveness in larger groups; if attention is mainly paid to the most attractive group members, larger parts of the group will be left out of the equation when judging the group as a whole. The moderating influence of group size might be one of the reasons why earlier studies supported the averaging rule; they used small groups of approximately three people (e.g., Anderson et al., 1973). In our studies, we do not find the GA-effect for most of the smaller groups of four or five people, which could indicate that the averaging rule is maybe more applicable when judging smaller groups. It thus remains to be seen whether group impressions of traits other than physical attractiveness follow an averaging rule when judged in relatively large groups.

Our findings suggest that the GA-effect is an automatic process that can be overridden when the motivation or ability to process all information is present. This idea is in line with previous findings on attentional capacity and focus. For instance, Maner and colleagues (2003) found that people overestimated the number of attractive individuals in a group only when they were presented with 15 faces simultaneously. When these faces were presented serially, participants could pay enough attention to each face and they did not overestimate the number of attractive individuals. Similarly, another line of research suggests that in group ratings under cognitive load, the presence of individuals with extreme traits is overestimated (Rothbart, Fulero, Jensen, Howard, & Birrell, 1978). It thus seems that people can form accurate impressions when they are given plenty of time to do so. But when people are not able to process all the information presented to them, differences between the group and individual ratings will be found.

In the introduction, we suggested that the selective attention account could produce consistent differences between genders. Gender effects in our studies were inconsistent, and often absent (also see the section on gender effects and Online Appendix III). We are not the first to find inconsistencies in terms of gender effects in attractiveness. Some studies do find differences between gender in terms of attractiveness ratings (e.g., Buss & Schmitt, 1993; Maner et al., 2003), some do not (e.g., Cunningham et al., 1995). Sometimes, it is even found that females look longer at female faces than men (Leder, Tinio, Fuchs, & Bohrn, 2010). In fact, gender effects pertaining to other judgments or behaviors seem to be limited to certain areas of study (Hyde, 2005).

Perhaps, more importantly, gender differences in preferences for attractiveness seem to depend on the context. Gender differences are found when people consider long-term relationships; men care more about attractiveness in a long-term partner than women do. However, in short-term contexts men and women both prefer attractive mates over less attractive mates (Meltzer et al., 2014). Because we did not assess whether our participants had long- or short-term interests, we cannot test whether this was indeed the case in our data, but it could explain why our gender analyses yielded inconsistent results.

## Group Gestalt

From the similarity principle in Gestalt psychology, one would expect that parts that appear more similar are more likely to be perceived as a group, suggesting that in groups where members are seen as more similar the GA-effect should be larger. We tested two types of similarity. First, in Study 5 we tested for homogeneity in terms attractiveness. If people perceive group members as more similar in terms of attractiveness, this should increase the GA-effect. This study actually revealed that not homogeneity, but heterogeneity was related to the size of the GA-effect. As mentioned above, controlling for perceived variance in terms of attractiveness eliminated the effect of rating, suggesting that this effect only occurs in groups where participants perceive enough variation in terms of attractiveness. This finding is thus not in line with the similarity principle and speaks more in favor of the selective attention account.

Second, we operationalized similarity in terms of similarity in dress style. In Study 6, we found that presenting the individuals in the group as more similar in terms of clothing did not increase the group rating, nor did it increase the size of the GA-effect. Concluding, we found no evidence for the similarity principles in Gestalt psychology that could account for the GA-effect in our data.

The present studies only examined two types of similarity in the group: to what extent the group members are perceived as similar in terms of attractiveness and clothing style. Of course other types of similarity could be examined, for instance, similarity in physical appearance (same skin or hair color) or posture (similar poses across group members). Also, future efforts could examine another important principle in Gestalt psychology, the principle of proximity, which suggests that the closer individual elements are in relation to each other, the more likely they are grouped (Wagemans, Elder, et al., 2012). So if the GA-effect is caused by a perception of "groupness" the effect should be greater when group members are portrayed with less interpersonal space between them. Because we used natural stimuli, we could not control and thus test for this factor, but future studies could manipulate interpersonal distance and see whether it increases the GA-effect.

Still, we think there are some Gestalt-like principles at play, because the selective attention account cannot explain why the attractiveness rating of the group is sometimes higher than the attractiveness rating of its most attractive member (see Figure 1, comparison between  $M_{group}$ - $M_{highest}$ ; none of these differences are significant though). If an averaging rule would apply to group ratings of attractiveness, one would expect a similar distribution of individual ratings below and above the group rating. Also, a selective attention account would suggest that the group rating would be an average of the ratings of the most attractive group members, but no higher than the most attractive group member. This finding needs further empirical scrutiny but seems to suggest that the whole is indeed more than the sum of its parts.

### Alternative Explanations

Another explanation for the GA-effect is that people unconsciously morph all the faces in the group into one average “group-face,” and that they judge the attractiveness of the group based on that morph. This mechanism is also used, but never tested, as an explanation for the occurrence of the cheerleader effect (Walker & Vul, 2014). The finding that average faces are seen as more attractive (Langlois & Roggman, 1990) would explain why the attractiveness ratings of the group do not follow the averaging rule. Direct support for the claim that people morph groups of faces (also referred to as ensemble coding) comes from a study in which people viewed sets of emoted morphed faces (of the same face) and subsequently judged the average emotion in the set of faces (Haberma & Whitney, 2009). In that study, people were rather accurate in assessing the mean displayed emotion in sets of 16 faces even at very short exposure times (500 ms or less). However, it is unclear whether this finding generalizes to attractiveness judgments. In addition, it is not clear whether this finding predicts more accurate judgments of attractiveness (conform the averaging rule) or whether the mental morph will be judged to be *more* attractive than the group members it is based on (conform the GA-effect). In the end, it thus remains open to investigation whether processes that imply mental morphing contribute to the GA-effect or not.

The GA-effect seems similar to the so-called sample size bias in judgments (Price, Smith, & Lench, 2006). Research on this bias consistently reveals that people judge the average of a group of targets as riskier, taller, larger, or more prone to experience a certain type of affect when these targets are presented in larger groups than in smaller groups (Price, Kimura, Smith, & Marshall, 2014; Price et al., 2006). These findings are similar to our results revealing that the larger the group of individuals, the larger the GA-effect. The aforementioned researchers explain their findings in terms of a priming effect in which “the sample size activates an internal representation of relative quantity or magnitude that directly affects the internal representations of magnitude and therefore affects the judgment of average” (Price et al., 2014, p. 1329). Moreover, in two studies these researchers ruled out selective attention to extreme targets within the presented group as an explanation. Most of the judgments Price and colleagues use are numeric in nature (more members in the group predict greater estimations of size in terms of millimeters), and thus we are not certain how the number of members in a group could prime high attractiveness ratings. However, the individuals in the group pictures were relatively attractive which, in a sample size bias way, could lead people’s judgment of the group’s attractiveness to be biased upward. Thus, to rule out priming as an explanation, attractiveness of the group members should be systematically varied.

Testing some of the above-mentioned hypotheses requires a fundamentally different approach than the one we took. One of the strengths of our studies is that we use

stimuli that have a very high external validity. The presentation of the groups resembles the way in which people naturally encounter and evaluate groups. Testing these additional hypotheses requires experimentally composed groups of pretested individuals (or portraits) or morphed/composite/computerized faces and groups. This approach may lower the ecological validity of future studies, but it seems worthwhile in further uncovering the psychological mechanism underlying the GA-effect. One could, for instance, manipulate to which group member a participant pays attention, and subsequently see whether drawing attention to the more attractive group members increases the GA-effect.<sup>12</sup> These and other tests of potential mechanisms require future attention.

In the introduction, we stated that differences in individual versus group perceptions may have implications for the formation of stereotypes (e.g., Ford & Stangor, 1992). So far, differences between group and individual impressions have been mainly studied by looking at how people use a set of traits to form impressions when this trait describes either a group as a whole or one individual (Hamilton & Sherman, 1996; Schneider, 2004). We took a different approach in line with research on the averaging rule where we compared the impressions of a group as a whole or of a group consisting of individuals. Our studies reveal that people create different first impressions based on how they encounter a group; their impressions of groups encountered as a whole are more extreme than their impressions of groups when they encounter individuals in that group. The differences in impressions seem to become larger the larger the groups are. Such contextual differences may indicate that stereotypes are formed differently depending on how others are encountered. Our research may not only have implications for how stereotypes are formed but also on how stereotypes could be changed. It is possible that if we make people more aware of the natural variation in social groups, instead of having a natural focus on more extreme group members (i.e., not all immigrants are criminal), we could change stereotypes. Such an approach could help circumvent the issue of subtyping because one does not present people with someone who does not fit the stereotype but presents a whole range of individuals who fit the social category but vary greatly in terms of the stereotypical trait (e.g., Hewstone & Hamberger, 2000).

### Conclusion

Nine studies provide evidence for the existence of a GA-effect: The perceived attractiveness of a group is greater than the average attractiveness of its members. This suggests that group judgments are not always based on the “averaging rule” as suggested by others (e.g., Anderson, 1965). The GA-effect is medium to large and robust across different methods and stimuli. We tested two potential mechanisms and found most support for the selective attention account: The data suggest that attractiveness ratings of the group are higher than an averaging rule would predict

because attention is drawn to the most attractive people in the group.

### Acknowledgments

We thank Pim Houben and field hockey team BHV Push 1, Linda Oosterwijk, Nina Spälti, Kevin van Kalker, Renske van der Linden, Bart van de Pasch, Lotte Hardenbol, Milou van Hal, and Emy Haagh for their help in obtaining stimulus material and collecting data. We thank the social-psychology lab at Tilburg University and Marcel Zeelenberg in particular for their helpful comments and suggestions.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Notes

- In popular media, this effect has also been called the “cheerleader-effect” and it was popularized by Barney Stinson, a character from the CBS show “How I met your mother.” In this show, the effect is introduced as follows: Ted Mosby: “What about those girls over there? They’re hot!” Barney Stinson: “They’re not hot! You’ve just become victims of the cheerleader effect. . . . The cheerleader effect is when a group of women seems hot, but only as a group. Just like with cheerleaders, they seem hot, but take each one individually: Sleddogs!” (How I met your mother, Season 4, Episode 7, CBS; Rashid & Fryman, 2008)
- Relatively unattractive groups were rated as less attractive than the mean attractiveness of the group members.
- All photographs can be obtained from the authors.
- Unlike all other studies, the numbers to identify the individual women in Study 1b in the group-member condition were presented below instead of on the photograph.
- All (12) comparisons between the group conditions (group in casual wear, group hockey apparel, group with numbers, and the group construed from portrait pictures) versus control conditions (members with red square, members with rest of group blurred in black, and the member portrait pictures) were significant (all  $ps < .001$  with Cohen’s  $d$  effect sizes ranging from 0.50 to 0.76).
- We conducted a nonparametric test because our dependent variable was not normally distributed.
- We used Photographs 3 and 5 from Study 1, and the photograph used in Study 6. Participants viewed two other photographs: one that was used by the experimenter as an example to familiarize them with the procedure; the other was the photograph of the male group used in Study 3. Unfortunately, many participants indicated to have recognized the male author of this article, and thus we excluded this photograph from our analyses.
- Due to reasons unknown to us, for a large group of participants attractiveness ratings were not saved by the Tobii software and therefore not analyzed.
- The eye tracker failed to collect sufficient gaze data for 6 of the 120 participants. This could indicate that participants moved away from the computer or moved too much in order for the eye tracker to continue following their eyes. One participant took part twice due to technical errors; these data were neither included in our analyses nor in the number of participants reported.
- The meta-analysis did not include the within-study effect sizes of Studies 5, 7, and 9b because these within-study effect sizes are not directly comparable with between-subjects Cohen’s  $d$  effect sizes.
- Please note that this definition of the cheerleader effect is not fully congruent with the original definition as cited in Footnote 1.
- We thank an anonymous reviewer for this suggestion.

### Supplemental Material

The online supplemental material is available at <http://pspb.sagepub.com/supplemental>.

### References

- Anderson, N. H. (1965). Averaging versus adding as a stimulus combination rule in impression formation. *Journal of Experimental Psychology*, *70*, 394-400.
- Anderson, N. H., Lindner, R., & Lopes, L. L. (1973). Integration theory applied to judgments of group attractiveness. *Journal of Personality and Social Psychology*, *26*, 400-408.
- Berry, D. S. (2000). Attractiveness, attraction, and sexual selection: Evolutionary perspectives on the form and function of physical attractiveness. *Advances in Experimental Social Psychology*, *32*, 273-342.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384-392. doi:10.1177/0956797610397956
- Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, *100*, 204-232.
- Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). “Their ideas of beauty are, on the whole, the same as ours”: Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, *68*, 261-279.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt.
- 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*, 878-902.
- Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology*, *63*, 356-367.
- Gangestad, S. W. (1993). Sexual selection and physical attractiveness. Implications for mating dynamics. *Human Nature*, *4*, 205-235.

- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology*, *35*, 718-734. doi:10.1037/a0013899
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336-355.
- Hewstone, M., & Hamberger, J. (2000). Perceived variability and stereotype change. *Journal of Experimental Social Psychology*, *36*, 103-124.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581-592.
- Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general ability, physical attractiveness, core self-evaluations, and income. *Journal of Applied Psychology*, *94*, 742-755.
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluations as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, *29*, 299-304.
- 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*, 390-423.
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*, 115-121.
- Leder, H., Tinio, P. L., Fuchs, I. M., & Bohrn, I. (2010). When attractiveness demands longer looks: The effects of situation and gender. *The Quarterly Journal of Experimental Psychology*, *63*, 1858-1871.
- Levy, L. H., & Richter, M. L. (1963). Impressions of groups as a function of the stimulus values of their individual members. *Journal of Abnormal and Social Psychology*, *67*, 349-354.
- Maner, J. K., Gaijlot, M. T., Rouby, D. A., & Miller, S. L. (2007). Can't take my eyes off you: Attentional adhesion to mates and rivals. *Journal of Personality and Social Psychology*, *93*, 389-401.
- Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., & Neuberg, S. L. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology*, *85*, 1107-1120.
- Martin, D. R., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (Vol. 2, pp. 416-423). IEEE.
- Meltzer, A. L., McNulty, J. K., Jackson, G. L., & Karney, B. R. (2014). Men still value physical attractiveness more than women: Rejoinder to Eastwick, Neff, Finkel, Luchies, and Gunt (2014). *Journal of Personality and Social Psychology*, *106*, 435-440.
- Miller, C. T., & Felicio, D. M. (1990). Person-positivity bias: Are individuals liked better than groups? *Journal of Experimental and Social Psychology*, *26*, 408-420.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353-383.
- Pinna, B. (2012). Perceptual organization of shape, color, shade and lighting in visual and pictorial objects. *i-Perception*, *3*, 257-281.
- Price, P. C., Kimura, N. M., Smith, A. R., & Marshall, L. D. (2014). Sample size bias in judgments of perceptual averages. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1321-1331.
- Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality and Social Psychology*, *90*, 382-398.
- Rashid, R., (Writer), & Fryman, P. (Director). (2008). Not a father's day [Television series episode]. In C. Bays & C. Thomas (Creators), *How I met your mother*. New York, NY: CBS.
- Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology*, *14*, 237-255.
- Schneider, D. J. (2004). *The psychology of stereotyping*. New York, NY: Guilford Press.
- Simmons, J., Nelson, L., & Simonsohn, U. (2012, Fall). *A 21 word solution*. Science Dialogue. Retrieved from [http://www.spsp.org/resource/resmgr/dialogue/dialogue\\_26\(2\).pdf](http://www.spsp.org/resource/resmgr/dialogue/dialogue_26(2).pdf)
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software*, *36*, 1-48.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von der Heydt, R. (2012). A century of Gestalt psychology in visual perception I: Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172-1217. doi:10.1037/a0029333
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., Van der Helm, P., & Van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception II. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*, 1218-1252. doi:10.1037/a0029334
- Walker, D., & Vul, E. (2014). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science*, *25*, 230-235. doi:10.1177/0956797613497969
- Watkins, L. M., & Johnston, L. (2000). Screening job applicants: The impact of physical attractiveness and application quality. *International Journal of Selection and Assessment*, *8*, 76-84.
- Wedell, D. H., & Pettibone, J. C. (1999). Preference and the contextual basis of ideals in judgment and choice. *Journal of Experimental Psychology*, *128*, 346-361.
- Willis, R. H. (1960). Stimulus pooling and social perception. *Journal of Abnormal and Social Psychology*, *60*, 365-373.