von Helversen, Bettina ; Herzog, Stefan, M. ; Rieskamp, Jörg

Haunted by a doppelgänger: Irrelevant facial similarity affects rule-based judgments

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3465

Publication date of this document: 15/11/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

von Helversen, B., Herzog, S. M., & Rieskamp, J. (2014). Haunted by a Doppelgänger: Irrelevant Facial Similarity Affects Rule-Based Judgments. Experimental Psychology, 61(1), 12–22. https://doi.org/10.1027/1618-3169/a000221

Haunted by a Doppelgänger

Irrelevant Facial Similarity Affects Rule-Based Judgments

Bettina von Helversen,[1] Stefan M. Herzog,[2] and Jörg Rieskamp[1]

[1]Department of Psychology, University of Basel, Switzerland, [2]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Abstract. Judging other people is a common and important task. Every day professionals make decisions that affect the lives of other people when they diagnose medical conditions, grant parole, or hire new employees. To prevent discrimination, professional standards require that decision makers render accurate and unbiased judgments solely based on relevant information. Facial similarity to previously encountered persons can be a potential source of bias. Psychological research suggests that people only rely on similarity-based judgment strategies if the provided information does not allow them to make accurate rule-based judgments. Our study shows, however, that facial similarity to previously encountered persons influences judgment even in situations in which relevant information is available for making accurate rule-based judgments and where similarity is irrelevant for the task and relying on similarity is detrimental. In two experiments in an employment context we show that applicants who looked similar to high-performing former employees were judged as more suitable than applicants who looked similar to low-performing former employees. This similarity effect was found despite the fact that the participants used the relevant résumé information about the applicants by following a rule-based judgment strategy. These findings suggest that similarity-based and rule-based processes simultaneously underlie human judgment.

Keywords: decision making, similarity, judgment, computational modeling

Imagine a personnel selection situation where the owner of a small business is looking for a new employee. The position is advertised and a large number of applications need to be evaluated. Prescriptive research on personnel selection advocates evaluating each applicant on a number of relevant criteria – such as work experience, academic performance, conscientiousness, or communication skills, to name but a few – weighting the criteria according to their importance, and then selecting the highest scoring applicant (e.g., Figueira, Greco, & Ehrgott, 2005; Hough & Oswald, 2000). This personnel selection policy indeed predicts actual employment decisions of experienced personnel managers (e.g., Graves & Karren, 1992; Moy & Lam, 2004). There is, however, ample evidence that human judgments and decisions can deviate from optimal policies because decision makers rely on irrelevant information. For instance, employment decisions are influenced by applicants' ethnic background (Stewart & Perlow, 2001), gender (Davison & Burke, 2000), or (un) attractiveness (Agthe, Spörrle, & Maner, 2010; O'Brien et al., 2008), even though these attributes generally are neither predictive of work performance nor accepted as valid grounds for making employment decisions. One explanation for these influences is that people hold stereotypical beliefs and attitudes about specific groups, which – explicitly or implicitly – color their judgment about members of that group (Davison & Burke, 2000; O'Brien et al., 2008; Stewart & Perlow, 2001).

In this article we suggest a further mechanism by which judgments can deviate from the optimal policy. Suppose that in the introductory example above, the first applicant invited for an interview looks just like the former, lazy employee that the business manager recently fired. Chances are that the business manager is reluctant to hire this person, even if she is convinced by the résumé. In this vein, we suggest that judgments about people can be influenced by their similarity to people whom the decision maker has previously encountered, even when this similarity is irrelevant for the task, other relevant information is available that can be used for rule-based judgment, and the decision maker has no stereotypical beliefs or attitudes about the person in question.

Cognitive Processes in Judgment

Evaluating a job applicant is an instance of a more general class of judgment problems called *multiple-cue judgments*. In such tasks the decision maker integrates information from several variables (i.e., cues) into a single judgment (for an overview see Doherty & Kurz, 1996). Typical multiple-cue judgment tasks involve evaluating other people, such as judging the credibility of witnesses, the mental health of patients, or the risks in granting parole to prisoners. Research suggests that people often rely on a linear judgment policy when making multiple-cue judgments and that this process can be described by a linear model estimated with linear regression methods (Cooksey, 1996). According to this linear model, people weight relevant information according to its importance and then form a judgment by integrating the weighted information in an additive manner. The linear model has been successfully used to describe how people render judgments in many areas, ranging from medical diagnoses (Wigton, 1996) to bail decisions (Ebbesen & Konečni, 1975).

Although linear models are well suited to describe judgment processes, they cannot explain why human judgments are often influenced by irrelevant information, such as appearance (e.g., O'Brien et al., 2008). We argue that the similarity of targets (e.g., job applicants, patients) to people whom the decision maker has previously encountered can systematically affect judgments even when the decision maker is following a linear judgment policy and similarity does not carry any information relevant for the judgment. Similarity is considered a key factor underlying cognitive processing in many areas of cognition (Hahn & Chater, 1998). Similarity-based theories, such as exemplar models, are prominent in categorization (e.g., Nosofsky, 1986) and have recently been extended to judgment and decision making (Juslin, Karlsson, & Olsson, 2008; Nosofsky & Bergert, 2007; Persson & Rieskamp, 2009; von Helversen, Mata, & Olsson, 2010; von Helversen & Rieskamp, 2008, 2009). Furthermore, people rely on similarity when evaluating others. Perceived similarity of another person to the self has been linked to increased attraction, sympathy, and liking (e.g., Berscheid & Walster, 1978; Byrne, Clore, & Worchel, 1966; Loewenstein & Small, 2007). In particular, facial similarity has been shown to influence evaluations. For instance, Bailenson, Iyengar, Yee, and Collins (2008) showed that voters preferred photographs of political candidates that resembled themselves. Also, facial similarity to people other than the self can influence evaluations (e.g., Lewicki, 1985; Zebrowitz & Montepare, 2008). Whereas unknown faces are more positively evaluated if they resemble significant-others (Günaydin, Zayas, Selcuk, & Hazan, 2012; Kraus & Chen, 2010) or persons about whom the decision maker has formed a positive opinion, negative experiences with a similar person can decrease liking (Gawronski & Quinn, 2013; Verosky & Todorov, 2010).

These effects of facial similarity on judgments were shown in situations when no further information about the person under evaluations was available. If information is available, however, similarity effects may disappear. Indeed, in the judgment and decision-making literature exemplar effects have generally been reported when it was impossible or very difficult to apply a rule-based process, but not when information was readily available that could be used for making an accurate rule-based judgment (Nosofsky & Bergert, 2007; Persson & Rieskamp, 2009; Platzer & Bröder, 2012; von Helversen & Rieskamp, 2009). This raises the question as to whether similarity – and, in particular, facial similarity – will still influence judgments, even if it is irrelevant for judgments and other relevant information is available that allows accurate rule-based judgments.

If people's judgments are influenced by facial similarity when other information is available, this could also affect how this other information is processed. Research in judgment and categorization often assumes that people use either similarity- or rule-based strategies (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Juslin et al., 2008; Karlsson, Juslin, and Olsson, 2007). This suggests that when people consider similarity they completely switch to an exemplar-based judgment process. According to this hypothesis they not only consider facial similarity, but rely on a similarity-based strategy to process all information available about the person. For instance, Juslin, Karlsson, and Olsson (2008) found that people used an exemplar-based judgment process that relied on a similarity measure based on all cues when a linear model was not suited for the task (see also, Karlsson et al., 2007). Alternatively, however, people could continue processing the information available according to a linear policy, but additionally integrate facial similarity into their judgment. This idea is supported by research in categorization suggesting that similarity and rules can simultaneously influence categorizations (Brooks & Hannah, 2006; Brooks, Norman, & Allen, 1991; Erickson & Kruschke, 1998; Hahn, Prat-Sala, Pothos, & Brumby, 2010; Hannah & Brooks, 2009; Thibaut & Gelaes, 2006). For instance, Hahn and colleagues showed that people were faster and made fewer errors classifying new items that were similar to training items (as compared to items that where dissimilar) – even if the categorization rule was equally applicable to all new items. Moreover, computational models, such as "attention to rules and instances in a unified model (ATRIUM)", have successfully captured categorization behavior by combining similarity- and rule-based processes (Erickson & Kruschke, 1998).

The goal of the present research is to investigate (1) whether similarity can influence judgments, even when people have relevant information that they can use to make rule-based judgments and similarity information is irrelevant, and (2) if people were to use similarity, whether they would completely switch to an exemplar strategy or whether they would follow a rule-based strategy influenced by similarity.

The Present Study

In two studies we tested the hypothesis that the evaluation of a person will be influenced by facial similarity to previously encountered people even if relevant information is available. The participants' task was to evaluate applicants for a job. To this end, they received information about the applicants' résumé that they could use to form a rule-based judgment. In addition, we manipulated similarity by morphing the faces of the applicants so that they resembled either high-performing former employees, low-performing former employees, or none of the former employees.

In addition, we varied the difficulty of the judgment problem across studies. In Study 1a we used a task that could be well solved using the provided cues. Thus using

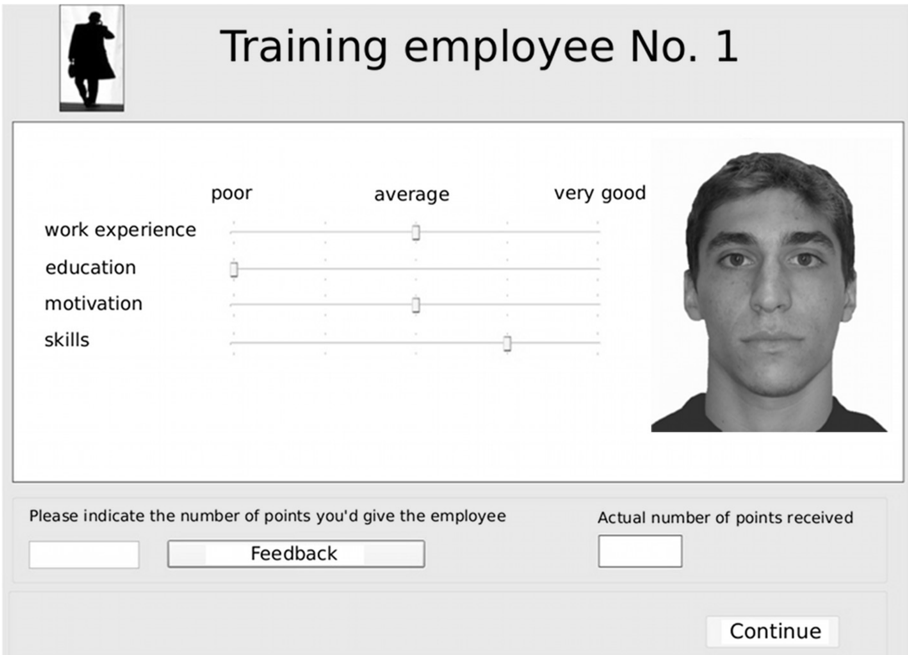

Figure 1. Illustrative example of a trial in the training phase. In each trial participants saw a training employee and gave a suitability judgment; then they received feedback. Training continued until participants could accurately evaluate all six training employees. Image from Minear and Park (2004).

similarity would not be necessary to reach good performance. To gauge the influence of similarity on real-world judgment, we used a less predictable task that is more representative of real-world judgment problems in Study 1b. Research in categorization shows that similarity influences responses, especially when categorization rules are imprecise (Brooks et al., 1991; Brooks & Hannah, 2006). If a task is less predictable, a linear judgment policy will lead to lower performance. This may, in turn, cause people to search for additional information, such as similarity, to guide their judgments and then to rely more heavily on it.

Method

Participants

In Study 1a, 30 students participated; 80% were female; mean age was 26.5 years (SD = 8.8). In Study 1b, 30 students participated; 57% were female; mean age was 23.7 years (SD = 4.9). The majority of participants were students from the University of Basel who took part for a book voucher (worth 10 Swiss Francs [CHF]) or course credit. In addition, they could earn a bonus depending on their performance in the judgment task (Study 1a: M = 7.0 CHF; Study 1b: M = 4.3 CHF). In Study 1b two participants were excluded from the analysis, because they did not reach the learning criterion.

Design and Procedure

Both studies consisted of two parts: a training and a test phase. In the training phase participants had to learn to correctly judge the suitability of six training employees on a scale from 0 (*not suitable at all*) to 100 (*very suitable*). For each training employee they saw a summarized version of the training employee's application documents consisting of a picture and résumé information on four cues.[1] The cues were quality of work experience, motivation, skills, and education. The order of the cues on the screen was randomized between participants. On each cue, training employees had a value between 0 (*poor*) and 4 (*very good*). Participants were informed that they could use the cues to help them make a judgment and that they would be paid according to their judgment accuracy in a test phase. In each training trial participants saw a training employee's application and then had to give a judgment (see Figure 1). After providing their judgment, participants received feedback about the true suitability of the training employee. Two of the six training employees performed at a high level, two at a moderate level, and two performed poorly. During training participants repeatedly judged the training employees in blocks consisting of all six training employees. Presentation order was randomized within a block. Training continued until a participant reached a learning criterion (the average deviation between the participants' judgments and the suitability of the training employees was two or less in the last training

[1] Job applications in Switzerland typically include a picture of the applicant.

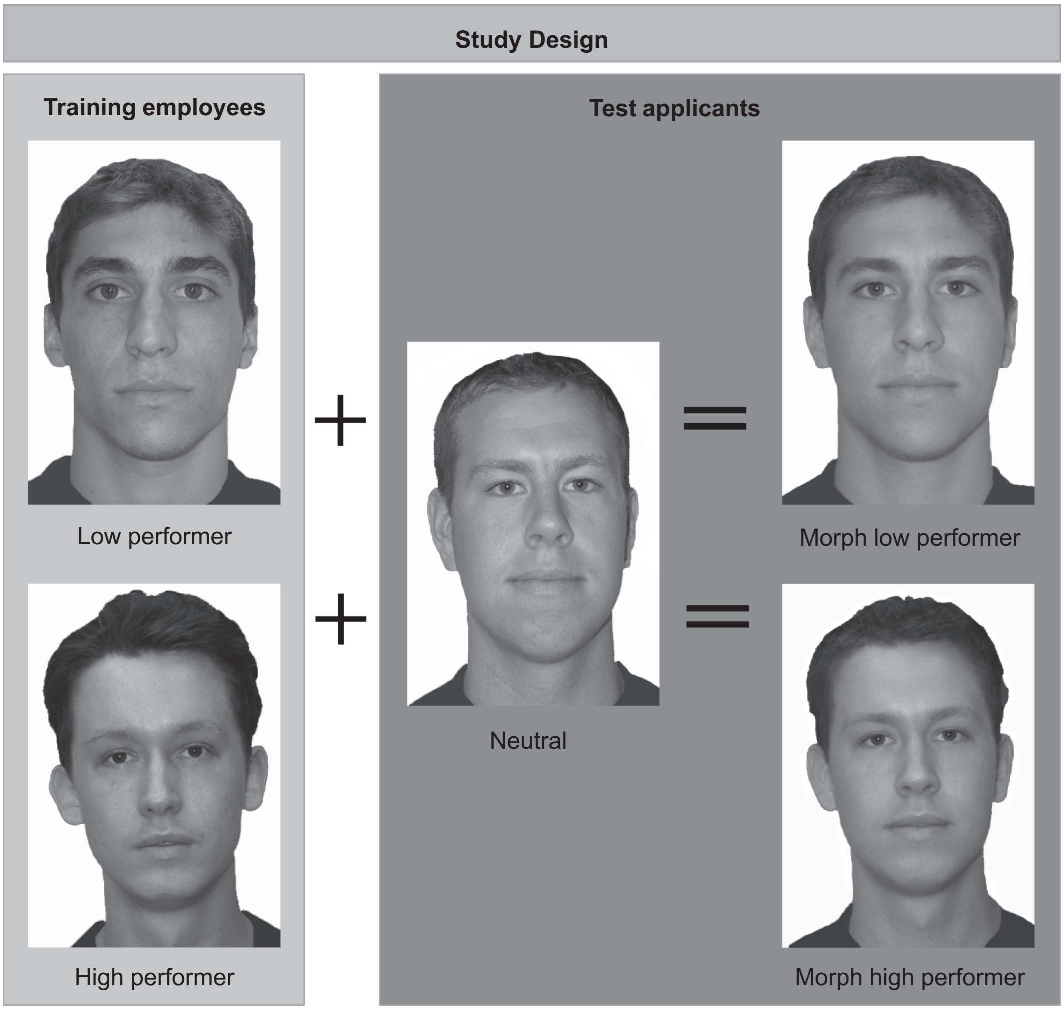

Figure 2. Illustrative example of the design in Studies 1a and 1b. In a training phase participants saw pictures and résumé information of high- and low-performing training employees (e.g., the two left pictures). In the subsequent test phase participants evaluated four sets of three test applicants who had the same résumé information but looked different: (1) test applicants who did not resemble any of the training employees (e.g., neutral test applicant, middle picture), (2) test applicants who looked similar to a low-performing training employee (e.g., upper-right picture), and (3) test applicants who looked similar to a high-performing training employee (e.g., lower-right picture). Similarity was manipulated by morphing the picture of the neutral test applicant with that of a training employee. Images from Minear and Park (2004).

block and one of the training blocks before) or had completed a maximum of 120 training trials (i.e., 20 blocks).

In the subsequent test phase participants made judgments about a set of new job applicants (i.e., test applicants). In this test phase we manipulated the facial similarity of test applicants to resemble that of the training employees. We asked participants to judge four *neutral test applicants* whose cue profiles indicated average suitability and who did not facially resemble any of the training employees. For each of the four neutral test applicants we created two variants with the same respective cue profiles but different pictures: one where the applicant looked similar to an unsuccessful training employee (i.e., *low morph test applicants*) and one where he looked similar to a successful training employee (i.e., *high morph test applicants*; Figure 2). We manipulated facial similarity by morphing images of a neutral test applicant with that of either a successful or an unsuccessful training employee. If facial similarity influences judgments, the four test applicants who looked similar to low-performing training employees should be judged as less suitable than the respective four test applicants who did not resemble any training employee or the four test applicants who looked similar to a high-performing training employee.

We also included 12 *model-fitting test applicants*, to computationally model participants' judgments and to investigate whether participants relied on a linear judgment or a similarity-based strategy to process the cues. These applicants all had unique cue profiles and neutral pictures that did not resemble any of the training employees or other test applicants. We also included the six training employees from the training phase.

The 30 test applicants were repeated three times in random order. In each test trial, participants received points depending on how close their judgment came to the criterion values, but they only received feedback about performance at the end of the test phase. If they estimated exactly the correct value, they received 100 points; the more they deviated from the correct criterion value, the fewer points they received.[2] Points were exchanged for Swiss francs (100 points = 0.10 CHF) after the experiment.

In Study 1b we additionally controlled for picture-specific effects by switching the photographs of the low-performing and high-performing training employees for half of the participants (i.e., counterbalancing which pictures were assigned to low- and high-performing training employees, respectively). This resulted in a two (assignment condition) by three (morph set 1, neutral, morph set 2) mixed design. Morphed test applicants who looked similar to high-performing training employees in the first assignment condition looked similar to low-performing training employees in the second assignment condition and vice versa. Furthermore, we added three questions to assess whether participants thought that they relied on appearance to guide their judgments. We asked participants to indicate how much they had relied on (1) the résumé information and (2) the photograph and (3) how much their judgment was influenced by facial appearance (on Likert scales ranging from 1 *not at all* to 5 *very much*). Because the latter two questions were highly correlated ($r = 0.83$), we merged them into a single "appearance" scale.

Materials

For each study we generated a judgment environment by first creating 1,000 persons described on four cues. The cues' values were drawn from uniform distributions (ranging from 0 to 4). We then calculated the criterion value for each applicant (i.e., suitability) as a linear function of the cues (i.e., the résumé information) with each cue receiving a weight of 6; the intercept was zero. We introduced random error by adding an error term to the criterion value of each person, drawn from a zero centered normal distribution. By using a narrow error distribution ($\sigma = 5$) in Study 1a, the criterion was highly predictable from the cues ($R^2 = 0.96$). In contrast, by using a wider error distribution ($\sigma = 20$) in Study 1b, the criterion was less predictable from the cues ($R^2 = 0.60$). For each study, we then selected training and test items using two constraints. First, a linear model (Cooksey, 1996; von Helversen et al., 2010) must explain a similar amount of variance of the training employees' suitability as of the suitability of all 1,000 persons. Second, a linear model and an exemplar model (von Helversen et al., 2010) must make different predictions for the model-fitting test applicants. The task structures can be found in the Appendix, Tables A1 and A2. We used images from the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips, Wechsler, Huang, & Rauss, 1998) and the Lifespan Database of Adult Facial Stimuli (Minear & Park, 2004) using the neutralized faces by Ebner (2008). We morphed the facial images using the Files Guard software Free Morphing 2.1, which produces a sequence of frames that show the transformation of the source image into the target image. We used the middle frames and post-edited them in Adobe Photoshop CS5 (Adobe Systems) to obtain natural-looking facial images. In a pretest participants rated the similarity of the morphed images to the two source images. We selected four sets of three pictures, so that the two ratings of similarity between an unprocessed picture (i.e., neutral test applicant) and two morphed images were comparable (see also Figure 2).

Results Study 1a

Performance

Participants quickly learned to evaluate the training employees: On average they reached the training criterion in 43 trials (range: 24–84). They also performed well during the test phase: participants' judgments explained, on average, 79% ($SD = 10$) of the criterion's variance. This suggests that participants used the cues because using facial similarity would not allow making accurate judgments at test.

Similarity Effects

In a first step we investigated whether facial similarity influenced the judgments in addition to the cues. To this end, we averaged – separately for the three types of test candidates (i.e., low morph, neutral, or high morph) – the four respective judgments of the test applicants and then conducted a repeated measurement ANOVA with type of test candidate as a within factor. Facial similarity systematically affected the judgments (see Figure 3): Test applicants who resembled *low*-performing training employees were judged as *less* qualified than neutral test applicants, $F(1, 29) = 4.77$, $p = .04$, partial $\eta^2 = .14$, whereas test applicants who resembled *high*-performing training employees were judged as *more* qualified than neutral test applicants, $F(1, 29) = 10.29$, $p = .003$, partial $\eta^2 = .26$ (means and *SD*s are reported in Table 1).

Cognitive Processes

To investigate participants' judgment strategies, we computationally modeled participants' judgments of the

[2] The number of points q was determined based on the quadratic loss function, $q = -x^2/4.45 + 100$, with x being the deviation of the judgment to the correct criterion value; x was truncated to a value of 25 for any x larger than 25. For the rationale of this feedback algorithm see von Helversen and Rieskamp (2008).

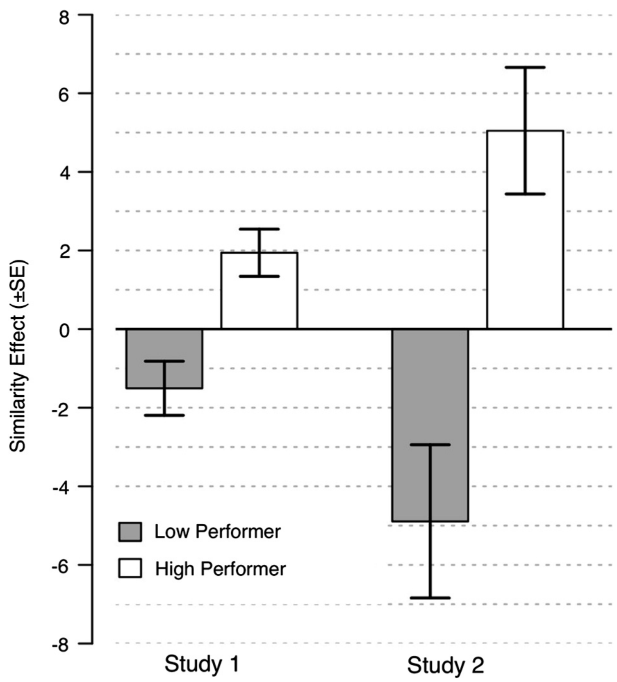

Figure 3. Mean effect of similarity induced by test applicants with pictures similar to low-performing training employees (low performer) and high-performing training employees (high performer), respectively. Study 1a included 30 participants; Study 1b included 28 participants. The similarity effect was operationalized as the difference in suitability judgments for low performers and high performers compared to neutral test applicants. Judgments were averaged across the four sets of test applicants.

model-fitting test applicants. To this end, we estimated (a) a linear model using multiple linear regression (Cooksey, 1996; Juslin et al., 2008) that predicted participants' judgments based on the four cues, (b) a *cue-based* exemplar model (i.e., an exemplar-model that predicts participants' judgments solely based on the cues and does not incorporate facial similarity; Juslin et al., 2008; von Helversen et al., 2010), and (c) a baseline model that always predicted a participant's mean judgment (i.e., ignored the cues). We estimated the best-fitting model parameters minimizing the squared deviation between model predictions and participants' judgments.[3] In a second step we used the estimated model parameters to predict participants' judgments for the neutral and the morphed test applicants (i.e., cross-validation).

We included a cue-based exemplar model to investigate whether participants completely switched to an exemplar-based judgment process (i.e., whether they also adopted a similarity-based strategy to process the cue information), or whether they still integrated the cues according to a linear policy. If the former were to hold, the exemplar model should be able to predict participants' judgments and should outperform the linear model. However, if the latter were to hold, the linear model should outperform the exemplar model.

The linear model fitted participants' judgments of the model-fitting applicants very well (R^2 = 0.93, *SD* = 0.05; RMSD = 7.2, *SD* = 2.6) and better than the two alternative models (Exemplar model: R^2 = 0.85, *SD* = 0.07, RMSD = 11.2, *SD* = 2.4; Baseline model: RMSD = 28.1, *SD* = 2.1; all *p*s < .001). More importantly, the linear model also best predicted participants' judgments of the morphed and neutral applicants (i.e., cross-validation; all *p*s < .05, see Table 1).

Results Study 1b

Performance

Also in Study 1b participants quickly learned to evaluate the training employees: On average they reached the training criterion in 47 trials (range 24–108); 2 of the 30 participants did not reach the learning criterion (and were excluded from further analyses). Although the task was much less predictable than in Study 1a, participants still performed comparatively well during the test phase: Participants' judgments explained, on average, 32% (*SD* = 18) of the criterion's variance. This suggests that participants still used the cues, even though they were less predictive than those in Study 1a.

Similarity Effects

Judgments were strongly influenced by similarity (see Figure 3). We again first averaged the respective judgments across the four sets of test applicants and then conducted a mixed analysis of variance (ANOVA) on the judgments with the within-factor image (morph set 1, neutral, or morph set 2) and the between-factor assignment condition. We again found a similarity effect, that is, an interaction between assignment condition and image (Greenhouse-Geisser corrected: $F(1.27, 32.96) = 10.37$, $p = .002$, partial $\eta^2 = .29$), but no main effect of image (Greenhouse-Geisser corrected: $F(1.27, 32.96) = 1.31$, $p = .27$, partial $\eta^2 = .05$) or assignment condition, $F(1, 26) = 0.12$, $p = .73$, partial $\eta^2 = .01$. Follow-up tests showed that morphed test applicants were rated as less qualified than neutral applicants when they resembled low-performing training employees, $t(27) = 3.04$, critical *t*-value (Tukey corrected)$_{\alpha = .05}$ = 2.48. In contrast, the same morphed test applicants were rated as more qualified than neutral applicants when they resembled the high-performing training employees, $t(27) = 2.59$, critical *t*-value (Tukey

[3] Modeling details can be downloaded from https://www.researchgate.net/publication/235735770_Supplementary_Material_HelversenHerzogRieskamp_2013.

Table 1. Mean judgments and model fit (RMSD) for the morphed and neutral test candidates in Study 1a

	Model-fitting test applicants	Low-performance morphs	Neutral	High-performance morphs
Judgment	–	47.9 (6.8)	49.4 (7.5)	51.4 (7.2)
Linear model fit	7.2 (2.6)	11.2 (4.9)	10.0 (3.9)	11.9 (5.0)
Exemplar model fit	11.2 (2.4)	21.9 (5.7)	21.3 (6.1)	21.9 (6.2)
Baseline model fit	28.1 (2.1)	13.1 (5.4)	12.0 (5.0)	13.4 (5.4)

Note. $N = 30$. Numbers in parentheses indicate standard deviations.

corrected)$_{\alpha = .05}$ = 2.48. Means and *SD*s are reported in Table 2.

Cognitive Processes

As in Study 1a, we estimated the best-fitting parameters for a linear model, a cue-based exemplar model, and a baseline model by fitting them to participants' judgments of the model-fitting applicants and then predicted judgments for the neutral and morphed applicants based on the estimated model parameters. As in Study 1a, the linear model fitted participants' judgments of the model-fitting test applicants very well ($R^2 = 0.81$, $SD = 0.17$; RMSD = 11.2, $SD = 3.7$) and better than the two alternative models (Exemplar model: $R^2 = 0.67$, $SD = 0.13$; RMSD = 17.1, $SD = 2.5$; Baseline model: RMSD = 28.1, $SD = 3.0$; *p*s < .001). More importantly, again, the linear model also best predicted participants' judgments of the neutral and morphed applicants (i.e., cross-validation; all *p*s < .001, see Table 2 and Footnote 3).

Self-Reports

Participants reported that they relied somewhat more on the résumé information ($M = 3.5$, $SD = 1.4$) than on appearance ($M = 2.6$, $SD = 1.4$), $t(27) = 1.96$, $p = .06$, $d = 0.37$. The more participants reported relying on résumé information, the lower was the size of their similarity effect, $r(28) = -.46$, 95%-CI [–.71, –.11], $p = .02$; in contrast, the more participants reported relying on appearance, the larger was the size of their similarity effect, $r(28) = 0.63$, 95%-CI [0.34, 0.81], $p = .001$.

Discussion

Judging people is a common and important task. Although people's judgments are generally reasonable, human judgment can be influenced by irrelevant context factors (e.g., Agthe et al., 2010; O'Brien et al., 2008; Stewart & Perlow, 2001). In two studies we found that facial similarity systematically affected judgments of job suitability in an employment context even though participants used relevant résumé information and using facial similarity led to worse judgments than ignoring it. Job applicants were judged as more qualified when they resembled high-performing employees previously encountered in the study. In contrast, job applicants were judged as less qualified when they resembled low-performing employees.

Our results demonstrate that facial similarity to previously encountered people can influence judgments even if facial similarity is irrelevant and other, relevant information is available. These results correspond with effects of facial resemblance on evaluations (e.g., Kraus & Chen, 2010; Verosky & Todorov, 2010) and show that they also apply when relevant information is available. The effect of facial similarity was evident in both studies, even though in Study 1a the judgment task could be solved almost perfectly based on a linear combination of the résumé information. This could suggest that the effect of facial similarity on evaluations is automatic and unconscious. Consistent with this conjecture, there is evidence that facial resemblance influences automatic processes (Gawronski & Quinn, 2013) and it has been argued that the effect of similarity is outside of deliberate control (Hahn et al., 2010). In Study 1b, we found, however, that the influence of similarity was larger the *more* people reported using appearance and the *less* they reported using the résumé information. Thus, even

Table 2. Mean judgments in both assignment conditions and model fit (RMSD) for the morphed and neutral test candidates in Study 1b

Measures	Model-fitting test applicants	Set 1	Neutral	Set 2
Judgments (assignment 1)	–	47.6 (10.7)	54.1 (7.4)	61.4 (14.5)
Judgments (assignment 2)	–	56.3 (14.0)	53.2 (11.8)	49.7 (10.8)
Linear model fit	11.2 (3.7)	17.6 (8.6)	14.3 (5.4)	15.6 (6.5)
Exemplar model fit	17.1 (2.5)	26.0 (10.8)	24.0 (8.7)	25.9 (10.3)
Baseline model fit	28.1 (3.0)	21.5 (7.3)	18.5 (5.8)	19.3 (6.2)

Notes. $N = 28$. Numbers in parentheses indicate standard deviations. In assignment condition 1, the morphed faces in set 1 resembled low-performing training employees and the morphed faces in set 2 resembled high-performing employees; vice versa for assignment condition 2.

though it seems plausible that facial similarity exerts an automatic influence on evaluations, people may also deliberately choose to use facial similarity as a basis for their judgments – for instance, when other information does not allow reliable judgments.

Our results also bear on the cognitive processes underlying human judgment. Even though participants in our studies were influenced by similarity, computational modeling of participants' judgments indicated that participants primarily used a linear combination of the résumé information for their judgments that was additionally influenced by similarity. This suggests that participants did not completely switch to an exemplar strategy. Instead, they simultaneously relied on similarity- and rule-based processes to render their judgments – resonating with findings in the categorization literature (Brooks & Hannah, 2006; Hahn et al., 2010). This result corresponds with the idea that people rely on both processes (Ashby et al., 1998) and supports a growing literature on hybrid models assuming a blending of rule-based and similarity-based processes (e.g., Erickson & Kruschke, 1998). Furthermore, our results suggest that similarity can still influence judgments, even if overall a rule-based strategy describes people's judgments or decisions better.

As facial similarity to previously encountered people influences a decision maker's judgments, our results imply that personal contact can impair objective decision making. Whenever a person is to be evaluated purely based on facts – as in legal, medical, or employment decisions – such an influence of facial similarity is arguably problematic. To alleviate this problem, such objective judgment tasks should be structured so that only relevant facts can enter the evaluation process. For instance, application documents should contain only relevant résumé information and not an applicant's picture – standard procedure in hiring decisions in, for example, the United States and the United Kingdom, but not in, for example, Switzerland or Germany.

Although relying on facial similarity may be inappropriate in many contexts, similarity is not always an undesirable influence: Incorporating similarity into judgments can improve accuracy if similarity adds valid and nonredundant information. For example, when judging the type of a tumor, physicians are well advised to use the similarity of the tumor to pictures of typical tumors. Furthermore, research in machine learning, forecasting, and psychology suggests that blending different methods or processes based on different representations is often advantageous (Herzog & Hertwig, 2009; Herzog & von Helversen, in press; Kuncheva, 2004).

In sum, our studies show that evaluative judgments about people (such as job applicants, patients, students, or clients) can be systematically influenced by their facial similarity to other people the decision maker has previously encountered even if similarity is irrelevant and using it can reduce judgment accuracy. Although we tested the influence of facial similarity, we speculate that many other aspects can influence judgments (e.g., clothing, accent and vocabulary, movement patterns, or the sound of the voice). The effect of similarity is particularly strong in difficult judgment problems akin to the problems we encounter in our daily lives. How people judge us depends not only on what we do in our lives, but also on what those people do that we resemble to. We are haunted by our doppelgängers.

Acknowledgments

This research was supported by a grant from the Swiss National Science Foundation (100014_130192/1). We would like to thank Loris Criscione for his help in preparing the morphed picture and Elias Niklaus for his support in collecting the data.

References

Agthe, M., Spörrle, M., & Maner, J. K. (2010). Don't hate me because I'm beautiful: Anti-attractiveness bias in organizational evaluation and decision making. *Journal of Experimental Social Psychology*, 1151–1154. doi: 10.1016/j.jesp.2010.05.007

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481. doi: 10.1037/0033-295X.105.3.442

Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. A. (2008). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly, 72*, 935–961. doi: 10.1093/poq/nfn064

Berscheid, E., & Walster, E. H. (1978). *Interpersonal attraction* (2nd ed.). Reading, MA: Addison-Wesley.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*, 278–287. doi: 10.1037/0096-3445.120.3.278

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules". *Journal of Experimental Psychology. General, 135*, 133–151. doi: 10.1037/0096-3445.135.2.133

Byrne, D., Clore, G. L., & Worchel, P. (1966). Effect of economic similarity-dissimilarity on interpersonal attraction. *Journal of Personality and Social Psychology, 4*, 254–258.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.

Davison, H., & Burke, M. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225–248. doi: 10.1006/jvbe.1999.1711

Doherty, M. E., & Kurz, E. M. (1996). Social judgment theory. *Thinking and Reasoning, 2*, 109–140. doi: 10.1080/135467896394474

Ebbesen, E. B., & Konečni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32*, 805–821. doi: 10.1037/0022-3514.32.5.805

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods, 40*, 130–136.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107–139. doi: 10.1037/0096-3445.127.2.107

Figueira, J., Greco, S., & Ehrgott, M. (2005). *Multiple criteria decision analysis: State of the art surveys series*. Boston, MA: Springer.

Gawronski, B., & Quinn, K. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic

evaluation. *Journal of Experimental Social Psychology, 49*, 120–125. doi: 10.1016/j.jesp.2012.07.016
Günaydin, G., Zayas, V., Selcuk, E., & Hazan, C. (2012). I like you but I don't know why: Objective facial resemblance to significant others influences snap judgments. *Journal of Experimental Social Psychology, 48*, 350–353. doi: 10.1016/j.jesp.2011.06.001
Graves, L., & Karren, R. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology, 45*, 313–340. doi: 10.1111/j.1744-6570.1992.tb00852.x
Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197–230. doi: 10.1016/S0010-0277(97)00044-9
Hahn, U., Prat-Sala, M., Pothos, E. M., & Brumby, D. P. (2010). Exemplar similarity and rule application. *Cognition, 114*, 1–18. doi: 10.1016/j.cognition.2009.08.011
Hannah, S. D., & Brooks, L. R. (2009). Featuring familiarity: How a familiar feature instantiation influences categorization. *Canadian Journal of Experimental Psychology, 63*, 263–275. doi: 10.1037/a0017919
Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231–237. doi: 10.1111/j.1467-9280.2009.02271.x
Herzog, S. M., & von Helversen, B. (in press). Blending and choosing within one mind: Should judgments be based on exemplars, rules or both? In M. Knauff, N. Sebanz, M. Pauen, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Berlin, Germany: Cognitive Science Society.
Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future – remembering the past. *Annual Review of Psychology, 51*, 631–664. doi: 10.1146/annurev.psych.51.1.631
Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633.
Moy, J. W., & Lam, K. F. (2004). Selection criteria and the impact of personality on getting hired. *Personnel Review, 33*, 521–535. doi: 10.1108/00483480410550134
Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106*, 259–298. doi: 10.1016/j.cognition.2007.02.003
Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review, 14*, 1140–1146. doi: 10.3758/BF03193103
Kraus, M. W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science, 21*, 518–522. doi: 10.1177/0956797610364949
Kuncheva, L. (2004). *Combining pattern classifiers: Methods and algorithms*. Hoboken, NJ: Wiley.
Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology, 48*, 563–574. doi: 10.1037//0022-3514.48.3.563
Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology, 11*, 112–126. doi: 10.1037/1089-2680.11.2.112
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57. doi: 10.1037/0096-3445.115.1.39
Nosofsky, R. M., & Bergert, F. B. (2007). Limitations of exemplar models of multi-attribute probabilistic inference. *Journal Experimental Psychology: Learning, Memory, and Cognition, 33*, 999–1019. doi: 10.1037/0278-7393.33.6.999
O'Brien, K. S., Latner, J. D., Halberstadt, J., Hunter, J. A., Anderson, J., & Caputi, P. (2008). Do antifat attitudes predict antifat behaviors? *Obesity, 16*, 87–92. doi: 10.1038/oby.2008.456
Persson, M., & Rieskamp, J. (2009). Inferences from memory: Strategy-and exemplar-based judgment models compared. *Acta Psychologica, 130*, 25–37. doi: 10.1016/j.actpsy.2008.09.010
Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing, 16*, 295–306. doi: 10.1016/S0262-8856(97)00070-X
Platzer, C., & Bröder, A. (2012). When the rule is ruled out: Exemplar and rules in decisions from memory. *Journal of Behavioral Decision Making*. Advance online publication. doi: 10.1002/bdm.1776
Stewart, L., & Perlow, R. (2001). Applicant race, job status, and racial attitude as predictors of employment discrimination. *Journal of Business and Psychology, 16*, 259–275. doi: 10.1023/A:1011113301301
Thibaut, J.-P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1403–1415. doi: 10.1037/0278-7393.32.6.1403
Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science, 21*, 779–785. doi: 10.1177/0956797610371965
von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology, 46*, 220–229. doi: 10.1037/a0016690
von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General, 137*, 73–96. doi: 10.1037/0096-3445.137.1.73
von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 867–889. doi: 10.1037/a0015501
Wigton, R. S. (1996). Social judgement theory and medical judgement. *Thinking and Reasoning, 2*, 175–190. doi: 10.1080/135467896394492
Zebrowitz, L., & Montepare, J. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*, 1–16. doi: 10.1111/j.1751-9004.2008.00109.x

Bettina von Helversen

Department of Psychology
University of Basel
Missionsstrasse 62a
4055 Basel
Switzerland
E-mail Bettina.vonhelversen@unibas.ch

Appendix

Table A1. Structure of the Tasks in Study 1a

Test applicant no.	Cue 1	Cue 2	Cue 3	Cue 4	Criterion	Item description
1	0	0	0	2	0	Model-fitting test applicant
2	0	1	0	1	14	Model-fitting test applicant
3	0	1	1	1	18	Model-fitting test applicant
4	0	1	0	3	21	Model-fitting test applicant
5	0	2	1	1	31	Model-fitting test applicant
6	0	4	0	2	39	Model-fitting test applicant
7	3	0	4	1	46	Model-fitting test applicant
8	3	4	1	2	61	Model-fitting test applicant
9	3	4	0	6	62	Model-fitting test applicant
10	4	1	4	2	70	Model-fitting test applicant
11	2	2	4	4	72	Model-fitting test applicant
12	4	4	4	4	98	Model-fitting test applicant
13	0	3	1	3	36	Morph (N1 & T6)
14	4	1	2	1	44	Morph (N2 & T5)
15	0	3	0	4	45	Morph (N3 & T6)
16	3	0	4	2	50	Morph (N4 & T5)
17	0	3	1	3	36	Neutral (N1)
18	4	1	2	1	44	Neutral (N2)
19	0	3	0	4	45	Neutral (N3)
20	3	0	4	2	50	Neutral (N4)
21	0	3	1	3	36	Morph (N1 & T2)
22	4	1	2	1	44	Morph (N2 & T2)
23	0	3	0	4	45	Morph (N3 & T1)
24	3	0	4	2	50	Morph (N4 & T1)
25	2	0	1	1	16	Training (T1)
26	3	1	0	0	21	Training (T2)
27	1	4	3	1	54	Training (T3)
28	2	1	3	3	53	Training (T4)
29	3	4	1	3	83	Training (T5)
30	4	4	3	4	90	Training (T6)

Notes. Cue value of 0 indicates a poor performance and cue value of 4 a very good performance. The abbreviations in parentheses indicate the pictures that were shown with a specific applicant's résumé information and which pictures were used to create the morphed pictures. For example, "Morph (N1 & T6)" indicates that this face is a morph between the pictures of neutral test applicant N1 and training employee T6.

Table A2. Structure of the Tasks in Study 1b

Test applicant no.	Cue 1	Cue 2	Cue 3	Cue 4	Criterion	Item description
1	0	0	1	0	14	Model-fitting test applicant
2	3	0	3	1	23	Model-fitting test applicant
3	3	1	3	4	26	Model-fitting test applicant
4	0	3	0	1	37	Model-fitting test applicant
5	4	0	4	0	40	Model-fitting test applicant
6	1	3	1	1	42	Model-fitting test applicant
7	3	4	0	1	59	Model-fitting test applicant
8	1	2	0	1	59	Model-fitting test applicant
9	4	1	4	3	67	Model-fitting test applicant
10	4	4	4	4	69	Model-fitting test applicant
11	4	4	0	4	71	Model-fitting test applicant
12	4	2	3	3	87	Model-fitting test applicant
13	4	0	3	0	29	Morph (N1 & T5)
14	4	1	3	0	48	Morph (N2 & T5)
15	3	3	0	4	77	Morph (N3 & T6)

(Continued on next page)

Table A2. (Continued)

Test applicant no.	Cue 1	Cue 2	Cue 3	Cue 4	Criterion	Item description
16	2	4	0	4	79	Morph (N4 & T6)
17	4	0	3	0	29	Neutral (N1)
18	4	1	3	0	48	Neutral (N2)
19	3	3	0	4	77	Neutral (N3)
20	2	4	0	4	79	Neutral (N4)
21	4	0	3	0	29	Morph (N1 & T2)
22	4	1	3	0	48	Morph (N2 & T1)
23	3	3	0	4	77	Morph (N3 & T2)
24	2	4	0	4	79	Morph (N4 & T1)
25	0	2	2	3	20	Training (T1/T5)
26	2	0	1	0	22	Training (T2/T6)
27	2	4	3	3	48	Training (T3)
28	4	2	1	4	49	Training (T4)
29	3	4	1	4	89	Training (T5/T1)
30	0	2	4	3	88	Training (T6/T2)

Notes. Cue value of 0 indicates a poor performance and cue value of 4 a very good performance. The abbreviations in parentheses indicate the pictures that were shown with a specific applicant's résumé information and which pictures were used to create the morphed pictures. For example, "Morph (N1 & T6)" indicates that this face is a morph between the pictures of neutral test applicant N1 and training employee T6. Pictures T1, T2, T5 and T6 were either shown with a high- or a low-performing training employee depending on the assignment condition. For example, "Training (T1/T5)" indicates that this employee's résumé information was shown with picture T1 in the first assignment condition and picture T5 in the second assignment condition.