Heliyon 5 (2019) e01736

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Arbitrary signals of trustworthiness – social judgments may rely on facial expressions even with experimentally manipulated valence

Ferenc Kocsor [a], Luca Kozma [a,*], Adon L. Neria [b], Daniel N. Jones [b], Tamas Bereczkei [a]

[a] Institute of Psychology, University of Pécs, Hungary
[b] Department of Psychology, University of Texas, El Paso, USA

ARTICLE INFO

Keyword:
Psychology

ABSTRACT

Generalization has been suggested as a basic mechanism in forming impressions about unfamiliar people. In this study, we investigated how social evaluations will be transferred to individual faces across contexts and to expressions across individuals. A total of 93 people (33 men, age: M = 29.95; SD = 13.74) were exposed to facial images which they had to evaluate. In the *Association* phase, we presented one individual with (1) a trustworthy, (2) an untrustworthy, (3) or an ambiguous expression, with either positive or negative descriptive sentence pairs. In the *Evaluation* phase participants were shown (1) a new individual with the same emotional facial expression as seen before, and (2) a neutral image of the previously presented individual. They were asked to judge the trustworthiness of each person. We found that the valence of the social description is transferred to both individuals and expressions. That is, the social evaluations (positive or negative) transferred between the images of two different individuals if they both displayed the same facial expression. The consistency between the facial expression and the description, however, had no effect on the evaluation of the same expression appearing on an unfamiliar face. Results suggest that in social evaluation of unfamiliar people invariant and dynamically changing facial traits are used to a similar extent and influence these judgements through the same associative process.

1. Introduction

Previous studies have demonstrated how evaluating novel facial images can be influenced by pairing individual faces with different types of stimuli. Studies using sounds (Jones et al., 2007), behavioral descriptions (Kocsor and Bereczkei, 2016; Verosky and Todorov, 2010, 2013), or perithreshold images (Kocsor and Bereczkei, 2017), found that people carry the associations they made for one facial image to novel faces that share traits with previously presented faces. It was proposed that the valence of the associated stimuli is generalized to other objects that have similar physical characteristics. In this way, repeated pairings during the experiments lead to the formation of facial prototypes to which unfamiliar faces can be compared at later encounters. In other words, the consequence of the repeated associations is that there will be a shift in what people expect from someone with a particular facial traits. These results echo the findings of other experiments using evaluative conditioning paradigms as a means to change attitudes towards others (e.g., FeldmanHall et al., 2018; Putz et al., 2018; Walther et al., 2005; Walther et al., 2011), and studies measuring face preferences in relation to long term interactions with personally significant people, such as parents (Bereczkei et al., 2002; Bereczkei et al., 2004; Kocsor et al., 2013; Kocsor et al., 2016) or partners (Günaydin et al., 2012).

The associative process through which evaluations of individual faces shift, and expectations about physical appearance of people with characteristic behavior are formed, are best understood within the model of Trait Inference Mapping (TIM) (Over and Cook, 2018). This model suggests that after encountering unfamiliar people, their facial features will be represented as vectors in the *face space*. When knowledge about their behavior accumulates through first hand experiences either by direct interactions, or indirectly by observing reactions of others or hearing behavior-relevant information, locations in the *trait space* will be mapped onto the representations in the face space. Thereby specific face-trait mappings emerge. On a longer time scale, non-specific face-trait contingencies, that is mappings between regions of the two representational spaces, lead to the emergence of "face types" that evoke stereotypical judgments (Over & Cook).

One might argue that if the TIM model is a valid theoretical framework and evaluations related to static facial features can be explained by a contingency between points and regions of the face and the trait space, a similar mapping between dynamic features and behavioral traits may

* Corresponding author.
E-mail address: kozma.luca@gmail.com (L. Kozma).

https://doi.org/10.1016/j.heliyon.2019.e01736
Received 17 July 2018; Received in revised form 7 March 2019; Accepted 10 May 2019

also occur. Traditionally, facial expressions were seen as an innate set of behavioral patterns that evolved to signal intentions and inner states to others (e.g., Horstmann, 2003, cf. Fridlund, 1991). Likewise, expression-recognition is also assumed to be automatic (e.g., Dimberg et al., 2000) with developmental origins in early life and some progress even in adolescence (Camras and Allison, 1985; Herba and Phillips, 2004; Kessels et al., 2014; McClure, 2000; Nelson, 1987; Widen, 2013). In this sense, the emotions that are evoked when people observe contractions of certain facial muscles rely on long-term personal experiences. In light of the aforementioned, the main question to be answered is whether the evaluation of facial expressions could be changed with the same types of stimuli that were effectively used in previous experiments to modify attitudes towards others. In the current study we tested, first, how behavioral descriptions affect the rating of facial expressions and, second, whether the shift in the ratings is generalized to other individuals showing the same expression. If generalization happens, it would support the view that the evaluation of facial expressions is partly built on the same low-level associative processes as that of invariant facial traits, which would also support the TIM model as a good theoretical model for explaining impression formation.

To this end, we designed an experiment where participants were exposed to a set of individual facial images showing various expressions of three possible categories (trustworthy, untrustworthy, and ambiguous). These images were presented along descriptive sentences of different behaviors that were either pro- or antisocial. Participants were exposed to two conditions. In the (1) *inconsistent* condition the set of facial expressions were easy to recognize as trustworthy (e.g., a smile) or untrustworthy (e.g., an angry face), but the descriptors were reversed such that prosocial actions were ascribed to the untrustworthy face and *vice versa*. In the (2) *consistent* condition the set of facial expressions were ambiguous (arbitrary grimaces with no clear social meaning) and the social descriptors were either pro- or antisocial. To be accurate, consistent condition was not entirely consistent, as neutral images were presented with negative or positive descriptions. However, they were not as inconsistent as the other conditions. We decided to use this wording for simplicity.

The experiment proceeded in two phases, an *Association* phase and an *Evaluation* phase. In the *Association* phase participants saw one set of emotional facial images and social descriptions. In the *Evaluation* phase, the participants were shown a different set of images depicting a combination of the same individuals from the *Association* phase with neutral (relaxed) faces, as well as novel faces with similar expressions to those in the *Association* phase (i.e., trustworthy, untrustworthy, ambiguous). The pairing of the images and descriptions was arranged in a manner that facilitated the association of the descriptions with the expressions, rather than with the individual facial traits (see *Methods*). Our primary aim was to investigate the generalization of the affective valence of social descriptions across individuals as well as facial expressions. We also set out to reveal which source of information participants are more willing to rely on: descriptions referring to previous social behaviors or facial expressions signaling current intentions. The following hypotheses have been created (phrases in brackets refer to the images the ratings of which particularly should be analyzed, see Table 2):

1. Valenced social descriptions associated to emotional facial images will influence the evaluation of these images (learned, emotional faces).
2. The affective valence of the descriptions would be transferred to the individual faces, influencing their evaluation even when neutral photographs have to be rated (learned, neutral faces.
3. The valence of the descriptions would be generalized to the particular expression and transferred across individuals (unfamiliar emotional faces).
4. These effects would be stronger in the consistent condition when the expressions are ambiguous, compared to the inconsistent condition (interaction with the factor *consistency*).
5. The ratings of neutral, unfamiliar faces would not be statistically different from each other (unfamiliar, neutral faces, i.e., control measure).

2. Methods

2.1. Participants

Four groups of undergraduates from a Hungarian university partook in the study, each one of them had different tasks. First, images from the *Facial Action Coding System* (Ekman et al., 2002) were presented to 33 independent raters (Group 1, 5 men; age: M = 20.2; SD = 0.96; 19–23 years) who judged the trustworthiness of these faces. The second group consisted of 20 men who volunteered as photo subjects to create the stimuli set, and facial images of 16 of those men (Group 2, age: M = 21.3; SD = 2.5; 18–27 years) were used in the experimental part. They provided written informed consent and agreed to their images being used as stimuli and in scientific publications. Another group of independent raters (Group 3, 30 people, 6 men; age: M = 20.23; SD = 0.73; 19–22 years) was asked to judge the trustworthiness of the facial images of Group 2. A total of 93 people (Group 4, 33 men, age: M = 29.95; SD = 13.74, 18–67 years) participated in the experimental part of our study. Our study had been approved by the Hungarian United Ethical Review Committee for Research in Psychology (approval number 2015/26).

2.2. Stimuli

To create the image pool, we presented 39 photos taken from Ekman's *Facial Action Coding System* (Ekman et al., 2002) to Group 1 who judged how trustworthy the person on each picture was. Based on their ratings, we could divide the pictures into three groups: trustworthy, untrustworthy and ambiguous. Out of the 39 Ekman photos, we chose 16 (5 trustworthy, 5 untrustworthy and 6 ambiguous, Table 1) and asked 20 male volunteers (members of Group 2) to copy these expressions. We photographed them with a Canon EOS 700D digital camera equipped with 100 mm fixed portrait lenses under standard lighting conditions before a non-reflecting white background. Each volunteer provided one picture mimicking each of the 16 Ekman photos plus a neutral one where they showed no expression, resulting in a total of 340 facial images. Members of Group 3 were asked to judge how trustworthy the 20 men were based on their neutral photo. A total of four men were eliminated from the stimuli set, one who was rated noticeably more trustworthy than average and 3 others because of poor picture quality.

However, upon subjective evaluation by the experiment leaders it was apparent that not all expressions could not be mimicked by participants with equal success. From all of the photographed expressions we have chosen four that appeared to be the easiest to be mimicked, that is, on all of our photographs they were indistinguishable from the original FACS faces. In the end, we had 16 men showing 4 expressions each: 1 trustworthy, 1 untrustworthy and 2 ambiguous (Fig. 1). Out of those expressions that most of our participants could mimic accurately, the one that raters judged most trustworthy was AU13 from FACS (Ekman et al., 2002). To perform this expression, the manual's instructions read: “Try to pull the inner corners of your lips straight up without letting yourself smile” (Ekman et al., 2002). In other publications, it is also called a “non-enjoyment display” that smiles may contain, or “listener smile” that signals involvement in the conversation (Bousmalis et al., 2009; Ruch, 2005). The most untrustworthy expression was one that mobilizes action units 10 and 25. These are components of the expression of anger and disgust (Wiggers, 1982; Rozin et al., 1999; Pantic and Rothkrantz, 2000). It is possible that our participants associated these emotions to the expression hence why they labeled it untrustworthy.

Using the 16 individuals' photos we created 24 picture-description pairs (see top 2 pictures in Fig. 2). Eight men showing ambiguous expressions were paired with both negative and positive descriptions – resulting in 16 sets. The other 8 men were divided into two groups: 4 of

Table 1
Descriptive statistics of ratings of faces from the FACS database. Rows in bold indicate the expressions which were used in the experimental part of the study.

	FACS database file names	FACS action unit codes	Mean score of trustworthiness	SD
Trustworthy faces	s6_12z26	6D+7C+12E+25D+26C	6.303	2.039
	s6121517	6E+7D+12+15B+17D	5.546	1.769
	s13a	**13B**	**5.485**	**1.788**
	s12x_23	12B+23D+38A	5.576	1.751
	sL14	L14C	5.788	1.833
Untrustworthy faces	s10y_17	10E+17D	2.212	1.431
	sL10x_25	L10B+25B	2.606	1.580
	s10y1625	**10C+16E+25E**	**2.394**	**1.273**
	s9_25	7C+9E+25C	2.061	1.144
	s10y2325	10C+16A+23E+25E	2.121	.992
Ambiguous faces	s20z	G20E+21B	4.091	1.684
	s6_15z17	**6D+G7D+15E+17E+38B**	**4.212**	**1.933**
	s17_24	17D+24D	3.758	1.542
	s4b	4D	4.212	2.043
	s2	**V2C+38A**	**4.182**	**1.310**
	s25	25B	3.546	1.787

Table 2
Number of trials and image types of the experimental conditions as presented in the *Evaluation* phase. Note that though this was a full factorial design, the factors *valence* and *consistency* are not shown here.

Expression type	*Novelty of the presented individual*	*Number of presented individuals*	*Number of trials (64 in total)*
emotional	learned	8 men	2 × 8
neutral	learned		2 × 8
emotional	unfamiliar	8 men	2 × 8
neutral	unfamiliar		2 × 8

them showing the aforementioned "listener smile" paired with negative descriptions; and another 4 men displaying the expression that mobilized action units used in anger and disgust – these images were paired with positive descriptions. We took negative and positive social descriptions from a previous study (Kocsor and Bereczkei, 2016). For the inconsistent condition, each untrustworthy expression (muscle movement akin to anger and disgust) was paired with two positive descriptions and the trustworthy expressions ("listener smile") with two negative sentences. For the consistent condition, one of the individual faces with ambiguous expressions was paired with positive, the other with negative descriptions.

2.3. Procedure

First, in the *Association* phase, the participants were asked to memorize eight randomly selected face and social description pairs. In the *inconsistent condition* they saw 2 men showing the same trustworthy expression while negative descriptions were presented on the screen. Another 2 men mimicked an untrustworthy expression that was accompanied by positive social descriptions. That is, 4 pictures were shown in the inconsistent condition.

In the *consistent condition* 4 men were seen with 2 ambiguous expressions, 8 pictures in total. Two men displayed one of these ambiguous expressions which was presented with negative descriptions, while the other expression, presented by other 2 men, was seen with positive descriptions. To half of the participants we showed these 4 men with the mentioned descriptions, while to the other half of the participants we switched the valence of the descriptions – men that were seen with negative description by half of the participants were seen with positive descriptions by the other half. Presenting images like this helped counterbalancing the trials. In the two conditions together, participants saw 8 men in the *Association* phase. These pictures were presented five times in five blocks, in random order within each block. To maintain the attention of the subjects and enhance their focus on the task, after the third block, they were asked to decide whether they find the individuals on the pictures trustworthy. Then the *Association* phase continued with the remaining two blocks.

After that, in the *Evaluation* phase, participants were asked to judge the trustworthiness of a new set of pictures on a 9-point Likert-scale. In this phase, we showed 32 pictures – 8 men that participants have seen before and 8 novel individuals. Emotional images from the *Association* phase were repeated *and* those 8 men reappeared with neutral faces as well, meaning 2 photos per man (8 emotional + 8 neutral images, 16 pictures in total). The 8 new individuals were shown with neutral faces *and* displaying expressions that have already been seen in the *Association phase* (8 emotional + 8 neutral images, 16 pictures in total). Each of the 32 pictures were presented twice, in random order (i.e., 64 trials altogether; see Table 2 for factorial design, and Fig. 2 for visual presentation).

The trials were counterbalanced in many respects. First, men who were shown with trustworthy expressions and negative descriptions to some subjects were shown with untrustworthy expressions and positive descriptions to other subjects, and *vice versa*. Second, ambiguous

Fig. 1. Four different expressions (from left to right, with action unit codes): one trustworthy (AU13B), one untrustworthy (10C+16E+25E) and two ambiguous expressions (6D+G7D+15E+17E+38B and V2C+38A).

Fig. 2. Top row: Picture shown in the *Association* task. Centre left: *Evaluation* task – a man with the expression previously shown. Centre right: Previously seen man with neutral face. Bottom left: Picture shown in the *Evaluation* task – previously shown expression on an unfamiliar person. Bottom right: neutral face of the unfamiliar person.

expressions varied with respect to whether they were paired with positive or negative descriptions. Third, only 8 of the 16 individuals were presented in the *Association* phase to any subject, allowing the other 8 to be presented as test faces in the *Evaluation* phase, in this way we were able to create alternative presentation scripts for an approximately equal number of subjects.

2.4. Data processing

As each expression was presented by two men, and each image was shown twice, we used the mean of these four scores given to these individuals in the test phase (see Supplementary Material "*dataset_arbitrary signals.xlsx*" for all aggregate scores). For simplicity, when we refer to individual faces in the forthcoming parts of the paper, we mean these average scores.

3. Results

We conducted a repeated measures ANOVA with four within-subject factors, each with two levels (see Table 2): 2 (*novelty*: the faces were either learned in the first phase of the experiment along with the descriptions, or they were unfamiliar stimuli) × 2 (*expression type*: faces either showing an expression or were neutral) × 2 (*consistency*: consistent or inconsistent) × 2 (*valence*: the faces were previously presented either with positive or negative descriptions). We added *sex* as a between-subject factor.

The analysis shows that the scores given by participants were significantly influenced by expression type ($F = 8.958$, $p = .004$, $\eta^2_p = .100$) and by the valence of the description ($F = 10.006$, $p = .002$, $\eta^2_p = .110$). There was also a significant two-way interaction between consistency and valence ($F = 15.951$, $p < .001$, $\eta^2_p = .165$), and a three-way interaction between expression type, consistency, and valence ($F = 16.718$, $p < .001$, $\eta^2_p = .171$). Neither the third within-subject factor, consistency, nor other interactions and the participants' sex had significant effects on the ratings (all p's > 0.05). However, the global analysis of the experimental factors does not show the effects separately for the image types, which would be critical to evaluate whether the results support the hypotheses. Therefore, the analysis was divided into four measurements according to *novelty* and *expression type* (2 × 2 ANOVA, see Table 3), using only *consistency* and *valence* as within-subject factors, and *sex* as a between-subject factor.

The results indicate a significant main effect of valence for the learned neutral (Fig. 3) and learned emotional (Fig. 4), and for the unfamiliar emotional images (Fig. 5), and a significant interaction between consistency and valence for the learned neutral (Fig. 3) and unfamiliar neutral

Table 3
Results of the 2 × 2 repeated measures ANOVA, grouped by novelty and expression type of the stimuli images.

Variables and interactions	Measures	F	*p*	η^2_p
consistency	Learned, neutral	0.253	.616	.003
	Learned, emotional	1.260	.265	.015
	Unfamiliar, neutral	2.559	.114	.031
	Unfamiliar, emotional	1.668	.200	.020
consistency × sex	Learned, neutral	1.312	.255	.016
	Learned, emotional	0.012	.914	.000
	Unfamiliar, neutral	0.199	.657	.002
	Unfamiliar, emotional	0.000	.986	.000
valence	**Learned, neutral**	**7.535**	**.007***	**.085**
	Learned, emotional	**8.816**	**.004***	**.098**
	Unfamiliar, neutral	0.834	.364	.010
	Unfamiliar, emotional	**4.116**	**.046***	**.048**
valence × sex	Learned, neutral	0.063	.803	.001
	Learned, emotional	0.014	.906	.000
	Unfamiliar, neutral	0.001	.981	.000
	Unfamiliar, emotional	0.637	.427	.008
consistency × valence	**Learned, neutral**	**21.621**	**< .001****	**.211**
	Learned, emotional	0.472	.494	.006
	Unfamiliar, neutral	**13.054**	**.001***	**.139**
	Unfamiliar, emotional	0.542	.464	.007
consistency × valence × sex	Learned, neutral	1.533	.219	.019
	Learned, emotional	1.541	.218	.019
	Unfamiliar, neutral	2.339	.130	.028
	Unfamiliar, emotional	0.035	.853	.000
sex	Learned, neutral	0.132	.717	.002
	Learned, emotional	0.152	.698	.002
	Unfamiliar, neutral	0.586	.446	.007
	Unfamiliar, emotional	0.314	.577	.004

Note: All significant main effects indicate higher scores for the positively valenced images, and all significant interactions indicate larger difference between the scores of positively and negatively valenced images in the not inconsistent condition.

* Effects are significant on a $p < .05$ significance level.

** Effects are significant on a $p < .001$ significance level.

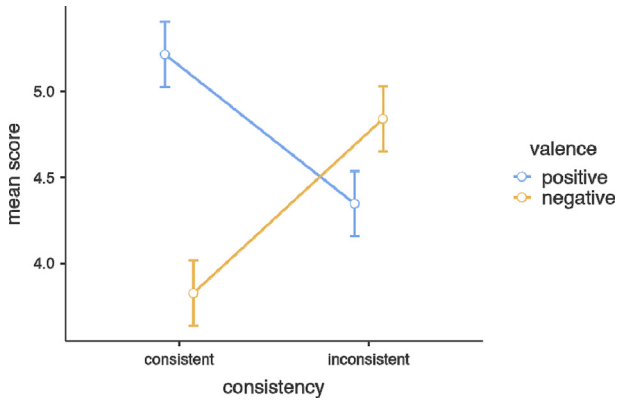

Fig. 3. Scoring of learned neutral images.

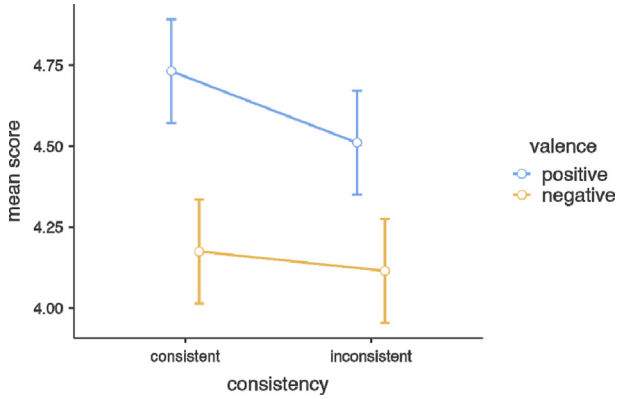

Fig. 4. Scoring of learned emotional images.

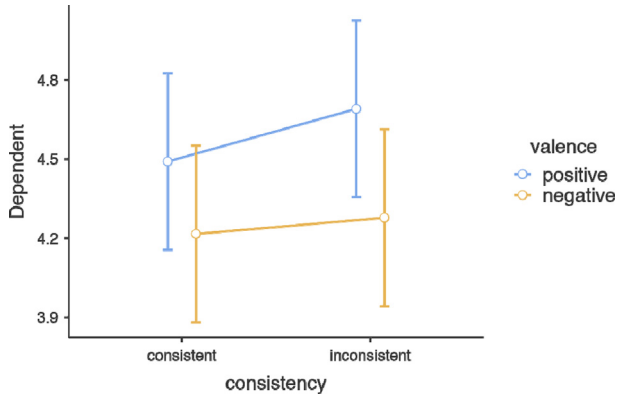

Fig. 5. Scoring of unfamiliar emotional images.

(Fig. 6) images. These interactions suggest that the mean score differences between positively and negatively valenced faces are higher in the consistent condition (PN: M = 1.440, SD = 2.433; UN: 1.170, SD = 2.686) than in the inconsistent condition (PN: M = -0.566, SD = 2.131; UN: -0.955, SD = 2.558). There were no main effects or interactions for participants' sex.

The findings support Hypothesis 4 that participants were more likely to transfer the valence of the descriptions to unfamiliar neutral faces when the expression presented in the *Association* phase was ambiguous, than when it showed cues of trustworthiness or untrustworthiness. However, with Hypothesis 4 we also conjectured that the effect of generalization would be stronger in the condition where the expression and the description were consistent, and both for emotional and neutral

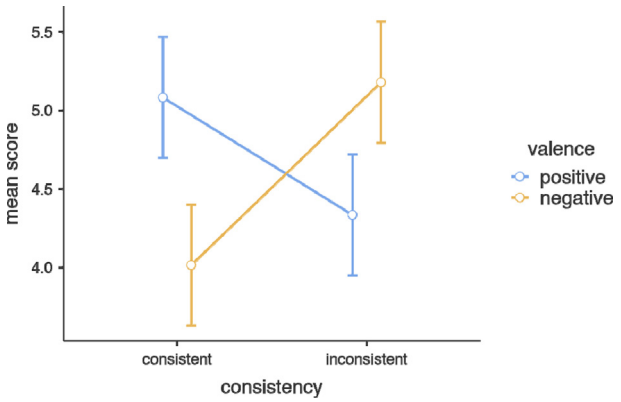

Fig. 6. Scoring of unfamiliar neutral images.

faces. As the data supported the hypothesis only for neutral faces, we ran paired-samples t-tests to double-check the lack of significant interactions in the ANOVA for the learned emotional and unfamiliar emotional images. The analysis confirmed that the score differences between the differently valenced faces were the same, irrespective of whether the facial expression was ambiguous (i.e., consistent condition) or easily recognizable (i.e., inconsistent condition), both for the learned emotional ($t = 0.954$, $df = 83$, $p = 0.341$) and the unfamiliar emotional images ($t =$ -0,700, $df = 83$, $p = 0.486$). This result means that the subjects learned the associations between descriptions and expressions and transferred this knowledge to unfamiliar faces that showed the same expressions. Unlike the transfer of valence to neutral faces, this was not influenced by the consistency between descriptions and faces.

4. Discussion

4.1. Global effects of the experimental factors

With the present study we investigated how descriptions with negative or positive content influence the social evaluation of facial expressions. Particularly, we focused on how social evaluation will be transferred to neutral facial images and generalized to the expression across individuals.

A full factorial analysis of variance showed that *expression type*, that is whether emotional or neutral photographs of the faces were presented, influenced the ratings significantly. More importantly, *valence* had also a significant effect, highlighting that the social information presented in the *Association* task had a global effect on the evaluation of the facial stimuli. Furthermore, the interaction between consistency and valence suggests that the facial expressions, which were categorized by independent raters as trustworthy or untrustworthy, were more resistant to valence transference than ambiguous expressions. This would support our fourth hypothesis; however, further analyses showed that this effect was due to the ratings of neutral faces.

4.2. Manipulation of trustworthiness ratings with descriptions – analysis of learned emotional faces

The second ANOVA showed a more detailed picture about how the particular factors affected ratings. The first hypothesis was that associated descriptions influence the trustworthiness of learned emotional faces, and this has been confirmed. Interestingly, and in contrast with what we had expected, consistency did not significantly interact with valence. This means that the participants have evaluated expressions, which were previously rated as trustworthy by independent judges, just as untrustworthy as an ambiguous grimace if the description suggested socially undesirable behavior. However, we would like to highlight that the effect of descriptions was only marginally significant when totally

new faces were shown with expressions (unfamiliar emotional faces). This means that learned information only moderately affected the evaluation of unfamiliar faces.

4.3. Generalization of trustworthiness ratings to individual facial trait – analysis of learned neutral faces

Supporting the second hypothesis, participants' evaluations were also transferred to the individual faces outside of the original context: when the neutral photographs of learned stimuli faces – i.e., faces that have been presented in the *Association* phase – were shown in the *Evaluation* phase, the trustworthiness ratings reflected the previously presented descriptions. That is, images of men paired with descriptions of trustworthiness scored higher, when averaged over the two conditions. However, consistency significantly interacted with valence, ambiguity of the expression being a promoter of the transfer of valence (see Fig. 3). This suggests that, though participants generalized the valence of the associated written information to the individual facial traits, the same also happened with the social desirability of the originally presented expression. Therefore, ratings of neutral faces from the inconsistent condition show no difference, reflecting that neither the valence of the expression, nor that of the descriptions had a significant effect. In other words, people make predictions about what can be expected from others, facial expression being one source of information for predicting intentions and subsequent behavior, and available social information being another.

Facial expressions signal both emotional states and intentions (e.g., Horstmann, 2003), which guide our behavior along with other relevant social information, such as knowledge about formerly observed acts. If these two sources of information contradict each other, as in the inconsistent condition of this study, the predictions will be uncertain. However, from our data we could hardly make firm conclusions about which source of information is more influential for trustworthiness judgments in a real-world situation. According to the presumed reliability of the information in the particular situation, people are likely to rely to some extent both on knowledge about previous social behaviors, and on facial expressions signaling current intentions.

4.4. Generalization of the valence of an expression – analysis of unfamiliar emotional faces

The third hypothesis was that facial expressions may be bestowed with affective valence values, and this might be independent of the individual facial traits. It has been already shown that not only the resemblance of static and invariant facial traits can mediate the transfer of valence across individual faces. With training, as evidenced by clinical practice, the perception of ambiguous facial expressions depicting emotions can be altered (e.g., Ekman and Friesen, 1976; Penton-Voak et al., 2012; Tottenham et al., 2009), and this effect generalizes across individuals (Dalili et al., 2016). In the present study we replicated this effect. In addition, as our results indicated no significant interaction between valence and consistency for the unfamiliar emotional faces, the recent findings go beyond the former results. These suggest that even in that case when the facial expression is close to a real, socially meaningful expression, for instance it resembles expressions of anger or happiness, the social desirability of the expression can be manipulated. As a note of caution, we would like to highlight that the depicted expressions were copied from static expressions from the FACS database (Ekman et al., 2002). Genuine, spontaneous facial expressions are likely to evoke stronger feelings in observers and might be more easily categorized as trustworthy or untrustworthy. Yet, it might be possible that humans' adherence to their assumptions about what behavior is expected when a well-known, basic expression appears on a face, might be much less rigidly ossified during childhood and adolescence than what we intuitively presume in our everyday life, and what has been suggested in many studies (e.g., McClure, 2000).

4.5. Control measures and limitations – analysis of unfamiliar neutral faces

The full factorial experimental design allowed us to test whether participants' ratings of facial images were influenced by the *a priori* trustworthiness of these faces. We expected that the unfamiliar neutral images would be rated equally in each condition, irrespective of valence and consistency. The reason for this was that these unfamiliar neutral faces have not been shown in the *Association* phase, so participants did not have any previous experience with them, and they did not depict any expression that could influence the ratings. This hypothesis (Hypothesis 5) has been only partially supported. Though valence did not influence the ratings globally, it significantly interacted with consistency. This means that in the consistent condition the mean score difference between the faces with positive and negative descriptions was higher than in the inconsistent condition. To highlight it again, the unfamiliar neutral faces were not shown in the *Association* phase, but during the *Evaluation* phase both neutral images and emotional images of the same men (i.e., with a familiar expression that was shown in the *Association* phase on a different individual) were shown. One explanation might be that the partial correspondence between the ratings of the learned emotional and neutral, and the unfamiliar emotional and neutral images was caused by the coincidence that the faces which appeared with positive descriptions were, by mere chance, more trustworthy than those with negative description. Hence, the results may reflect this difference rather than the effect of the experimental manipulation.

However, there are several reasons to think that this explanation is unlikely. First, the faces we used were rated by independent judges prior to the main experiment, and the only face with an extreme trustworthiness score was dropped from the image pool. Hence, the individual faces had similar trustworthiness scores. Second, the counterbalanced arrangement of the stimuli individuals across valence and consistency (we used 4 different presentation scripts to show the photos) very likely eliminated this potentially confounding effect. The most probable explanation for the results of the unfamiliar neutral images is that it is an artifact of the randomized appearance of the images. Namely, as the test images appeared on the screen in a random order, the ratings of the four image types were not fully independent. As we used photographs of two persons for each image type, and each individual image appeared twice, it may well have happened for several participants that the unfamiliar emotional images preceded the unfamiliar neutral faces. In this case the valence value of an expression – which has already been shifted in the *Association* phase – might have been transferred to the neutral image of the same individual. Though due to the randomization the chance that each of the eight unfamiliar faces with a previously seen expression appeared before their neutral counterparts is pretty low (appr. 0.34%), the likelihood that a single unfamiliar emotional face appeared before its neutral version is quite high (appr. 49.2%). This already might have distorted the ratings of the unfamiliar and neutral faces. To avoid this kind of confusion, in future studies it might be useful to present stimuli in a different order, for instance in a block design in which emotional faces precede all neutral faces.

5. Conclusions

With this study we set out to find further support for the assumption that social evaluation of unfamiliar people relies on the generalization of the affective valence associated to facial traits. The results correspond to the predictions of the TIM model (Over and Cook, 2018) as well, namely that information about the expected behavior (represented in the trait space) will be mapped onto representations in the face space. Beyond invariant facial features we extended this assumption to facial expressions. Most of our hypotheses have been confirmed. The affective valence of social descriptions was transferred both to individual faces and expressions. These results may not be surprising given that – in contrary to what was assumed by early models of face perception (e.g., Bruce and

Young, 1986) and empirical research (e.g., Bobes et al., 2000; Humphreys et al., 1993) – cortical areas responsible for the recognition of invariant and dynamically changing facial traits are not fully independent (Lander and Butcher, 2015; Rhodes et al., 2015). Therefore, the overlap in the neural structures of the recognition of facial expressions and static facial traits makes them likely to be affected by the same general cognitive processes. More specifically, during real-life interactions and in experimental settings, if representations of either invariant or dynamic facial features are associated with personality traits, they provide space to face-trait mappings to the same extent. Using these two types of facial information for modeling behavioral outcomes enables a more accurate and flexible estimation of intentions. Indeed, this is what the results of the current experiment also suggest: valence of socially relevant visual cues (i.e., facial expressions) were transferred to familiar faces that presently show no expression (generalization to individuals across contexts), and to unfamiliar faces with expressions (generalization to expressions across individuals).

Surprisingly, the effect of valence transference to unfamiliar emotional faces was not stronger in the consistent condition than in the inconsistent condition – when social descriptions were in line with the presented facial expression vs. when they were not. Because across their lives people are extensively exposed to faces and expressions, and their personal experience endows these with various levels of trustworthiness, these interactions were expected to influence people's decisions more deeply than a one-time social description linked to this face. In our study this was not the case. However, as a note of caution we would like to highlight that the effect of valence (i.e., pairing with either positive of negative descriptions) was only marginally significant with a weak effect size.

Though expressions signaling basic emotions like anger or happiness may be more easily and rapidly recognized than ambiguous facial configurations, the cognitive apparatus responsible for detecting socially relevant information seems to be very flexible. The same was suggested in a study by Heerey and Velani (2010), who showed that even arbitrary behavioral cues, such as subtle muscle contractions on the face, can be used as socially predictive signals. The current results echo this finding.

Declarations

Author contribution statement

Ferenc Kocsor: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Luca Kozma: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Adon L Neria, Daniel N Jones, Tamas Bereczkei: Analyzed and interpreted the data; Wrote the paper.

Funding statement

This work was supported by the Hungarian Scientific Research Fund, Hungary (grant number OTKA K112673), and the European Social Fund, European Union (EFOP-3.6.1.-16-2016-00004 – Comprehensive Development for Implementing Smart Specialization Strategies at the University of Pécs) and Institutional Excellence Grant, University of Pécs, Hungary (17886-4/2018 FEKUTSTRAT). Ferenc Kocsor received funding from the National Excellence Program of the Ministry of Human Capacities, Hungary (ÚNKP-17-4 -I.-PTE-298).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01736.

References

Bereczkei, T., Gyuris, P., Weisfeld, G.E., 2004. Sexual imprinting in human mate choice. Proc. Biol. Sci. 271 (1544), 1129–1134.
Bereczkei, T., Gyuris, P., Koves, P., Bernath, L., 2002. Homogamy, genetic similarity, and imprinting; parental influence on mate choice preferences. Pers. Indiv. Differ. 33 (5), 677–690.
Bobes, M.A., Martín, M., Olivares, E., Valdés-Sosa, M., 2000. Different scalp topography of brain potentials related to expression and identity matching of faces. Cogn. Brain Res. 9 (3), 249–260.
Bousmalis, K., Mehu, M., Pantic, M., 2009. Spotting agreement and disagreement: a survey of nonverbal audiovisual cues and tools. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–9.
Bruce, V., Young, A., 1986. Understanding face recognition. Br. J. Psychol. 77 (3), 305–327.
Camras, L.A., Allison, K., 1985. Children's understanding of emotional facial expressions and verbal labels. J. Nonverbal Behav. 9 (2), 84–94.
Dalili, M.N., Schofield-Toloza, L., Munafò, M.R., Penton-Voak, I.S., 2016. Emotion recognition training using composite faces generalises across identities but not all emotions. Cognit. Emot. 0 (0), 1–10.
Dimberg, U., Thunberg, M., Elmehed, K., 2000. Unconscious facial reactions to emotional facial expressions. Psychol. Sci. 11 (1), 86–89.
Ekman, P., Friesen, W.V., 1976. Pictures of Facial Affect. Consulting psychologists Press.
Ekman, P., Friesen, W.V., Hager, J.C., 2002. Facial Action Coding System: the Investigator's Guide. Research Nexus, Salt Lake City, UT.
FeldmanHall, O., Dunsmoor, J.E., Tompary, A., Hunter, L.E., Todorov, A., Phelps, E.A., 2018. Stimulus generalization as a mechanism for learning to trust. Proc. Natl. Acad. Sci. Unit. States Am. 201715227.
Fridlund, A.J., 1991. Evolution and facial action in reflex, social motive, and paralanguage. Biol. Psychol. 32 (1), 3–100.
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. J. Exp. Soc. Psychol. 48 (1), 350–353.
Heerey, E.A., Velani, H., 2010. Implicit learning of social predictions. J. Exp. Soc. Psychol. 46 (3), 577–581.
Herba, C., Phillips, M., 2004. Annotation: development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. JCPP J. Child Psychol. Psychiatry 45 (7), 1185–1198.
Horstmann, G., 2003. What do facial expressions convey: feeling states, behavioral intentions, or actions requests? Emotion 3 (2), 150–166.
Humphreys, G.W., Donnelly, N., Riddoch, M.J., 1993. Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. Neuropsychologia 31 (2), 173–181.
Jones, B.C., Debruine, L.M., Little, A.C., Feinberg, D.R., 2007. The valence of experiences with faces influences generalized preferences. J. Evol. Psychol. 5 (1), 119–129.
Kessels, R.P.C., Montagne, B., Hendriks, A.W., Perrett, D.I., de Haan, E.H.F., 2014. Assessment of perception of morphed facial expressions using the Emotion Recognition Task: normative data from healthy participants aged 8–75. J. Neuropsychol. 8 (1), 75–93.
Kocsor, F., Bereczkei, T., 2016. First impressions of strangers rely on generalization of behavioral traits associated with previously seen facial features. Curr. Psychol.
Kocsor, F., Bereczkei, T., 2017. Evaluative conditioning leads to differences in the social evaluation of prototypical faces. Pers. Indiv. Differ. 104, 215–219.
Kocsor, F., Gyuris, P., Bereczkei, T., 2013. The impact of attachment on preschool children's preference for parent-resembling faces — a possible link to sexual imprinting. J. Evol. Psychol. 11 (4), 171–183.
Kocsor, F., Saxton, T.K., Láng, A., Bereczkei, T., 2016. Preference for faces resembling opposite-sex parents is moderated by emotional closeness in childhood. Pers. Indiv. Differ. 96, 23–27.
Lander, K., Butcher, N., 2015. Independence of face identity and expression processing: exploring the role of motion. Front. Psychol. 6.
McClure, E.B., 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychol. Bull. 126 (3), 424–453.
Nelson, C.A., 1987. The recognition of facial expressions in the first two years of life: mechanisms of development. Child Dev. 58 (4), 889–909.
Over, H., Cook, R., 2018. Where do spontaneous first impressions of faces come from? Cognition 170, 190–200.
Pantic, M., Rothkrantz, L.J.M., 2000. Expert system for automatic analysis of facial expressions. Image Vis Comput. 18 (11), 881–905.
Penton-Voak, I.S., Bate, H., Lewis, G., Munafò, M.R., 2012. Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. Br. J. Psychiatry 201 (1), 71–72.
Putz, Á., Kocsor, F., Bereczkei, T., 2018. Beauty stereotypes affect the generalization of behavioral traits associated with previously seen faces. Pers. Indiv. Differ. 131, 7–14.
Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., et al., 2015. How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. Cognition 142, 123–137.

Rozin, P., Lowery, L., Imada, S., Haidt, J., 1999. The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). J. Personal. Soc. Psychol. 76 (4), 574–586.

Ruch, W., 2005. Extraversion, alcohol, and enjoyment. In: Ekman, P., Rosenberg, E.L. (Eds.), What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS), 2nd ed, pp. 112–132. New York, NY, USA: Oxford University Press.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., et al., 2009. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatr. Res. 168 (3), 242–249.

Verosky, S.C., Todorov, A., 2010. Generalization of affective learning about faces to perceptually similar faces. Psychol. Sci. 21 (6), 779–785.

Verosky, S.C., Todorov, A., 2013. When physical similarity matters: mechanisms underlying affective learning generalization to the evaluation of novel faces. J. Exp. Soc. Psychol. 49 (4), 661–669.

Walther, E., Nagengast, B., Trasselli, C., 2005. Evaluative conditioning in social psychology: facts and speculations. Cognit. Emot. 19 (2), 175–196.

Walther, E., Weil, R., Düsing, J., 2011. The role of evaluative conditioning in attitude formation. Curr. Dir. Psychol. Sci. 20 (3), 192–196.

Widen, S.C., 2013. Children's interpretation of facial expressions: the long path from valence-based to specific discrete categories. Emotion Review 5 (1), 72–77.

Wiggers, M., 1982. Judgments of facial expressions of emotion predicted from facial behavior. J. Nonverbal Behav. 7 (2), 101–116.