Psychologia, 2019, **61**, 252–268 doi: 10.2117/psysoc.2019-A114

SOCIAL DIMENSIONS IN SIMILARITY JUDGMENT OF FACES

Sho ISHIGURO[1), 2)] and Satoru SAITO[1)]

[1)]*Kyoto University, Japan*
[2)]*Japan Society for the Promotion of Science, Japan*

We spontaneously infer the social traits of people from their appearance. In the current study, the possibility that the perceived similarity of faces is based on perceived social traits, more specifically, evaluations on valence and power dimensions, was tested. Pilot studies provided the dissimilarity data of Japanese female faces with the similarity judgment task and validated them with a memory task. The current study demonstrated that two axes provided by multidimensional scaling analysis to the dissimilarity data could be interpreted in terms of valence and power dimensions. As participants were not explicitly asked to focus on social dimensions when rating the similarities of faces, the results suggested that participants automatically used social dimensions in the similarity judgment task. The present study suggests that the similarity of faces across social dimensions affects the perceived similarity of faces.

Key words: face, appearance, dissimilarity, memory, multidimensional scaling

Introduction

People often infer social traits from faces (Hassin & Trope, 2000). Such appearance-based inferences of social traits are thought to be automatic and rapid (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008). In the present study, we hypothesized that people would spontaneously compare face stimuli in terms of inferred social traits, and tested this hypothesis based on Todorov and colleagues' two-dimensional model of appearance-based inference (Todorov et al., 2008). We begin by explaining appearance-based inferences and its dimensions. Thereafter, we clarify the aim of the current study and outline the methodology of the pilot studies.

Appearance-Based Inference

Although the accuracy of appearance-based inferences or reading traits from faces has been debated (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015), appearance-based inference is a prevalent phenomenon (Hassin & Trope, 2000). Previous literature

This work was supported by a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) fellows [grant number 17J05372].

The authors thank Yoshiyuki Ueda for providing access to the Kokoro Research Center database, Shiora Ikeda and Hideyuki Futatsugi for running the experiment in Pilot Study B, and Nozomu Kutomi for his advice on statistical analysis in the main study.

Correspondence concerning this article should be addressed to Sho Ishiguro, Graduate School of Education, Kyoto University, Japan (e-mail: ishiguro.sho.grocio@gmail.com).

suggest that appearance-based inferences of social traits from observing faces are automatic and rapid (Todorov et al., 2008). In relation to the automaticity, it has been demonstrated that the effect of appearance-based inferences is hard to change (Hassin & Trope, 2000; Olivola & Todorov, 2010; Todorov et al., 2015). Moreover, neuroimaging studies have provided clear evidence for its automaticity: the activity in the amygdala in response to face stimuli during tasks unrelated to judgment of personality traits, such as an age assessment task (Winston, Strange, O'Doherty, & Dolan, 2002) or a memory task (Engell, Haxby, & Todorov, 2007), explains the rated trustworthiness of face stimuli. These findings suggest that participants automatically infer personality traits from presented face stimuli even when they are not explicitly asked to infer personality traits (e.g., trustworthiness). Furthermore, Bar, Neta, and Linz (2006) showed that presentation of 39 ms of face stimuli led to accurate judgments of trustworthiness, demonstrating the rapidity of appearance-based inference (see also Rule & Ambady, 2008). Therefore, a variety of studies have suggested that appearance-based inference is a robust and significant phenomenon because people automatically and rapidly infer social traits from observing faces and rely on this type of inference.

Two Dimensions in Appearance-Based Inference

Todorov and colleagues investigated the process of appearance-based inference from neutral faces and proposed that appearance-based inference on social traits is based on two dimensions: valence/trustworthiness and power/dominance dimensions (Oosterhof & Todorov, 2008; Todorov et al., 2008; we call them valence and power dimensions, hereafter). Oosterhof and Todorov (2008) demonstrated that 13 personality traits inferred from the appearance of neutral faces could be reduced to two components using principal component analysis (PCA). Their results showed that the correlation coefficients between the first principal component (PC) and personality traits such as trustworthy, emotionally stable, responsible, sociable, and caring were *r*s > .90, suggesting that the first PC reflected inference on the valence dimension. The second PC was highly correlated to dominance (r = .93), moderately to aggressive and confident (*r*s > .65), suggesting that the second PC reflected inference on the power dimension. Sutherland et al. (2013) replicated these two dimensions with a larger sample of diverse face stimuli taken from the internet but also found a novel youthful-attractiveness dimension. Todorov et al. (2015) pointed out that the faces used in their previous experiments (e.g., Oosterhof & Todorov, 2008) did not vary depending on age and suggested that the variability in age of faces might determine the presence of the third dimension relating to age (i.e., youthful-attractiveness dimension). Importantly, the results of Sutherland and colleagues (2013) replicated the two dimensions assumed by Todorov and colleagues (Oosterhof & Todorov, 2008; Todorov et al., 2008), and suggested an additional dimension relating to age with diverse face stimuli that varied depending on age.

In a review by Todorov et al. (2008), valence and power dimensions were believed to originate from adaptive mechanisms. The evaluation of valence provides a signal for determining approach or avoidance behavior. Power evaluation provides a signal for a target's ability to realize intentions (e.g., whether a person has the ability to harm). They suggest that inferences based on evaluations of valence and power are even applied to neutral faces (i.e., emotion face overgeneralization, see also Zebrowitz & Montepare, 2008).

The Aim of the Current Study

We assumed that the two dimensions suggested by Todorov and colleagues (Oosterhof & Todorov, 2008; Todorov et al., 2008) would play an important role in perception of faces. Based on this assumption, we hypothesized that participants would spontaneously use valence and power dimensions while comparing face stimuli. The aim of the current study is to test this hypothesis by examining whether the similarity data collected in the similarity judgment task, which requires comparing face stimuli, can be explained in terms of valence and power dimensions.

To employ a data-driven approach to find the dimensional structure of data, we used a combination of the similarity judgment task and multidimensional scaling (MDS) analysis (Abe, Ohkawa, & Takano, 2008; Hirschberg, Jones, & Haggerty, 1978). By the similarity judgment task, data on similarities or dissimilarities of stimulus pairs can be collected. Based on that dissimilarity data, MDS can place stimuli on a multidimensional space. Note that similarity values collected in the similarity judgment task can be transformed to dissimilarity values, which represent the psychological distances between stimuli.

A strength of the similarity judgment task is that a researcher does not need to orient participants to specific dimension (s), nor does a researcher have to predetermine any particular dimension (s) to be targeted. Instead, participants can judge the similarity of stimulus pairs and the researcher can then investigate which dimension (s) they use by MDS analysis to the dissimilarity data collected in the similarity judgment task. Hirschberg et al. (1978) applied MDS analysis to the dissimilarity data collected through the similarity judgment task of face stimuli and suggested that the dimensions provided by MDS were interpretable as affective dimensions such as desirability and masculine attractiveness.

Nevertheless, as far as we know, valence and power dimensions suggested by Todorov and colleagues' two-dimensional model (Oosterhof & Todorov, 2008; Todorov et al., 2008) have not yet been tested by MDS analysis to the data of the similarity judgment task. Therefore, in the main study, we tested the above-mentioned hypothesis that participants would spontaneously use valence and power dimensions in comparing face stimuli, by conducting the similarity judgment task, applying MDS to the dissimilarity data acquired by the similarity judgment task, and examining whether the two dimensions provided by MDS can be explained by valence and power dimensions. More specifically, we rotated the constellation provided by MDS along with valence and power variables that were estimated by personality trait ratings from an existing database with PCA. Then, we tested whether the axes of the constellation could be interpreted in terms of valence and power dimensions.

Pilot studies. To address the purpose of the main study, we also conducted two pilot studies (Pilot Studies A and B). In Pilot Study A, we asked participants to rate the similarity of pairs to collect the dissimilarity data (i.e., pairwise similarity, Chang, Nemrodov, Lee, & Nestor, 2017; Nishimura, Maurer, & Gao, 2009). The values of the similarity ratings were transformed to dissimilarity values and these dissimilarity data were used for Pilot Study B and the main study.

In Pilot Study B, we validated the dissimilarity data with a memory task as detrimental effects of similarity of faces on memory performance had been demonstrated (Light, Kayra-Stuart, & Hollander, 1979; Smyth, Hay, Hitch, & Horton, 2005). By cluster analysis of the dissimilarity data (Nagata & Munechika, 2001), we constructed two sets for Pilot Study B: the "similar" set that consisted of similar stimuli and the "dissimilar" set that consisted of dissimilar stimuli. We predicted that memory performance should be higher for the dissimilar set than for the similar set if the dissimilarity data were valid.

For the memory task, we selected the serial reconstruction task, which required participants to select a series of stimuli in the presented order (Smyth et al., 2005). In this task, participants do not have to recall stimuli (i.e., item memory) because the all stimuli are presented at recall but have to recall the presentation order of stimuli (i.e., order memory; for the relationship between the serial reconstruction task and order memory, see also Saint-Aubin & Poirier, 1999). We considered that the serial reconstruction task was an appropriate memory task to validate the dissimilarity data for three major reasons. First, the serial reconstruction task methodologically makes participants focus on the similarity between given stimuli. In our study, participants selected each face stimulus in the correct order from six face stimuli presented at recall as candidates, which let them view and compare these six candidates. Second, the serial reconstruction task was likely to detect the possible effect of similarity. The effect of similarity is generally detrimental to order memory whereas it can be both facilitative and detrimental to item memory (Logie, Saito, Morita, Varma, & Norris, 2016). If we had used a memory task measuring item memory, facilitative and detrimental effects would have offset each other, which could have made it difficult to detect the possible effect by similarity. Furthermore, the serial reconstruction task enables presenting the same set of stimuli repeatedly across trials with different presentation orders, which minimizes the effect of item memory on memory performance. Therefore, it was likely to detect the possible effect by similarity. Lastly, the detrimental effect of similarity of faces was already confirmed by the serial reconstruction task. A study by Smyth et al. (2005) using the serial reconstruction task of face stimuli demonstrated that memory performance was greater for the lists in which face stimuli were dissimilar to each other than for the lists in which face stimuli were similar to each other.

Although we assumed that there was a commonality in memory and perception of faces (Chang et al., 2017), and that the serial reconstruction task could validate the dissimilarity data, participants might have used mnemonic strategies that would diminish the similarity effect in the serial reconstruction task. We particularly addressed one such strategy, verbal labeling. If participants verbally label face stimuli at encoding, labels would help distinguish stimuli at recall. Critically, labels would be made arbitrarily, independent of visual features and/or social traits inferred based on visual features. For example, participants could arbitrarily assign numbers to six stimuli of a set, because the same set of six face stimuli were repeatedly used in our serial reconstruction task (see also, the method section of Pilot Study B). In Pilot Study B, we used articulatory suppression to minimize the effect of verbal labeling (Smyth et al., 2005). Articulatory suppression is a dual-task technique that requires participants to utter irrelevant sounds or words during the primary task (e.g., a memory task) and that is thought to prevent verbal labeling. In Pilot Study B, participants repeatedly uttered "da" when viewing sequences of face stimuli.

Pilot Study A

The purpose of Pilot Study A was to collect similarity ratings for face stimuli using the similarity judgment task. These similarity ratings were transformed to dissimilarity values, which were used in Pilot Study B and the main study.

Method

Participants

Thirty undergraduate and graduate students (14 women and 16 men) from Kyoto University participated. The mean age was 21.33 years (SD = 2.06). The ethics application for the current study (for Pilot Studies A and B and the main study) had been approved by the Graduate School of Education, Kyoto University, and participants who confirmed their participation were a part of the study.

Materials

Pictures of 24 neutral Japanese female faces from the facial expression database of Kokoro Research

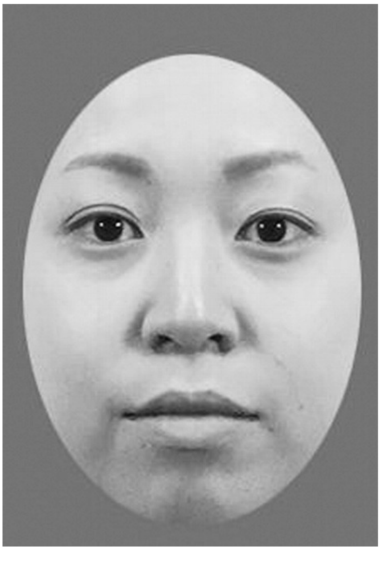

Fig. 1. An example of face stimuli

Center (KRC; Ueda, Nunoi, & Yoshikawa, 2019) were used. The KRC database has data of face stimuli of 74 Japanese university students (50 female and 24 male students). For each person, photographs of seven facial expressions (neutral, happy, angry, disgusted, sad, fearful, and surprised) across three gaze directions (divert, left-averted, and right-averted), three face directions (front, left-oriented, and right-oriented at 45 degree angles) and a front view of neutral expression with closed-eyed face were taken (64 in total for each person). Details of the database are reported in Ueda et al. (2019) and a researcher needs to contact Dr. Yoshiyuki Ueda at Kyoto University, the first author of Ueda et al. (2019), for permission to use the KRC database. The database can be used free of charge for non-commercial academic use.

To prevent the effect of hairstyle and/or hair color, the cropped face pictures used by Fujino et al. (2017) were used in the present study. Fig. 1 is an example of face stimuli. Fujino et al. (2017) selected these 24 stimuli from the KRC database based on two criteria. First, clear pictures were selected considering the lighting and focus. Second, pictures of emotions of which were discriminable were selected: the neutral face of a person was discriminable from the angry and smile faces of the same person. Top 24 emotionally discriminable faces were used in Fujino et al. (2017). The ID numbers of these 24 stimuli were as follows: 01, 06, 07, 08, 09, 13, 14, 15, 17, 18, 20, 21, 24, 25, 31, 32, 33, 34, 35, 38, 39, 42, 46, and 47.

Task

A similarity judgment task was conducted. Similar to a previous study (Nishimura et al., 2009), participants were required to rate the dissimilarity of a pair of two face stimuli on a 7-point Likert scale from 1 (*very dissimilar*) to 7 (*very similar*).

Procedure

Stimuli were 5.5 cm wide and 7.5 cm high. The distance from participants' eyes to the screen was approximately 50 cm. Prior to the test trials, participants took part in practice trials with six male face stimuli not used in the test trials. First, six face stimuli were presented sequentially, at a rate of 1 s per stimulus. Then, two randomly chosen face stimuli were presented, and participants rated the similarity between them. After providing their rating, participants were presented with all six stimuli and were asked to select the combination of the most similar pair of face stimuli. They were instructed to give a rating of 7 (i.e., "very similar") when presented with such a pair in the test trials. They were also instructed to select the most dissimilar pair and give a rating of 1 (i.e., "very dissimilar") to such a pair in the test trials. After participants confirmed their understanding of the task, the test trials were conducted.

In the test trials, all combinations of 276 pairs from a set of 24 stimuli were rated (${}_{24}C_2 = 276$). First, to familiarize participants with the set of stimuli, 24 face stimuli were presented one by one at the rate of 1s. Next, a pair of stimuli was presented on the screen (Fig. 2). To minimize the possible effect of position, the

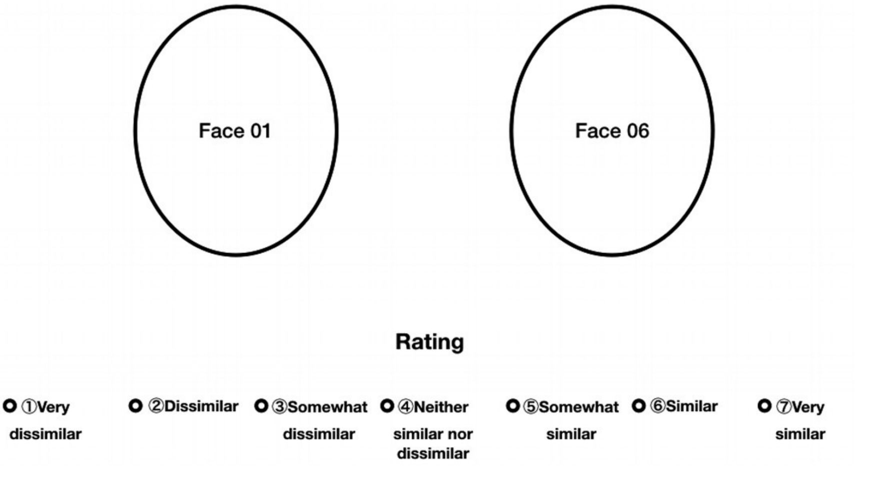

Fig. 2. Schematic illustration of the rating task in Pilot Study A

positions of stimuli on the screen were counterbalanced (e.g., for the 01 and 06 pair, half of the participants saw stimulus 01 on the left side and stimulus 06 on the right, whereas the other half saw stimulus 06 on the left and stimuli 01 on the right). During the rating phase, 276 pairs were presented, one pair after another, with order randomized by participant. Participants set their own pace in rating each pair and completed the task voluntarily. Participants finished the experiment within 40 min and received a book coupon of 1,000 JPY in return for their participation.

Result and Discussion

The rating for each pair was transformed for dissimilarity value by the expression of "8 – *x*," where *x* is the rating. The dissimilarity data can be available on request to Dr. Yoshiyuki Ueda, who is responsible for the KRC database.

Pilot Study B

Pilot Study B aimed to validate the dissimilarity data collected in Pilot Study A, using the serial reconstruction task. We validate the dissimilarity data by testing the prediction that serial order memory performance for the dissimilar set should be higher than for the similar set. Alternatively, if the dissimilarity data are not valid, we would not find the differences of serial order memory performance for the two stimulus sets.

Method

Participants

Twenty undergraduate and graduate students who confirmed their participation (4 women and 16 men) from Kyoto University were part of the experiment. The mean age of the participants was 22.05 years (SD = 2.01). The sample size was determined by planned interim analysis (e.g., Michaelson & Munakata, 2016), in which an interim power analysis was conducted with a planned sample size to determine the ideal sample size (see the procedure section below).

Materials

Two sets of six face stimuli from among the 24 stimuli in Pilot Study A were used (12 stimuli in total). The two sets were constructed by hierarchical cluster analysis with Ward's method. Six clusters were obtained by applying cluster analysis to the dissimilarity data and cutting the cluster tree at the height of five (Fig. 3). Because cluster analysis groups similar items together into a cluster and separates dissimilar items into different clusters (Nagata & Munechika, 2001), selecting stimuli from the same cluster leads to a set of similar stimuli while choosing stimuli from different clusters results in a set of dissimilar stimuli. Accordingly, six face stimuli (IDs: 14, 15, 21, 24, 33, 34) from the same cluster were allocated to the similar set and other six face stimuli (IDs: 01, 06, 07, 17, 20, 31) from six different clusters were allocated to the dissimilar set.

Task

With reference to Smyth et al. (2005), the serial reconstruction task was used. In this task, six stimuli were presented one by one. At the end of a trial, participants were asked to recall the presentation order of the stimuli. Although the similar and dissimilar sets were used repeatedly across trials, the presentation

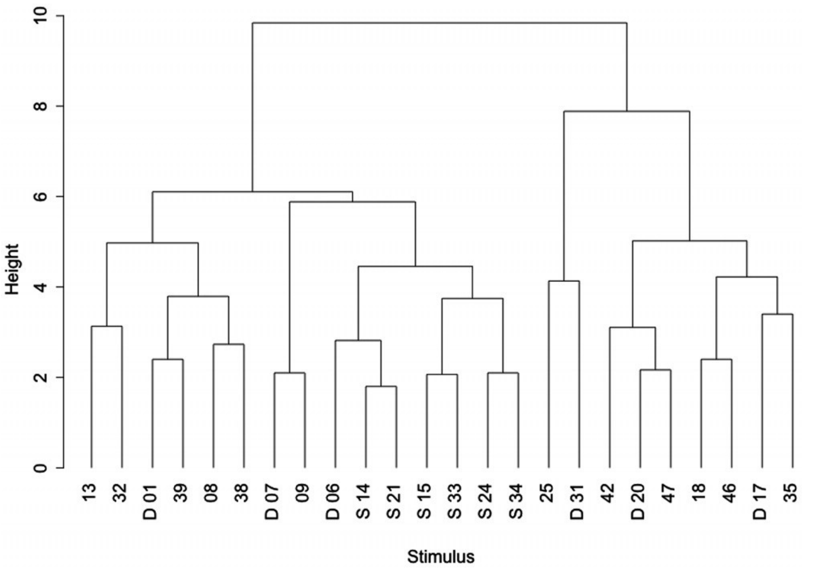

Fig. 3. Cluster tree based on dissimilarities. Numbers indicate stimulus IDs. Letters indicate similar (S) or dissimilar (D) sets.

order of stimuli varied. Using a Latin square, combinations of two adjacent stimuli were repeated only once. The test session included two blocks of 12 test trials. The similar set was used in one block and the dissimilar set in the other. The presentation order of blocks was counterbalanced. The order of trials within each block (i.e., trials of the similar set or these of the dissimilar set) was randomized by participant.

Procedure

As in Pilot Study A, stimulus size was 5.5 x 7.5 cm. The distance from participants' eyes to the screen was approximately 50 cm. The experiment consisted of a brief instruction, a practice of articulatory suppression, practice trials, then followed by the test session. Following the explanation phase, participants practiced articulatory suppression by saying "da" every 0.5 s, in time with a metronome, for 20 s. Participants then completed three practice trials with male faces, and the 24 test trials of the test session with female faces. A test trial is described as follows (Fig. 4).

Following a 200 ms beep sound, a fixation cross was presented at the center of the screen for 500 ms, and the participant commenced articulatory suppression (i.e., they started and kept saying "da" every 0.5 s). Subsequently, a blank screen was presented for 500 ms, followed by the face stimuli presented at the rate of 1000 ms per stimulus, with an inter-stimulus interval of 1000 ms. When the sixth stimulus disappeared, a blank screen was presented for 2000 ms, and then all six stimuli from the current trial (with randomized positions on the screen) and a question mark "?" were presented. Immediately, the participant stopped articulatory suppression and began selecting the stimuli in the correct order of their presentation. If the participant could not remember the correct stimulus for a position in the order, they selected "?". The selected face stimuli then disappeared from the screen. Once six responses were given, a button with a message: "Move to the next trial" appeared on the screen. When ready, the participant clicked the button and moved on to the next trial. The experiment was completed within 40 min. Participants received a book coupon (1,000 JPY) in return for their participation.

Planned interim analysis was used to determine the sample size and prevent inflation of a Type I error. Prior to the experiment, a sample size of 20 was set, and the effect size was calculated based on this sample size as an interim analysis. The effect size was dz = 0.88 and power analysis showed that 13 participants were required to secure a statistical power of $1 - \beta = .8$ with a significance level of $\alpha = .05$. Therefore, a final sample size of $N = 20$ was recruited.

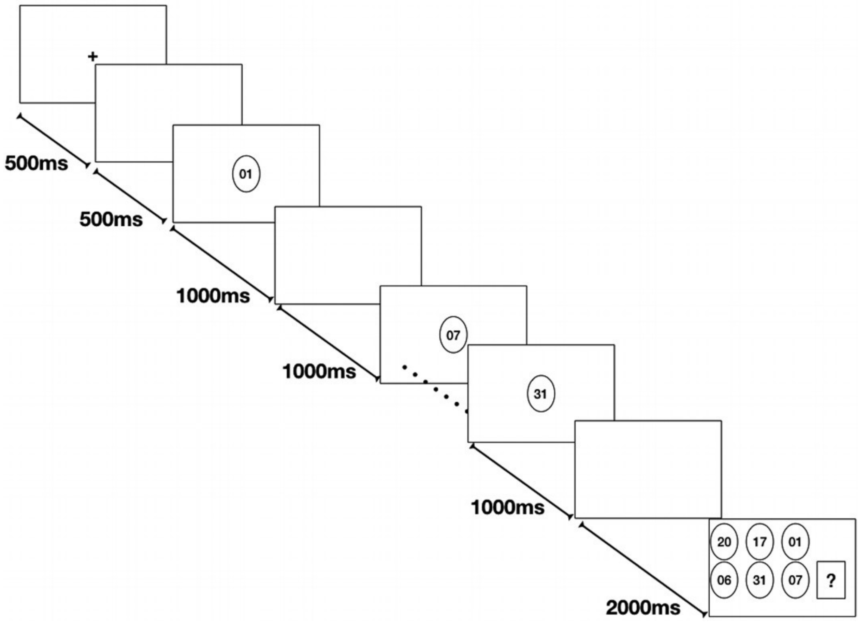

Fig. 4. Schematic illustration of a task trial in Pilot Study B

Result and Discussion

Correct-in-position scoring was used, where one point was given for each stimulus selected in its correct position. For example, if stimuli 01, 06, 07, 17, 20, and 31 were presented in that order, and a participant selected "01, ?, 17, 07, 20, ?", two points would be given. The scores were summed for each of the two sets, with a theoretical range of 0–72. Table 1 shows the mean scores and standard deviations for the two sets.

The results were consistent with our prediction. Mean scores for the dissimilar set were higher than those of the similar set, $t(19) = 3.93$, $p < .001$, $dz = 0.88$. Pilot Study B suggested that the ratings collected in Pilot Study A were significant data explaining the differences in serial order memory performance and thus, validated the dissimilarity data.

Table 1. Memory Performance for Dissimilar Set and Similar Set ($N = 20$)

List type	*Mean*	*SD*	Range
Dissimilar set	39.4	10.36	25 - 66
Similar set	30.2	8.12	14 - 48

Note. SD = Standard deviation.

The Main Study

The main study tested the possibility that participants rated similarity based on social dimensions, considering the perspective that inference on social traits is automatic and rapid (Todorov et al., 2008). Oosterhof and Todorov (2008) suggested that two components, valence and power, formed the basis for social trait inference from neutral faces (see also Sutherland et al., 2013). Within the framework of Todorov and colleagues' two-dimensional view (2008), we tested whether participants rated similarity in terms of valence and power through the following three steps: (1) Reducing the seven personality traits in the KRC database into two principal components by PCA to acquire values on two dimensions (i.e., valence and power), (2) obtaining the two-dimensional constellation of face stimuli by applying MDS to the dissimilarity data, and (3) rotating the MDS constellation of face stimuli along with the first component of PCA. Subsequently, the relationship between the principal components and the axes of the rotated constellation of face stimuli was investigated.

The personality traits in the KRC database and the dissimilarities in Pilot Study A were collected independently. With respect to the data units, the KRC database's personality traits are ratings of individual stimuli, whereas the dissimilarities obtained in Pilot Study A are ratings of pairs of stimuli, representing the differences between them. More importantly, the task of judging similarity in Pilot Study A did not explicitly require inference of personality traits. Given the independence of the observations and the difference in data units and task instructions, observation of a relationship between the PCA results for personality traits and the MDS results for dissimilarities would indicate a common psychological process for both judgments of personality traits and rating of similarity.

Method

Materials

The following seven personality traits for all 50 female faces in the KRC database were used for PCA: attractiveness, compassion, competence, dominance, extraversion, maturity, and trustworthiness. These traits were rated by 30 participants (Ueda et al., 2019). Dissimilarity data collected in Pilot Study A was used to get the MDS constellation.

Procedure

Step 1. The seven personality traits were reduced into principal components using the *prcomp* function in R (R Core Team, 2019). Loadings on the principal components are shown in Table 2. The positive traits loaded strongly on the first component, whereas traits relating to power, such as dominance and maturity, loaded strongly on the second component. These two components also well explain the overall variance (62.5% and 22.7%, respectively). The third component, however, was difficult to interpret in terms of personality traits, explaining only 8.7% of the variance. Consistent with a previous study, we concluded that the first two components were critical (see also, Ueda et al., 2019). The first component was labeled "valence", and the second "power." Although not all the seven personality traits in the KRC database were the same as in a previous study by Oosterhof and Todorov (2008), patterns of loadings in the current study were similar. Trustworthiness (or trustworthy) was highly correlated to the first PC (r = .88 in

Table 2. Loading of KRC Database Trait Judgments on the First Three Principle Components

Trait	Component 1	Component 2	Component 3
Attractiveness	.91	–.21	–.19
Compassion	.79	–.52	.21
Competence	.96	.08	.01
Dominance	.53	.76	–.31
Extraversion	.90	.20	–.23
Maturity	.36	.72	.58
Trustworthiness	.88	–.36	.19
Variance explained	62.9%	22.7%	8.7%

Note. Loadings denote correlations between the principle components and trait judgments.

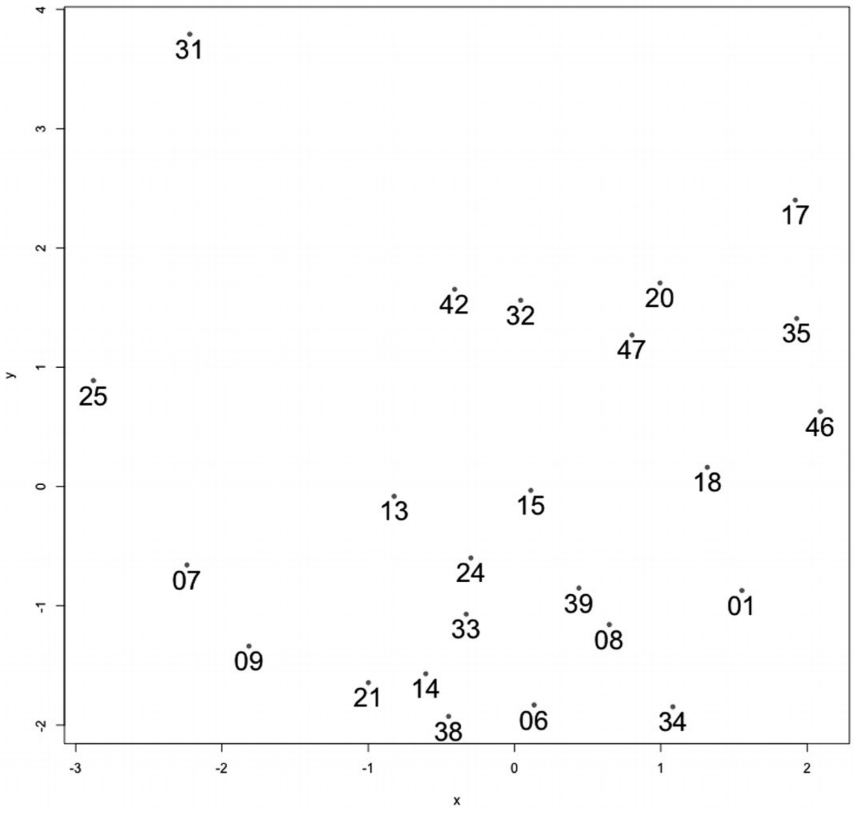

Fig. 5. Multidimensional scaling constellation of stimuli after rotation

the current study and $r = .94$ in Oosterhof and Todorov's study) and dominance was highly correlated to the second PC ($r = .76$ in the current study and $r = .93$ in Oosterhof and Todorov's study).

Step 2. A two-dimensional constellation of stimuli by MDS was obtained. Stimuli judged as similar are located close to one other whereas dissimilar stimuli are separated. We regarded that the MDS constellation represents the dissimilarity data appropriately.

Step 3. The MDS constellation of faces by Step 2 was rotated so that an axis was maximally correlated to the valence variable calculated in Step 1 (i.e., the first PC). As the critical factor in MDS is dissimilarity, rotation does not affect the MDS results, and is considered an acceptable method for interpretation (Nagata & Munechika, 2001). Rotation was performed using *MDSrotate* in the *vegan* R package (Oksanen et al., 2019). The results of Step 3 are presented in Fig. 5.

Result and Discussion

Correlation coefficients were calculated between the values of the two principal components (i.e., valence and power variables) from Step 1 and the values of the *x*- and *y*-axes of the rotated constellation from Step 3 (Table 3). As the constellation was rotated to ensure that the *x*-axis was maximally correlated to valence, their correlation was high, $r = .60$ ($p < .01$). More importantly, the *y*-axis was strongly correlated to power, $r = .47$ ($p < .05$). As the *y*-axis was determined by rotation based on the *x*-axis, there was no a priori statistical constraint associating the *y*-axis with power; this implies that the *y*-axis of the MDS constellation reflects perceived power from face stimuli. The *x*- and *y*-axes were selectively associated with valence and power, respectively, because the *x*-axis was barely correlated with power, $r = -.19$ ($p = .39$, *ns*), and the *y*-axis was almost non-

Table 3. Correlation of Principle Components and with *x*- and *y*-axes of Multidimensional Scaling Constellation After Rotation Based on Component 1

	Constellation x-axis	Constellation y-axis
Component 1: Valence	.60	.05
Component 2: Power	–.19	.47

Table 4. Correlation of Principle Components and with *x*- and *y*-axes of Multidimensional Scaling Constellation After Rotation Based on Component 2

	Constellation x-axis	Constellation y-axis
Component 1: Valence	–.21	.53
Component 2: Power	.52	.07

correlated with valence, $r = .05$ ($p = .81$, *ns*).

Rotation based on an axis and power provided similar results (Table 4). The correlation between the *x*-axis and power was $r = .52$ ($p < .01$), and correlation between the *y*-axis and valence was $r = .53$ ($p < .01$). This verified our hypothesis that two axes in the MDS constellation correspond to valence and power. We also split the data into two subsets and applied the analyses in Steps 1–3 to keep a check on the robustness of the results. The stimuli were sorted by ID and assigned numbers. For both the even-number and the odd-number subset, critical correlations between the *x*-axis and valence and between the *y*-axis and power were $r\text{s} > .34$ except that a weak correlation between valence and *y*-axis ($r = .12$) when rotating MDS constellation by power dimension for odd-number subset. Considering the limitations of the above-mentioned procedure based on only half of the 24 stimuli, these results, in general, supported the robustness.

Power dimension was relatively weakly related to an axis of MDS constellation ($r\text{s} = .47$ or $.52$), compared with the correlation between valence dimension and an axis of MDS constellation ($r\text{s} = .60$ or $.53$). As masculinity is associated with dominance (Boothroyd, Jones, Burt, & Perrett, 2007; Sutherland et al., 2013), excluding male faces might have reduced the variance on power dimension. As a consequence, power dimension in the current study might not have represented power dimension well.

The results of the combined PCA and MDS analysis implied that participants rated similarities between face stimuli based on social dimensions such as valence and power, even if they were not explicitly asked to pay special attention to valence and power dimensions.

General Discussion

We collected similarity ratings for face stimuli from the KRC database (Pilot Study A), validated the ratings in terms of serial order memory performance (Pilot Study B), and interpreted them along with principal components based on personality traits (the main study). In the main study, to test a possible effect of social dimensions on the perceived similarity, the axes of the MDS constellation for the dissimilarity data were interpreted according to two social dimensions, valence and power, drawn from Todorov and colleagues' two-dimensional model (Oosterhof & Todorov, 2008). Values on the axes were highly and selectively correlated to valence or power, respectively. These results verified that the perceived similarity was at least partly affected by social dimensions, even though participants were not explicitly asked to focus on them. This is consistent with the argument that appearance-based inference is automatic (Todorov et al., 2008). Since face stimuli in Pilot Study A were presented without additional information, such as verbal information about the personalities of people, we hypothesized that social inference was primarily based on visual information of face stimuli and social inference, in turn, affected the perception of similarity. That is, similarity on social dimensions or social similarity mediated the relationship between visual and perceived similarity.

Nevertheless, given the importance of visual information of facial features in inference on social traits (Abe et al., 2008; Dotsch & Todorov, 2012; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013; Yamada, 1993), an alternative account to the results of the current study is possible. That is, perceived similarity was solely based on the similarity of visual features but not based on social similarity. Because visually similar faces were evaluated similarly on social dimensions, there is an apparent correlation between social similarity and perceived similarity. Visual similarity determines perceived similarity while social similarity does not mediate the effect of visual similarity. Although the present data cannot rule out this alternative account, there is a possible causal role of social similarity in the perceived similarity because social similarity can affect the perceived similarity independent of visual similarity. Hassin and Trope (2000) manipulated impression for faces with verbal descriptions and demonstrated that pairs of face stimuli presented with descriptions of the same social trait (e.g., for two face stimuli, two different descriptions both indicated kindness) were perceived more similar than pairs of face stimuli presented with descriptions showing different social traits respectively (e.g., a description showed kindness and the other unkindness). Hassin and Trope (2000) interpreted the results as demonstrating that the similarity of social traits provided by descriptions directly affected the perceived similarity. In addition, studies of memory confusion paradigm (Bor, 2017; van Leeuwen, Park, & Penton-Voak, 2012) have demonstrated that verbal information describing people's traits directly affects the similarity of memory for people, independent of visual information. Therefore, previous studies suggest that social similarity would directly lead to the perceived similarity in perception and memory, which supports the assumption on the mediation effect of social similarity.

We also argue that appearance-based inference on social traits could have spontaneously occurred prior to or simultaneously with perception of similarity based on the view that appearance-based inference is rapid and automatic (e.g., Oosterhof & Todorov, 2008). Considering the rapidity and automaticity of social inference and the possible causal effect of social similarity on perceived similarity, we suggest that social similarity mediates the relationship between visual and perceived similarity.

An additional alternative account would be that critical visual dimensions or visual cues that were used for inference on social traits also played an important role in the perception of similarity. Previous studies suggest that some facial features such as the mouth and eyebrow are particularly important for inference on social traits (Dotsch & Todorov, 2012; see also Todorov et al., 2013). Takano and Abe (1996) also demonstrated that visual dimensions provided by MDS analysis to dissimilarity data of face were associated to visual features such as contour, eye, and eyebrow, suggesting that these features are the basis for visual dimensions of perceiving similarity of faces. More recently, two visual dimensions of faces based on visual features have been particularly highlighted and they have been interpreted as balance and form dimensions (i.e., "babyfaceness" and roundness; e.g., Takano, 2001). Abe et al. (2008) demonstrated that balance and form dimensions were correlated to maturity and approachability, respectively. The view by Takano, Abe, and associates (e.g., Takano, 2001; Abe et al. 2008) would indicate that critical visual dimensions were utilized for both perceptions of similarity and inferences on social traits. If this is the case, association of social and perceived similarity can be attributable to common visual dimensions and our assumption on the mediation effect of social similarity might be redundant. However, our assumption is not incompatible with the alternative account and offers a further explanation as to why common visual dimensions emerge in social inference and perception of similarity, by supposing the influence of rapid and automatic inference on social traits over subsequent cognitive processes such as similarity judgment.

Based on the above reasoning and the view that inference on social traits is consequential (Todorov et al., 2015), we propose that perceived similarity reflects not merely visual similarity but also social similarity. In addition, results imply that social dimensions are automatically used, because we did not explicitly orient participants to social dimensions in the similarity judgment task. Given that the perceived similarity was validated by the serial reconstruction task in Pilot Study B, our view casts doubt on the interpretation of the effect of facial similarity in terms of merely visual similarity, which is commonly accepted in memory research (Hurlstone, Hitch, & Baddeley, 2014; Smyth et al., 2005; see also Majerus et al., 2010). For example, Smyth et al. (2005) used sets of face stimuli that consisted of dissimilar faces in terms of properties such as gender, hairstyle, and face shape, as "visually dissimilar memory sets". Our view suggests that face stimuli convey above and beyond visual information (Hirschberg et al., 1978) and that internal processing for face stimuli is not only visual processing (e.g., Stolier, Hehman, Keller, Walker, & Freeman, 2018). While it is beyond the scope of the current study, future studies which manipulate visual and social similarity independently would be able to critically analyze the relationship between the critical dimensions of visual, social and perceived similarity.

AUTHOR'S CONTRIBUTION

S.I. and S.S. conceived the experiments. S.I. contributed to acquisition of data. S.I. analyzed the data. S.I. and S.S. contributed in interpreting the results. S.I. and S.S. wrote the paper. Both authors gave final approval for publication.

DISCLOSURE OF INTEREST

The authors report no conflicts of interest.

REFERENCES

Abe, T., Ohkawa, M., & Takano, R. (2008). Youbou no inshoukeisei ni oyobosu kahanka no eikyo [Effect of overgeneralization on evaluation of facial impressions]. *Nihon Kaogakkai Shi*, *8*(1), 87–96.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278. doi: 10.1037/1528-

3542.6.2.269

Boothroyd, L. G., Jones, B. C., Burt, D. M., & Perrett, D. I. (2007). Partner characteristics associated with masculinity, health and maturity in male faces. *Personality and Individual Differences*, *43*(5), 1161–1173. doi: 10.1016/j.paid.2007.03.008

Bor, A. (2017). Spontaneous categorization along competence in partner and leader evaluations. *Evolution and Human Behavior*, *38*(4), 468–473. doi: 10.1016/j.evolhumbehav.2017.03.006

Chang, C.-H., Nemrodov, D., Lee, A. C. H., & Nestor, A. (2017). Memory and perception-based facial image reconstruction. *Scientific Reports*, *7*(1), 6499. doi: 10.1038/s41598-017-06585-2

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562–571. doi: 10.1177/1948550611430272

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519. doi: 10.1162/jocn.2007.19.9.1508

Fujino, M., Ueda, Y., Inoue, W., Ooishi, Y., Kitagawa, N., & Nomura, M. (2017, December). *Dousatsu meiso no tanki kainyuu ga hyoujou heno chuui baiasu wo henyou saseru* [Short-term intervention of open monitoring meditation alters attentional bias for facial expressions]. Paper presented at the Japanese Association of Mindfulness 4th Annual Meeting, Tokyo, Japan.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852. doi: 10.1037/0022-3514.78.5.837

Hirschberg, N., Jones, L. E., & Haggerty, M. (1978). What's in a face: Individual differences in face perception. *Journal of Research in Personality*, *12*(4), 488–499. doi: 10.1016/0092-6566(78)90074-0

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. doi: 10.1037/a0034221

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 212–228. doi: 10.1037/0278-7393.5.3.212

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition*, *44*(4), 590–607. doi: 10.3758/s13421-015-0580-9

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., van der Linden, M., Collette, F., . . . Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–2593. doi: 10.1162/jocn.2009.21378

Michaelson, L. E., & Munakata, Y. (2016). Trust matters: Seeing how an adult treats another person influences preschoolers' willingness to delay gratification. *Developmental Science*, *19*(6), 1011–1019. doi: 10.1111/desc.12388

Nagata, Y., & Munechika, M. (2001). *Tahenryo kaisekiho nyumon* [Introduction to multivariate analysis]. Tokyo, Japan: Saiensu-sha.

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology*, *103*(3), 355–375. doi: 10.1016/j.jecp.2009.02.005

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., . . . Wagner, H. (2019). *Vegan: Community ecology package* (Version 2.5-4) [R software package]. Available from https://CRAN.R-project.org/package=vegan

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324. doi: 10.1016/j.jesp.2009.12.002

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. doi: 10.1073/pnas.0805664105

R Core Team. (2019). *R: A language and environment for statistical computing* [Software]. Retrieved from https://www.R-project.org

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105. doi: 10.1016/j.jesp.2007.12.001

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology Section A*, *52*(2), 367–394. doi: 10.1080/713755814

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual—spatial domain: Reconstructing sequences of unfamiliar faces. *The Quarterly Journal of Experimental Psychology Section A*, *58*(5), 909–930. doi: 10.1080/02724980443000412

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, *115*(37), 9210–9215. doi: 10.1073/pnas.1807222115

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. doi: 10.1016/j.cognition.2012.12.001

Takano, R. (2001). Mekyappu no saiensu [Science of makeup]. In I. Daibou (Ed.), *Kesho kodo no shakai shinrigaku* [Social psychology of cosmetic behavior] (pp. 90–101). Kyoto, Japan: Kitaohji-shobo.

Takano, R., & Abe, T. (1996, September). *Kaono ruijisei to keitai tokucho no kanren* [Relationship between facial similarity and facial features]. Paper presented at the 60th Conference of the Japanese Psychological Association, Tokyo (p. 700). Tokyo, Japan.

Todorov, A., Dotsch, R., Porter, J., Oosterhof, N. N., & Falvello, V. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–738. doi: 10.1037/a0032335

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. doi: 10.1146/annurev-psych-113011-143831

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. doi: 10.1016/j.tics.2008.10.001

Ueda, Y., Nunoi, M., & Yoshikawa, S. (2019). Development and validation of the Kokoro Research Center (KRC) facial expression database. *Psychologia*, *61*(4), 221–240. doi: 10.2117/psysoc.2019-A009

van Leeuwen, F., Park, J. H., & Penton-Voak, I. S. (2012). Another fundamental social category? Spontaneous categorization of people who uphold or violate moral norms. *Journal of Experimental Social Psychology*, *48*(6), 1385–1388. doi: 10.1016/j.jesp.2012.06.004

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283. doi: 10.1038/nn816

Yamada, H. (1993). Visual information for categorizing facial expression of emotions. *Applied Cognitive Psychology*, *7*(3), 257–270. doi: 10.1002/acp.2350070309

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. doi: 10.1111/j.1751-9004.2008.00109.x

(Manuscript received 29 May, 2019; Revision accepted 22 November, 2019; Released online in J-STAGE as advance publication 30 June, 2020)