Effects of observers' characteristics on impression formation on face

Kuangzhe Xu*† and Toshihiko Matsuka†‡
* Graduate School of Science and Engineering, Chiba University, Chiba, Japan
E-mail: jo@chiba-u.jp Tel/Fax: +81-43-290-2277
† ASCON Co. Ltd., Fukuyama, Japan
‡ Department of Cognitive and Information Science, Chiba University, Chiba, Japan
E-mail: matsuka@chiba-u.jp Tel/Fax: +81-43-290-3578

***Abstract*—The vast majority of previous research on facial impressions dealt with how individual features (e.g. shapes of eyes) and/or their configurations affect impression formation on face. These research implicitly assume that the effects are uniformly applicable to many if not all observers. But, we believe the validity of this assumption questionable. For example, a shy person, who is hesitant to look at other individuals eyes, and an outgoing person, who constantly looks at others eyes, may form different impressions to a particular face whose eyes are very characteristic, simply because those two individuals look at different areas of the face. In the present research we hypothesized that some features of observers also affect impression formation. In order to examine our hypothesis, we conducted two experiments. In Experiment 1, we examined how observers personalities and observational behaviors, recorded by an eye-tracking device, affected formed impression. Our data were analyzed with hierarchical Bayesian models. The result showed that both the observers personality and observational behaviors influences formed impressions. In order to distinguish the effects of observers personalities and observational behaviors, we induced participants to look at particular areas of faces in Experiment 2. The results indicated that observers personalities influence observational behaviors, which in turn influence impression formation on face.**

I. Introduction

The vast majority of previous research on facial impressions dealt with how individual features (e.g. shapes of eyes) and/or their configurations affect impression formation on face. For example, one study showed that the overall impression of face could be predicted by aggregating impressions of individual facial parts [1]. Todorov and his colleague [2] developed a model that generates facial computer graphics that gives particular types of impression and validated the model using realistic social situations such simulated election and job interview. In machine learning literature, a neural network model was used to predict first impression from facial images, and it showed promising results [3]. This series of research focusing on the features of facial parts and their configurations built on the basis of an implicit assumption that faces with particular features give particular sets of impression uniformly to observers.

However, the validity of this assumptions was questioned by a few studies that examined the effects of observers' characteristics on the impression on faces. For example, in one study, that used composite images of faces that were composited to vary on impression on degree of agreeableness, showed that observers personality trait affect how the images were seen. The results indicated not only facial features but also observers' personality affect how impressions on faces were formed [4]. In other study that collected data on where and how long observers looked at particular areas of faces showed that observers' observational behaviors also influences impression formation on faces [5]. In particular, they found signification relationships between the amount of time participants spent looking at mouth (and nose) and the impression rating on "intelligence" (and "extraversion").

Although this series of research focusing on the characteristics of observers clearly indicated that there are some effects of observers' characteristics on impression formation on face, only a limited number of research was conducted. In addition each of these research focused on a limited number of observers characteristic, and there is no single study that systematically compared effects of observers' personality traits and observational behaviors. Accordingly, the present research was conducted to compare effects of personality traits and observational behaviors using an eye tracking device. In particular, data were analyzed using hierarchical bayesian generalized linear mixed effect models to compare how well observers personality traits and observational behaviors accounted for impression ratings on faces.

II. Experiment 1

A. Overview

In Experiment 1, we conducted simple impression rating tasks asking participant to freely observe facial images. We recorded participants' eye movements using an eye tracking device to collect data about where and how long participants looked at particular areas of faces while observing facial images. In addition, we collected data on participants personality traits using a questionnaire. We then analyzed data to see wether participants' personality traits or observational behaviors could account their impression ratings on faces.

B. Method

1) Participants: Thirty four students from Chiba University participated in Experiment 1. Among them, there were 16

male and 18 female participants. All participants received gift certificates for participation.

2) Stimuli: We collected 50 pictures (25 male and 25 female) of East Asian faces through Internet. All pictures were taken from the front without any emotional expressions. The brightness of pictures were corrected with Photoshop. The sizes of pictures were adjusted so that the distances between the left and right eyes of all pictures were approximately equal. We, then, cropped and resized these pictures to 412 x 558 pixel.

3) Impression ratings: In the present experiment, we used the most famous set of personality traits, namely big five to follow previous studies [5]. The five personality traits were Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience.

In order to avoid the mere exposure effect and other unwanted effects, each personality trait was rated with 10 pictures that were not appeared in other rating tasks, and each picture was appeared exactly once in the present experiment.

4) Apparatus: We used Tobii T120 eye-tracker to present stimuli and collect eye movements. Experiment was controlled by TobiiStudio. To imitate a real interpersonal communication scene, the distance between the monitor and participants' heads were held at 65cm and the visual angle was set at 13-degree [9]. Participants' heads were fixed using a jaw stand.

5) Procedure: There were a total fo 50 sessions in Experiment 1. Each session started with a randomly selected question item asking participants to rate the face displayed immediately on one of the five personality traits. When participant click a mouse to confirmed the question, then a fixation marker (i.e., "+") was presented at the center of the monitor for one second, followed by a randomly selected face for 3 seconds. After seeing each face, participants were asked to rate the face on the personality trait asked at the beginning of the session using 7-point likert scale.

After completing the impression rating task, participants were asked to complete Japanese version of Ten Item Personality Inventory (TIPI) to collect participants five personality traits, namely Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience [6].

6) Data preprocessing: We analyzed the relationships among observers' personality traits, observational behaviors, and facial impression using hierarchical bayesian models, which are described in detail later. In these analyses, we used participants' eye-movement data that were collected when facial stimuli were presented. Based on previous study [8], we then counted the numbers of fixations at six areas of interest (AOI), namely, glabella, eyebrows (both left and right), eyes (both left and right), forehead, mouth and nose for each stimulus. Those six AOIs of each face was defined using Wacom's Intuos Pro PTH-660 (see Fig. 1 for example).

C. Data Analysis

We examined the effects of observers' personalities and observational behaviors on impression rating using models shown in Eqs. 1 - 3. The criterion variables (y) were ratings

Fig. 1. Areas of Interest in Experiments 1 & 2

on personality traits. Each of the five personality ratings was analyzed separately. The predictor variables were either five personality traits (Eq. 2) or fixations on six areas of interest (Eq. 3), where i,k,l indicate one particular data point, subject, and picture, respectively. We assumed that the criterion variable were distributed as Bernoulli. We treated observers' personality traits and observational behaviors (fixation on AOIs) as fixed effects and subjects (r_{ik}^{subj}) and pictures (r_{il}^{pic}) as random effects [7][8]. The models that included random subject effects accommodated subjects who systematically rate pictures more positive or negative than other without affecting the fixed effects (i.e., effects on personalities or observational behaviors). Likewise, the models with random picture effects allowed pictures that were systematically rated positively or negatively.

We assumed the numbers of fixations were distributed as Bernoulli, even though we using 7-point likert scale to collect impression ratings. This was because, as shown in Fig. 2, the distributions of all impression rating were bimodal. We interpret this bimodality as that there were two mixture of, perhaps Normal, distributions where 0 was their boundary. For a sake of simplicity, instead of treating that there as two separate distribution, we treated that one Bernoulli distribution. One class of our Bernoulli indicated negative, i.e., rated less than 4 which was the midpoint of the likert scale. Other class was non-negative (or simply positive), rate more than or equal to 4.

We used Rstan [9] for parameter estimation. The uniform prior was used for fixed effects, and weakly informative prior (gamma with $\alpha = 10$, $\beta = 10$) for random effects. We used Rstan's default settings for MCMC sampling. For each model, there were four chains each of which had 1000 warmup steps, 2000 iterations and thin factor being one. Thus there were 4000 MCMC samples for each model.

To verify whether MCMC samplings had converged, we checked $\hat{R}$ values. $\hat{R}$ values for all coefficients in all 36 models were less than 1.1, which is a typically used criterion, and we considered that our MCMC sampling had converged.

$$y_{ikl} \sim Bernoulli(q_{ikl}) \tag{1}$$

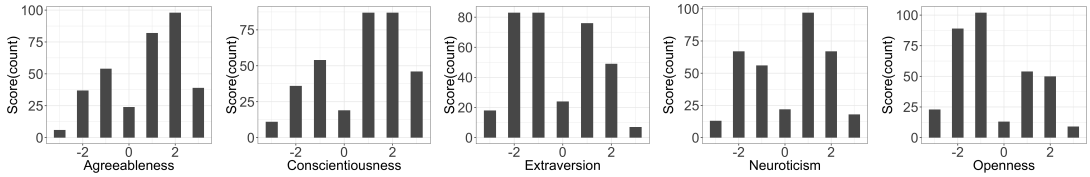

Fig. 2. Distributions of five personality rating items in Experiment 1

$$q_{ikl} = \frac{1}{1+\exp\left(-\left(b_0 + \sum_{j=1}^{5} b_j P_{ijk} + r_{ik}^{subj} + r_{il}^{pic}\right)\right)} \quad (2)$$

$$q_{ikl} = \frac{1}{1+\exp\left(-\left(b_0 + \sum_{j=1}^{6} b_j F_{ijk} + r_{ik}^{subj} + r_{il}^{pic}\right)\right)} \quad (3)$$

D. Result

Heat-maps and violin plots fixations are shown in Fig 3. In all conditions, participants tended to looked at either eyes or nose across all rating items.

1) Model Comparison: After confirming that all parameters estimations converged for all models, we calculated widely applicable information criterion or WAIC for all models. A WAIC value indicates the relative quality of statistical model, accounting for both model's fit and complexity. A model with a smaller WAIC is better than one with a higher WAIC. Table I shows WAIC values for all 10 models (5 personality rating items by 2 sets of predictors). For all personality rating items, models that incorporated observational behaviors (i.e, fixation on AOIs) resulted in better fits. None of the coefficients in models with observers personality traits were significant (i.e., 95% Highest Density Interval, or HDI included 0). The HDI indicates which points of a distribution are most credible. Thus, the HDI specifies an interval that spans most of the distribution such that every point inside the interval has higher credibility than any point outside the interval.

Table II shows significant predictor variables for models with observational behaviors. There was no significant predictors for extraversion and openness. Higher frequencies of fixation on the eyebrows and forehead were associated with lower ratings on conscientiousness. It should be noted that this particular significant relationship might have been caused by a particular set of faces we used in Experiment 1. Nonetheless, there seemed some relationship between fixations on particular areas of the face and impression rating on conscientiousness. Similarly, we found significant relationship between fixation on glabella and impression ratings on neurosis.

TABLE I
WAIC VALUES IN EXPERIMENT 1

Impression Rating	Personality	Obs. Behaviors
Agreeableness	318.861	**277.656**
Conscientiousness	305.978	**205.427**
Extraversion	321.056	**246.739**
Neurosis	421.525	**340.624**
Openness	313.952	**249.888**

TABLE II
SIGNIFICANT PREDICTORS IN EXPERIMENT 1

Impression Rating	Predictor	Mean	95% HDI
Agreeableness	Intercept	1.376	0.225 - 2.542
	Intercept	2.314	0.871 - 3.867
Conscientiousness	Eyebrows	-0.080	-0.138 - -0.026
	Forehead	-0.051	-0.100 - -0.003
Neurosis	Glabella	-0.048	-0.083 - -0.017

E. Discussion

The results of Experiment 1 indicated that impression ratings on faces were better accounted by observational behaviors than personality traits. However, this finding needed to be interpreted carefully. Participants saw 10 faces for each personality trait, and those 10 pictures could have been observed in 10 different ways depending on faces. On the other hand, participants' personality traits remained the same while observing the 10 faces. Thus, there were more variabilities in observational behaviors than personality traits as predictor variables, and this might have been a main reason why the former resulted in better fits. In other words, the comparison of observers' observational behaviors and personality traits in Experiment 1 might have been unfair to the latter, undermining their effects on impression rating on faces.

To overcome this limitation, we conducted Experiment 2. In Experiment 2, we asked participants to look at particular areas of faces to control and limit the areas that participants looked at during impression rating tasks.

III. EXPERIMENT 2

A. Overview

In Experiment 1, we tried to compare the effects of observers' personality traits and observational behaviors on impression formation on faces while participants were able to observe facial images freely. Unrestricted observations might have caused more variabilities in observational behaviors than personality traits, which in turn might have caused the former to be a better set of predictor variables than latter, resulting in unfair comparisons. In order to make fairer comparisons, we instructed participant to look at particular areas of faces during impression rating tasks in Experiment 2. In particular, we instructed participants to look at either eyes, nose, or mouth, depending on experimental conditions. Restricted observations would weaken the effects of observational behaviors by reducing their variabilities, and would allow us to examine the

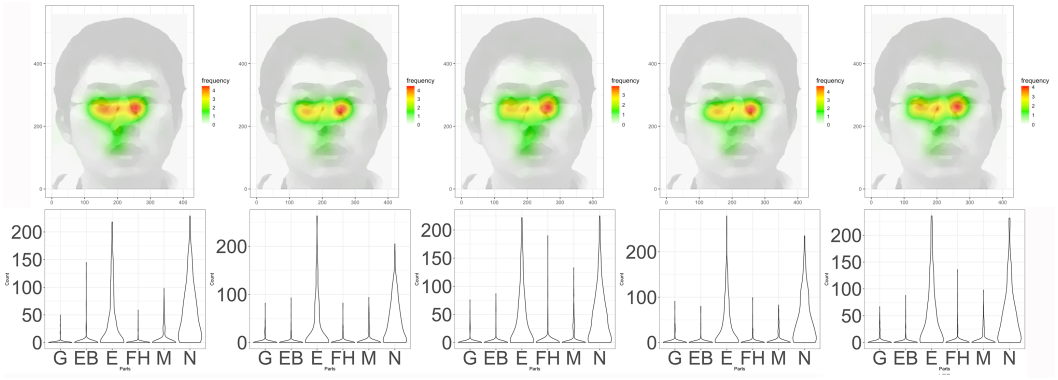

Fig. 3. Top: Heat-maps of fixation points for five personality rating items. Bottom: Violin plot of numbers of fixations on six areas of interest (G = glabella; EB = eyebrows; E = eyes; FH = forehead; M = mouth; N = nose). From left to right: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience.

effects of personality traits on impression formation on faces.

B. Method

1) Participants: In Experiment 2, 102 students from Chiba University participated. They were randomly assigned to one of the three condition. The numbers of participants in Eye, Nose, and Mouth conditions were 34 (17 male and 17 female), 34 (11 male and 23 female) and 34 (15 male and 19 female), respectively. All participants received gift certificates for participation.

2) Stimuli, rating items, apparatus: The same set of pictures of faces used in Experiment 1 was used in Experiment 2. Rating items and experimental apparatus of Experiment 2 was identical to those of Experiment 1.

3) Procedure: The general procedure of Experiment 2 was identical to that of Experiment 1, except that participants were requested to look at either eye, nose, or mouth depending on their experimental condition. A reminder instruction asking participants to see a particular area was inserted after a question item in each session.

C. Data preprocessing and Analysis

We used participants' eye-movement data that were collected when facial stimuli were presented, and counted the numbers of fixations at six areas of interest. The same set of definitions of AOIs for Experiment 1 were used in Experiment 2.

The criterion variables (y) were ratings on personality traits. Each of the five personality ratings was analyzed separately. The predictor variables were either five personality traits (Eq.5) or fixations on six areas of interest (Eq. 6). As in Experiment 1, we assumed that the criterion variable were distributed as Bernoulli.

$$y_{gikl} \sim Bernoulli(q_{gikl}) \tag{4}$$

$$q_{gikl} = \frac{1}{1+\exp\left(-\left(b_{0g}+\sum_{j=1}^{5} b_i P_{gijk} + r_{gik}^{subj} + r_{gil}^{pic}\right)\right)} \tag{5}$$

TABLE III
WAICS IN EXPERIMENT 2

Impression Rating	Personality	Obs. Behaviors
Agreeableness	938.097	**849.768**
Conscientiousness	918.221	**809.015**
Extraversion	1072.821	**931.194**
Neurosis	1336.904	**1152.504**
Openness	1112.198	**969.171**

TABLE IV
SIGNIFICANT PERSONALITY EFFECTS IN EXPERIMENT 2

Impression Rating	Cond.	Pred.	Mean	95% HDI
Agreeableness	Eye	Intcpt	5.071	1.297 - 9.124
Agreeableness	Nose	Consci.	-0.447	-0.847 - -0.053
Conscientiousness	Eye	Ext.	4.032	1.249 - 6.955
Extraversion	Eye	Ext.	3.521	0.319 - 6.493
Openness	Mouth	Open	0.434	0.042 - 0.862

$$q_{gikl} = \frac{1}{1+\exp\left(-\left(-b_{0g}+\sum_{j=1}^{6} b_i F_{gijk} + r_{gik}^{subj} + r_{gil}^{pic}\right)\right)} \tag{6}$$

Note that gs in Eqs. 4 - 6 indicate the experimental conditions (i.e., eye, nose, or mouth). We used Rstan for parameter estimation. The uniform prior was used for fixed effects, and weakly informative prior (gamma with $\alpha = 10$, $\beta = 10$) for random effects. As in Experiment 1, we used Rstan [9] for parameter estimation and its default settings for MCMC sampling (i.e, 4 chains, 1000 warmup steps, 2000 iterations and thin factor being 1). All $\hat{R}$ values in all 10 models were less than 1.1, indicating that MCMC samplings had converged.

D. Result

1) Manipulation check: Heat-maps and violin plots fixations are shown in Figs. 4 - 6. In general, participants in Nose and Mouth conditions extensively looked at nose and mouth, respectively, confirming that participant followed the instruction. In Eye condition, participants tended to look at both eyes and nose. However, as compared with Experiment 1, fixations on nose were decreased in Experiment 2, and we

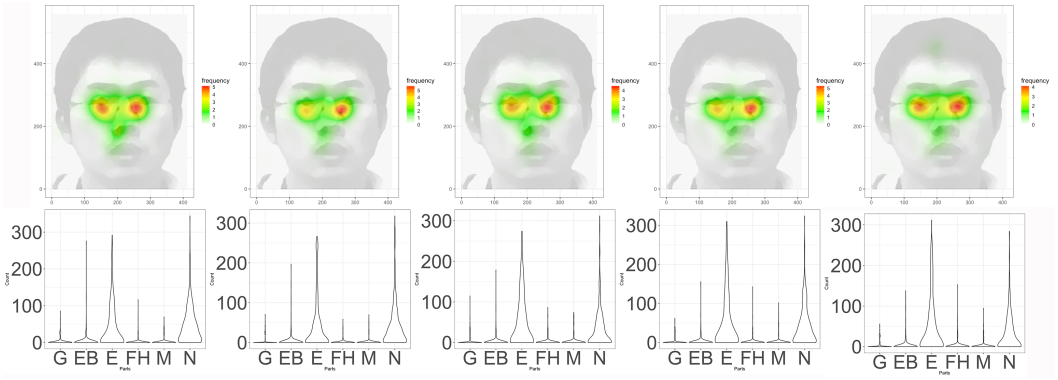

Fig. 4. Result of Eye condition in Experiment 2. Top: Heat-maps of fixation points for five personality rating items. Bottom: Violin plot of numbers of fixations on six areas of interest. From left to right: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience.

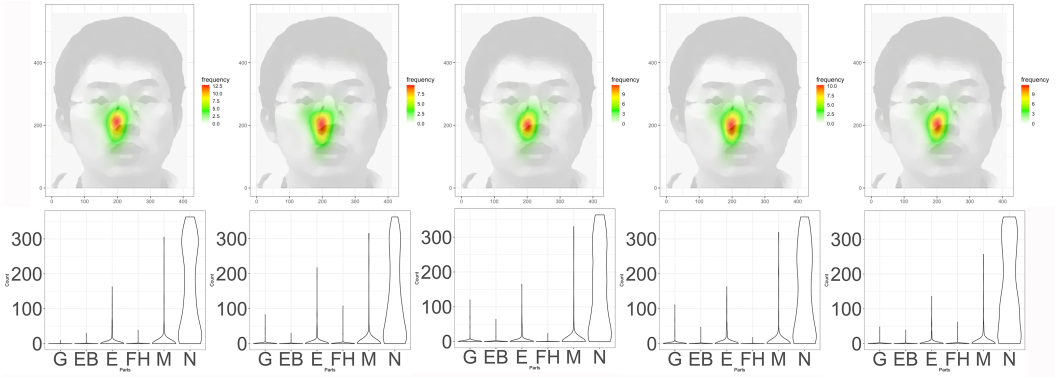

Fig. 5. Result of Nose condition in Experiment 2. Top: Heat-maps of fixation points for five personality rating items. Bottom: Violin plot of numbers of fixations on six areas of interest. From left to right: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience.

interpret this results as that fixations on eyes were increased, confirming experimental manipulation.

2) Model Comparison: Table III shows WAIC values for all 10 models in Experiment 2. For all personality rating items, models that incorporated observational behaviors (i.e, fixation on AOIs) resulted in better fits, even though participants were instructed to looked at particular areas of faces.

Table IV shows significant predictor variables for models with observers' personality. Contrary to our prediction, we found only a limited number of significant effects on observers' personality traits on impression rating. For example, a positive observers' extraversion effect on impression of extraversion of others' faces (when observers were instructed to look at eyes). This results confirmed the result of Experiment 1 in that observers' personality trait have rather weak effect on impression rating of others' faces.

Table V shows significant predictor variables for models with observers' observational behaviors (number of fixations). In Eye condition, we found significant effects of glancing at glabella on impression ratings on agreeableness, conscientiousness, and neurosis as well as effects of glancing at forehead on agreeableness and conscientiousness. In Nose condition, we found no effect of glancing. In mouth condition, we found effects of glancing at nose on conscientiousness and eyes and mouth on openness.

E. Discussion

In Experiment 2, we instructed to participants to look at particular areas of faces, either eyes, nose, or mouth, in order to weakens effect observational behaviors on impression ratings. However, instead of reducing the effect, it happened to strengthened the effects in contrary to our intention. Interestingly, looking at areas that were instructed to look at did not affect impression ratings. In this sense, we were able to weaken the effect of observational behaviors associated with these areas. However, in other sense, limiting observational behaviors enhanced the effects of observing other areas on impression ratings. This might have been caused by that limiting observational behaviors made participants to look at areas of faces that they really wanted to look at, which in turn strengthened the effect of glancing those areas on impression ratings. Each of the five impression rating items was significantly associated with at least one facial area.

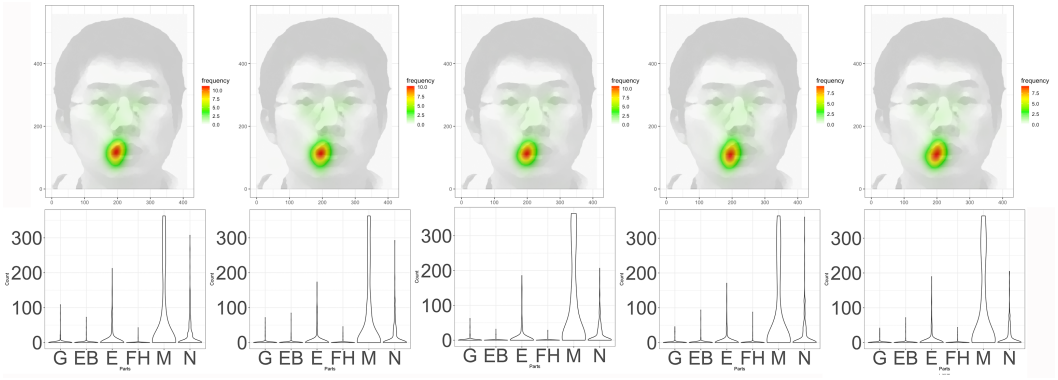

Fig. 6. Result of Mouth condition in Experiment 2. Top: Heat-maps of fixation points for five personality rating items. Bottom: Violin plot of numbers of fixations on six areas of interest. From left to right: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to experience.

TABLE V
SIGNIFICANT OBSERVATIONAL BEHAVIORS EFFECTS IN EXPERIMENT 2

Impression	Predictor	Condition					
		Eye		Nose		Mouth	
		Mean	95% HDI	Mean	95% HDI	Mean	95% HDI
Agreeableness	Intercept	1.330	0.394 - 2.295	n.s.		n.s.	
	Forehead	0.979	0.006 - 1.956	n.s.		n.s.	
	Glabella	1.508	0.494 - 2.451	n.s.		n.s.	
	Eye	79.510	0.131 - 195.363	n.s.		n.s.	
Conscientiousness	Intercept	1.190	0.071 - 2.220	n.s.		n.s.	
	Forehead	1.099	0.037 - 2.179	n.s.		n.s.	
	Glabella	1.257	0.168 - 2.310	n.s.		n.s.	
	Nose	n.s.		n.s.		0.014	0.001 - 0.029
Neurosis	Glabella	0.724	0.016 1.380	n.s.		n.s.	
Openness	Eye	n.s.		n.s.		-0.040	-0.076 - -0.002
	Mouth	n.s.		n.s.		0.003	0.001 - 0.006

IV. CONCLUSIONS

The vast majority of previous research on facial impressions dealt with how individual features (e.g. shapes of eyes) and/or their configurations affect impression formation on face. This line of research focusing on the features of facial parts and their configurations built on the basis of an implicit assumption that faces with particular features give particular sets of impression uniformly to observers. We assessed the validity of this assumption by conducting two experiment that examined effects of observers' characteristics on impression rating on faces.

In Experiment 1, we compared effects of observers' personality traits and observational behaviors on impression ratings on faces, and found that the latter effects were stronger than the former. However, there were higher degrees of variabilities in observational behaviors as a set of predictors than personality traits, resulting in potentially unfair comparisons. In order to conduct a fairer comparison, we restricted participants' observational behaviors by instructing them to look at particular areas of faces in Experiment 2. Although participant extensively looked at the areas that they were told to looked at, we found much stronger effects of observational behaviors on impression ratings on faces, suggesting their robust effects. With two experiments, we found robust significant effects of observers' characteristics, namely observational behaviors, on impression formation on faces.

REFERENCES

[1] Yamada, Y. & Sasayama, I. (1998). A study of the correlation between the impression formed from each features and the impression formed from face. *Bulletin of Fukuoka University of Education*. 48. 229-239.
[2] Todorov, A. T., Olivola, C.Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review on Psychology*. 66, 2015
[3] Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T.2014. Modeling first impressions from highly variable facial images. *Proceedings of National Academy of Science of the United of America* 11132. 3353-3361.
[4] Little A. C. & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology* 98. 111-26.
[5] Xu, K. & Matsuka, T. (2016), On the relationship between interpersonal impressions and where and how long people look as different regions faces. *Journal of Japanese Academy of Facial Studies.* 16. No.2:45-53.
[6] Oshio, A. Abe, S. & Cutrone, P. (2012). Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J). *The Japanese Journal of Personality* 21, 4052.
[7] Matsuura, K., Bayesian Statistical Modeling Using Stan and R. Kyoritsu Shuppan Co., Ltd 2016
[8] Kruschke, J. K.(2014). Doing Bayesian data analysis: A tutorial with R, JGAS, and Stan (2nd ed). Burlington, MA: Academic Press/Elsevier.
[9] Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.17.3. http://mc-stan.org/.