Age-related differences in the activation of the mentalizing- and reward-related brain regions during the learning of others' true trustworthiness

Check for updates

Atsunobu Suzuki [a,b,*], Mika Ueno [c], Kenta Ishikawa [d], Akihiro Kobayashi [d], Matia Okubo [d], Toshiharu Nakai [c]

[a] Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, Bunkyo-ku, Tokyo, Japan
[b] Department of Cognitive and Psychological Sciences, Graduate School of Informatics, Nagoya University, Chikusa-ku, Nagoya, Aichi, Japan
[c] Neuroimaging and Informatics Laboratory, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan
[d] Department of Psychology, Senshu University, Tama-ku, Kawasaki, Kanagawa, Japan

ABSTRACT

Behavioral studies suggest that older adults may be less adept than younger adults at remembering information contradicting their first impressions about others' trustworthiness. To identify the neural bases associated with such age-related differences, we measured the brain activity of older and younger participants using functional magnetic resonance imaging while they processed feedback on whether their initial trustworthiness impressions of stimulus persons, whose true trustworthiness had been predetermined, were right or wrong. Of special interest was the activation in mentalizing- (e.g., medial prefrontal cortex) and reward-related brain regions (e.g., striatum), which are known to be involved in impression formation and feedback learning, respectively. The reduction in the striatal responses to impression-contradicting versus impression-confirming feedback was greater in older than in younger participants. The activation of some mentalizing-related regions (medial prefrontal cortex and pre-cuneus) was lower in older than younger participants; however, it was not modulated by impression-feedback congruency. The results suggest that age-related differences in the striatum engagement may underlie older adults' inefficiency in learning impression-incongruent information about others' trustworthiness.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

The association between older age and fraud victimization is a critical yet controversial issue (Ross et al., 2014). Neuroscientists have therefore been exploring age-related brain differences that could potentially increase the risk of fraud victimization in older adults (OAs) (Castle et al., 2012; Samanez-Larkin et al., 2010; Spreng et al., 2017). A fundamental social skill that helps one avoiding deceitful exploitation is the ability to accurately judge others' trustworthiness, namely, to accurately distinguish "cooperators" from "cheaters" (Cosmides and Tooby, 1989). Therefore, age-related differences in brain function involving trustworthiness judgments would be of particular interest.

People initially assess others' trustworthiness from superficial cues, particularly from faces (Todorov, 2017). First impressions, however, are often inaccurate and require updating according to the actual behaviors of the target individuals (Suzuki and Suga, 2010). With respect to the initial face-based judgment, compared to younger adults (YAs), OAs tend to overestimate others' trustworthiness, possibly reflecting lower insular responsiveness to untrustworthy-looking faces (Castle et al., 2012) and higher amygdalar and striatal responsiveness to trustworthy-looking faces (Zebrowitz et al., 2018). Regarding the updating process, OAs seem to be less concerned than YAs with information that contradicts their first trustworthiness impressions about others, resulting in inefficient correction (Bailey et al., 2016; Bell et al., 2013; Rasmussen and Gutchess, 2018; Suzuki, 2018). This learning difficulty can increase OAs' vulnerability to "the wolf in sheep's clothing" (Suzuki and Suga, 2010) and is worth deeper scrutiny. The present study therefore aimed at elucidating its neural mechanisms, which are yet to be identified.

---

* Corresponding author at: Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan. Tel.: +81-3-5841-3859; fax: +81-3-5841-8969.
E-mail address: atsuzuki@l.u-tokyo.ac.jp (A. Suzuki).

Neural correlates of impression updating have been actively studied in young populations. Typically, participants are asked to form impressions of stimulus persons while reading a series of fictive sentences depicting their behaviors, of which earlier and later details imply opposite traits (e.g., Ma et al., 2012), or when viewing their faces while being paired with behavioral descriptions that are incompatible with facial impressions (e.g., Cassidy and Gutchess, 2015). Past research using these tasks has demonstrated that impression-incongruent information enhances the activity of a set of brain regions thought to be involved in "mentalizing," or making inferences about the minds of others (Koster-Hale and Saxe, 2013). These regions include the medial prefrontal cortex (mPFC), especially its dorsal portion (dmPFC), the temporoparietal junction (TPJ), and medial parietal regions such as the precuneus and posterior cingulate cortex (Cassidy and Gutchess, 2015; Hackel et al., 2015; Ma et al., 2012; Mende-Siedlecki et al., 2013; Stanley, 2016; but see also; Ames and Fiske, 2013). The increased neural activity is thought to reflect more elaborate attributional processing of incongruent than of congruent information as additional explanations are necessary to reconcile the discrepancy (Wyer, 2014; e.g., "Why on earth did a nice guy like him, who had been very kind to his customers, suddenly cheat them?"). In other words, low engagement of the mentalizing-related regions while processing impression-contradictory information can impede the learning of trustworthiness (Ferrari et al., 2016).

However, the traits of the stimulus persons were fictitious in the abovementioned studies (Cassidy and Gutchess, 2015; Hackel et al., 2015; Ma et al., 2012; Mende-Siedlecki et al., 2013; Stanley, 2016). Consequently, the receipt of impression-incongruent information did not imply that the participants' first impressions were necessarily inaccurate; instead, it suggested that they were merely different from the arbitrary rules set by researchers. When people judge the trustworthiness of "real cooperators" or "real cheaters" (e.g., Okubo et al., 2012) and are informed of their correct traits, they may possibly show greater brain responses to impression-confirming information rather than to contradictory information, especially in reward-related regions like the striatum (Bhanji and Delgado, 2014). This is because the confirmation corresponds to positive (intrinsically rewarding) feedback, indicating their possession of good social cognitive ability (Tricomi and DePasque, 2016). The resulting relative reduction in the striatal activity for impression-incongruent feedback can degrade its encoding by either attenuating the effortful cognitive processing (Cohen et al., 2014; Schmidt et al., 2012) or suspending the dopaminergic enhancement of memory consolidation (Shohamy and Adcock, 2010), thereby potentially retarding the update of trustworthiness impressions.

Based on both these theoretical considerations and OAs' difficulty in trustworthiness learning (Bailey et al., 2016; Bell et al., 2013; Rasmussen and Gutchess, 2018; Suzuki, 2018), a plausible hypothesis would be that OAs may under-recruit the mentalizing- and reward-related regions while processing impression-incongruent information when compared to YAs. Up to date, mixed results regarding age-related functional differences of these regions have been reported. In fact, while Moran et al. (2012) demonstrated that dmPFC activation during mentalizing tasks was lower in OAs than in YAs, Cassidy et al. showed both an age-invariant engagement of the dmPFC, TPJ, and precuneus when evaluating others in social domains (Cassidy et al., 2012) and some valence-dependent differences, namely, preferential neural responses to positive and negative information, in OAs and YAs, respectively (Cassidy et al., 2013). Moreover, the striatal activity for monetary feedback during value learning has been found to be lower in OAs than in YAs (see Samanez-Larkin and Knutson, 2015 for review), whereas that for performance feedback during a simple cognitive task was reported to be age-equivalent (Drueke et al., 2015). Importantly, the past investigations have not focused on the learning of trustworthiness; therefore, direct empirical testing of the above hypothesis is lacking.

For this reason, in the present study, we measured the brain activity of both older and younger participants using functional magnetic resonance imaging (fMRI) during the judgments of stimulus persons' trustworthiness (i.e., whether each person appeared cooperative or selfish) from their faces, and received immediate feedback regarding the correct answers, for which memory was later tested outside the scanner. Crucially, the actual trustworthiness of the stimulus persons had been measured in advance in the form of likelihood of cooperating with partners in a social exchange game (Okubo et al., 2012). Participants were informed of this fact and knew that the feedback was based on the actual trustworthiness of each person. Our primary interest was to examine age-related differences in brain activity for impression-contradictory information (i.e., feedback indicating that the participants' face-based judgments were incorrect), especially in the mentalizing- and reward-related regions of the brain including the mPFC, TPJ, precuneus, and striatum. As discussed previously, age-related differences in the disengagement of these regions during the processing of impression-incongruent feedback when compared to that of congruent feedback may be observed (hypothesis 1), and such neural disengagement may be associated with memory failure (hypothesis 2).

2. Material and methods

2.1. Participants

Twenty-five OAs (13 males and 12 females, 62–76 years of age [$M \pm SD = 69.80 \pm 3.32$]) and 25 YAs (14 males and 11 females, 20–33 years of age [22.28 ± 3.34]) participated in this study, with payment (7000 yen [about 65 US dollars]). They were recruited through local Silver Human Resources Centers, briefings at universities and community facilities, or by word of mouth. One OA and one YA were excluded from the analysis due to noncompliance with task instructions or recent psychiatric medication. Two OAs were further removed due to distinctive loss of MRI signals possibly associated with structural atrophy. The remaining participants (OAs: 11 males and 11 females, 62–76 years of age [$M \pm SD = 69.86 \pm 3.51$]; YAs: 13 males and 11 females, 20–33 years of age [22.21 ± 3.39]) were free of neurological and psychiatric illness, had normal or corrected-to-normal vision, and achieved scores of at least 27 on the Mini-Mental State Examination (Folstein et al., 1975). Other sociodemographic and psychological characteristics of the participants showed typical age-related differences (Salthouse, 2004): When compared to those of YAs, OAs had fewer years of education (OAs: $M \pm SD = 13.25 \pm 1.89$, YAs: $M \pm SD = 15.13 \pm 1.48$, $t(44) = -3.675$, $p = 0.001$), slower speed of processing, as measured by the Wechsler Adult Intelligence Scale-III (WAIS-III) Digit Symbol subtest (OAs: 71.86 ± 16.77, YAs: 100.75 ± 12.20, $t(44) = -6.571$, $p < 0.001$), poorer short-term memory, as measured by the WAIS-III Digit Span subtest (OAs: 15.73 ± 2.67, YAs: 20.96 ± 4.16, $t(44) = -4.918$, $p < 0.001$), and an equivalent level of general knowledge, as measured by the WAIS-III Information subtest (OAs: 17.41 ± 3.81, YAs: 18.29 ± 3.21, $t(44) = -0.833$, $p = 0.409$).

2.2. Face stimuli

Grayscale images depicting neutral faces of 32 "cooperative" and 32 "selfish" individuals were used as stimuli. They were a subset of participants in a previous study (Okubo et al., 2012), for which they

had played a trust game. On each trial of the game, they decided whether or not to cooperate with a hypothetical partner. Cooperation yielded a medium monetary reward (650 yen) both to themselves and to their partner, whereas noncooperation resulted in a maximum reward (1000 yen) to themselves but a minimum reward (300 yen) to their partner. A total of 84 young males played this game in the Okubo et al. study. The cooperative and selfish persons in the present study were the top 32 participants who showed higher cooperation rates (0.42–1.00) and the bottom 32 participants who showed lower cooperation rates (0.00–0.31), respectively.

Each facial image was cropped into a tall rectangular shape (198 × 280 pixels) so that the head was almost completely visible. Presentations of the faces and other task-relevant information were controlled by E-Prime 2.0 software (https://pstnet.com/). The stimuli were back-projected onto a screen in the scanner room, and participants viewed them using a mirror mounted on the head coil. The facial images subtended a visual angle of about 3.1°×4.4°.

2.3. Procedures and tasks

The study protocol was approved by the Ethics Committee of the National Center for Geriatrics and Gerontology, Japan. Written informed consent was obtained from each participant at the beginning of the experiment. The main tasks were a face-based trustworthiness judgment task, performed in the MRI scanner, and a memory-based trustworthiness judgment task, performed outside of the scanner.

2.3.1. Face-based trustworthiness judgment task

Before entering the scanner, for familiarization purposes, participants played the trust game (Okubo et al., 2012) with several fictitious partners, whose faces were not presented in the subsequent tasks. They were then told that the faces seen in the scanner were of people who had previously played the same game, and that those who had chosen "cooperation" (or "noncooperation") more than average were defined as "cooperative individuals" (or "selfish individuals") in the upcoming face-based trustworthiness judgment task. They were also encouraged to perform the task with effort as if they were testing themselves on their ability to judge people.

At each trial of the face-based trustworthiness judgment task, a male face was first presented for 6.5 seconds together with the appearance of two alternative judgments, "Cooperative" and "Selfish," during which participants answered whether they thought the male was a cooperative or selfish person, based on their impression of his face. The responses were made by pressing either a button under the left or right index fingers. The button assignments to the 2 alternatives were counterbalanced across the participants. After a 1-second waiting period, feedback as to whether the male was actually a cooperative or selfish person was displayed for 4 seconds, during which participants reformed impressions about him. A fixation interval occurred between trials, the duration of which was randomly chosen to be either 3.5 or 11 seconds, at a ratio of 7 to 1 (see Supplementary Fig. 1 for an example of the trial sequence).

The 64 face stimuli were divided into 4 sets so that each set included 8 cooperative and 8 selfish individuals. In each fMRI run, lasting for 290 seconds, 16 faces from 1 set were presented serially in a random order. Thus, there were 4 runs in total, and the order of the sets was counterbalanced across participants.

2.3.2. Memory-based trustworthiness judgment task

This task was administered outside of the scanner using a laptop computer. On each trial, a male face was presented along with the 2 alternative judgments, "cooperative" and "selfish," as in the face-based trustworthiness judgment task. Participants were asked to answer whether the male was cooperative or selfish by recalling the feedback they had received during the in-scanner task. No time limit was imposed, and the next trial started after a response was made. Each of the 64 face stimuli was presented serially in a random order.

2.4. Imaging protocol

Magnetic resonance images were acquired on a 3T scanner (Siemens MAGNETOM Trio, Erlangen, Germany) with a 12-channel head coil. Functional images were acquired using a T2*-weighted gradient echo planar imaging sequence with the following parameters: repetition time = 2500 ms, echo time = 30 ms, voxel size = 3 × 3 × 3.6 mm, field of view = 192 mm, and flip angle = 90°. Following the completion of the face-based trustworthiness judgment task, a high-resolution, magnetization-prepared, rapid-acquisition gradient echo image was acquired for anatomical details (repetition time = 2500 ms, echo time = 2.63 ms, field of view = 256 mm, flip angle = 7°, matrix 256 × 256, 208 slices per slab, slice thickness = 1 mm, and distance factor = 50%). Other structural images (e.g., fluid-attenuated inversion-recovery images) were also obtained, the analysis of which will be reported elsewhere.

2.5. fMRI data preprocessing and analyses

Image preprocessing and statistical analyses were conducted using Statistical Parametric Mapping 12 (http://www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB R2015b (https://www.mathworks.com/). Preprocessing of the functional images included slice-time correction, motion correction, co-registration with the magnetization-prepared, rapid-acquisition gradient echo anatomical image, spatial normalization to the Montréal Neurological Institute space including reslicing to a 2-mm isotropic voxel grid, and spatial smoothing with a 6-mm full width at half maximum.

In the first-level analysis, a voxel-by-voxel general linear model was applied to the preprocessed functional images. The model had four regressors of interest covering the 4-second feedback period, each of which corresponded to a unique combination of feedback congruence (congruent or incongruent [i.e., whether the feedback confirmed or contradicted the participant's judgment]) × correct trait (cooperative or selfish). In addition, regressors indicating the face onsets and button presses (0-second duration), as well as 6 head movement parameters, were included in the model as nuisance variables. The regressors were convolved with a canonical hemodynamic response function.

At the second level, the first-level contrast images representing brain activity during the feedback period were submitted to an analysis of variance (ANOVA) with 3 factors, including age group (older or younger), feedback congruence (congruent or incongruent), and correct trait (cooperative or selfish). As mentioned in the Introduction section, the activation in the mentalizing- and reward-related regions was of primary interest. Therefore, "reverse inference maps" were generated using Neurosynth (http://neurosynth.org/) with the search terms "mentalizing" and "reward" (see Supplementary Fig. 2). The regions included in the maps are presumably involved in the mentalizing and reward processing as their activation has been reported rather selectively in studies concerning these mental operations (http://neurosynth.org/faq/#q15; Yarkoni et al., 2011). The maps were combined and binarized to create a mask, and the second-level ANOVA was performed within it. The statistical threshold was set to $p < 0.001$

(uncorrected) at the voxel level and $p < 0.05$ (familywise error corrected) at the cluster level.

Age-related differences in the under-recruitment of the mentalizing- and reward-related regions during the processing of impression-incongruent person information when compared to that of congruent information (hypothesis 1) were identified by examining the age × feedback interaction. Regarding those regions showing this interaction, activation differences according to whether the feedback was retrieved correctly or incorrectly in the memory-based trustworthiness judgment task was further tested to clarify their role in the learning of trustworthiness (hypothesis 2). Specifically, regions of interest were defined as 6-mm radius spheres around the peak voxels of the regions showing the significant interaction (Tong et al., 2016). Later, the first-level general linear model was reapplied to each participants' functional images with feedback congruence (congruent or incongruent) and memory performance (correct or incorrect) as the factors of interest. Correct trait (cooperative or selfish) was excluded from this analysis as its inclusion yielded cells with a small number of trials, making the estimation unreliable. The main above-described second-level ANOVA revealed that correct trait did not have any significant effect on brain activity (see the Results section), suggesting that its omission would have a minimal impact. Finally, using the first-level images, mean contrast values were computed within the spherical regions of interest for each participant and the differences of the means according to memory performance were examined by t-tests.

3. Results

3.1. Behavioral data

In the face-based trustworthiness judgment task, the frequency of missing responses was very low and did not significantly differ between the 2 age groups (OAs: $M \pm SE = 0.136 \pm 0.075$ trials, YAs: 0.417 ± 0.225 trials, $t(44) = -1.142$, $p = 0.260$). In addition, mean response times did not significantly differ between the 2 groups either (OAs: 2.385 ± 0.126 seconds, YAs: 2.263 ± 0.121 seconds, $t(44) = 0.699$, $p = 0.489$). Furthermore, OAs and YAs tended to agree on which faces looked cooperative or selfish (see Supplementary Fig. 3).

By regarding the "cooperative" judgments given to cooperative and selfish people as hits and false alarms, respectively, the accuracy of performance (d') in the face- and memory-based trustworthiness judgment tasks was computed and submitted to an age group × task ANOVA (OAs and face-based: $M \pm SE = 0.370 \pm 0.057$, OAs and memory-based: 0.304 ± 0.065, YAs and face-based: 0.111 ± 0.066, YAs and memory-based: 0.230 ± 0.072). Only a significant main effect of age group was observed [$F(1, 44) = 5.135$, $p = 0.028$], indicating that OAs' judgments were more accurate than those of YAs in both the tasks. The absence of a significance main effect of task [$F(1, 44) = 0.210$, $p = 0.649$] and interaction [$F(1, 44) = 2.642$, $p = 0.111$] suggests an absence of between-tasks improved accuracy for both OAs and YAs. Nevertheless, although YAs' performance was at a chance level (i.e., $d' = 0$) in the face-based task [$t(23) = 1.685$, $p = 0.106$], indicating their inability to distinguish cooperative people from the selfish ones by their faces, it became significantly above chance in the memory-based task [$t(23) = 3.175$, $p = 0.004$]. In contrast, the d' of OAs surpassed the chance level in both tasks [face-based: $t(21) = 6.499$, $p < 0.001$, memory-based: $t(21) = 4.705$, $p < 0.001$; see also Supplementary Table 1 for information on response bias in the trustworthiness judgment].

3.2. fMRI data

The activation in the mentalizing- and reward-related regions (see Supplementary Fig. 2) during the feedback period of the face-based trustworthiness judgment task was examined using an ANOVA with 3 factors, including age group, feedback congruence, and correct trait. Of special interest for our study were age-related differences in the disengagement of these regions during the processing of impression-incongruent feedback when compared to that of congruent feedback (hypothesis 1), which should be identified as the age × feedback interaction. This interaction was significant in the bilateral striata (Table 1; Fig. 1A). As shown in Fig. 1B, the striatal activity was lower for incongruent than for congruent feedback in OAs [left/right: $t(21) = 2.113/3.444$, $p = 0.047/.002$], while the difference was not significant in YAs [left/right: $t(23) = -0.853/-0.384$, $p = 0.403/.705$]. Age-related differences in the reduction in the striatal activity for impression-incongruent feedback relative to congruent feedback were significant [left/right: $t(44) = 2.068/2.257$, $p = 0.045/.029$].

We further examined whether the striatal activity differed according to the correct or incorrect retrieval of the feedback in the memory-based trustworthiness judgment task (Fig. 2) to test the relationship between brain activation and memory performance

Table 1
Results of the group-level ANOVA of brain activity in the (A) mentalizing-related and (B) reward-related regions during the feedback period of the face-based trustworthiness judgment task

Effect	Cluster		Peak		
	Size (voxels)	$p_{\text{FWE-corrected}}$	Anatomical label	MNI coordinates	Z
A					
Younger > older	88	0.038	R Superior Medial Frontal	[4, 46, 34]	4.31
	151	0.008	R Precuneus	[4, −56, 34]	4.28
Congruent > incongruent	264	0.001	L Anterior Cingulate	[−6, 46, 6]	6.43
	826	<0.001	L Posterior Cingulate	[−2, −52, 28]	5.69
	92	0.034	R Precuneus	[18, −52, 16]	5.30
	267	0.001	L Middle Temporal	[−50, −68, 16]	5.19
	236	0.002	R Middle Temporal	[48, −60, 20]	4.67
B					
Age × feedback	154	0.001	R Putamen	[16, 12, 0]	4.35
	130	0.003	L Putamen	[−16, 12, 0]	4.23
Younger > older	162	0.002	R Anterior Cingulate	[10, 32, 18]	4.20
Congruent > incongruent	3447	<0.001	R Caudate	[10, 10, −4]	Inf
	74	0.030	L Middle Cingulate	[0, −30, 40]	6.47

The ANOVA includes three factors: age group, feedback congruence, and correct trait. The statistical threshold was set to $p < 0.001$ (uncorrected) at the voxel level and $p < 0.05$ (FWE corrected) at the cluster level. Only those effects revealing suprathreshold regional activation are presented. Anatomical labels are based on Tzourio-Mazoyer et al. (2002).
Key: ANOVA, analysis of variance; FWE, familywise error; MNI, Montréal Neurological Institute; L, Left; R, Right.

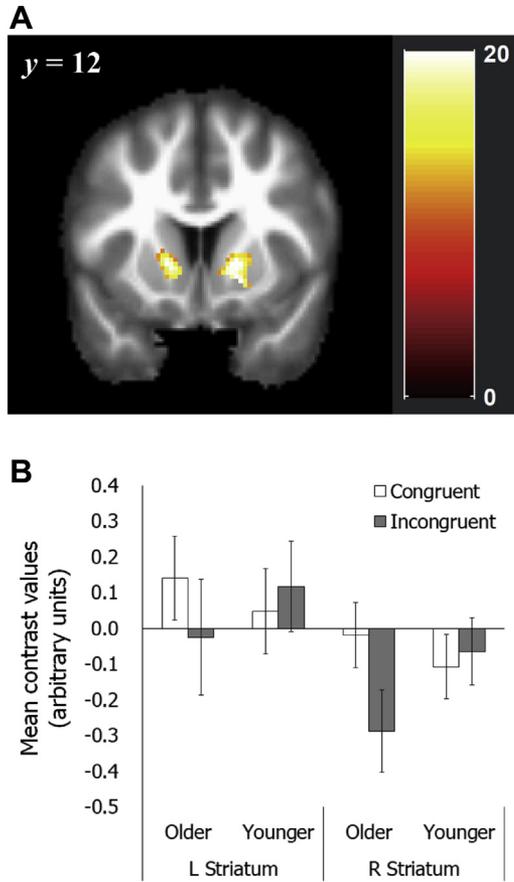**Fig. 1.** Age group × feedback congruence interaction. (A) The clusters of voxels in the bilateral striata where this interaction reached significance. (B) Mean contrast values within 6-mm radius spheres around the peak voxels in the striatal clusters above ([x, y, z] = [−16, 12, 0] [16, 12, 0]; see Table 1), as a function of age group and feedback congruence. Error bars represent standard errors. Abbreviations: L, Left; R, Right.

(hypothesis 2). In OAs, incorrectly retrieved incongruent feedback was associated with significantly lower striatal activity than was correctly retrieved feedback [left/right: $t(21) = 2.182/2.949$, $p = 0.041/.008$]. In contrast, the difference was not significant for congruent feedback [left/right: $t(21) = 1.335/0.255$, $p = 0.196/.801$].
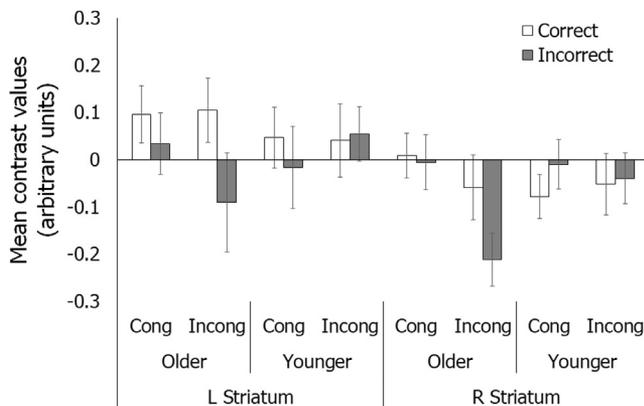

Fig. 2. Mean striatal activity as a function of age group, feedback congruence, and memory performance. The means were calculated by averaging contrast values across voxels within 6-mm radius spheres around the striatal peaks ([x, y, z] = [−16, 12, 0] [16, 12, 0]; see Table 1 and Fig. 1) that showed a significant age × feedback interaction. Error bars represent standard errors. Abbreviations: Cong, Congruent; Incong, Incongruent; L, Left; R, Right.

In YAs, the difference was not significant for either congruent [left/right: $t(23) = 0.720/−1.522$, $p = 0.479/.142$] or incongruent feedback [left/right: $t(23) = −0.259/−0.193$, $p = 0.798/.849$].

The age group × feedback congruence × correct trait ANOVA on the fMRI data also detected significant main effects of the age group and feedback congruence (Table 1). Specifically, YAs activated the dmPFC, precuneus (mentalizing-related regions), and anterior cingulate cortex (ACC) (reward-related region) more than OAs (Fig. 3A). However, the reverse contrast did not reveal suprathreshold activation in any region. In addition, congruent feedback elicited greater activation across a range of mentalizing- and reward-related regions than did incongruent feedback (Fig. 3B). The reverse contrast did not reveal suprathreshold activation in any region. Moreover, none of the effects involving correct trait (including the highest, three-way interaction) reached significance in any region.

The same ANOVA was also performed at the whole-brain level, reproducing the significant age × feedback interaction in the bilateral striata, significantly greater dmPFC and precuneus activation in YAs than in OAs, and widespread higher brain activity in response to congruent than to incongruent feedback (see Supplementary Table 2).

4. Discussion

We compared the brain activity of older and younger participants during a task in which they were required to intuitively judge the stimulus persons' trustworthiness and then received immediate feedback. The main hypotheses were that age-related differences in the disengagement of the mentalizing- and reward-related regions during the processing of impression-incongruent feedback when compared to that of congruent feedback would be observed (hypothesis 1), and that such neural disengagement would be associated with memory failure (hypothesis 2). The hypothesized patterns of activity were observed in the bilateral striata. Specifically, the reduction in the striatal activity in response to impression-incongruent versus impression-congruent feedback was greater in older than in younger participants. In addition, OAs' low striatal activity during encoding of incongruent feedback was associated with subsequent retrieval failure. On the other hand, we found a larger activation of the dmPFC, precuneus, and ACC in younger than in older participants during the processing of both congruent and incongruent feedback. The Neurosynth database (Yarkoni et al., 2011) indicates that these identified regions are relatively selectively involved in reward processing (striatum and ACC) or mentalizing (dmPFC and precuneus).

Greater modulation of striatal responses by the impression-feedback congruency in older than in younger participants may reflect age-related differences in the motivational impact of the impression-contradictory feedback. A primary functional characteristic of the striatum is its consistent activity for various types of rewards (Bhanji and Delgado, 2014), including positive feedback indicating good performance on cognitive tasks (Tricomi and DePasque, 2016). For OAs, who value "communal goals" of promoting cooperation and social bonds more strongly than do YAs (Fung et al., 2016), having the interpersonal skill to distinguish cooperators from cheaters should be more critical (Cosmides and Tooby, 1989). Thus, older participants may have found impression-incongruent feedback that signaled the lack of such an ability more frustrating and threatening than did younger participants, resulting in a larger decrement of striatal activity (Deveney et al., 2013; Seymour et al., 2015). The demotivating nature of such a feedback can then impede remembering it by either attenuating the mental effort put into the task (Cohen et al., 2014; Schmidt et al., 2012) or suppressing the dopaminergic
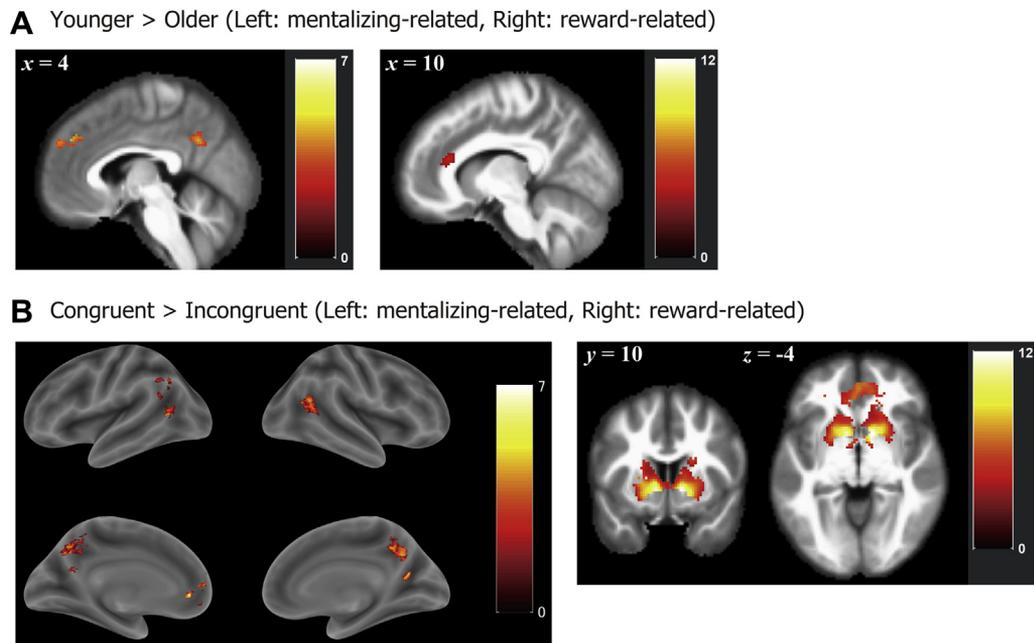
Fig. 3. (A) Brain regions showing significantly greater activation in younger than in older participants. The reverse contrast did not reveal any suprathreshold regional activation. (B) Brain regions showing significantly greater activation for congruent relative to incongruent feedback. The reverse contrast did not reveal any suprathreshold regional activation.

enhancement of memory consolidation (Shohamy and Adcock, 2010). In fact, the incongruent feedback that elicited smaller striatal activation tended to be forgotten in older participants.

The above explanation for the reduced striatal activity seen in OAs is in line with recent claims that socioaffective brain functioning may be greatly modified by motivational shifts with advancing age (Kensinger and Gutchess, 2017). It has also been argued, however, that older age is associated with a decline in dopaminergic neuromodulation (Eppinger et al., 2011), which may be responsible for both a disturbed striatal activity during reward learning (Chowdhury et al., 2013) and episodic memory deficits (Chowdhury et al., 2012) in OAs. Therefore, although representing a speculation, age-related differences in trustworthiness learning (Bailey et al., 2016; Bell et al., 2013; Rasmussen and Gutchess, 2018; Suzuki, 2018) might be resolved either by some motivational manipulations emphasizing the importance of correcting false impressions (DePasque and Tricomi, 2015) or by restoring the optimal dopamine levels (Chowdhury et al., 2012, 2013), which are intriguing possibilities to be explored in future studies.

Although activation in the mentalizing-related regions was also expected to show the age × feedback interaction, only the main effect of age was detected in dmPFC and precuneus activity, suggesting OAs' general tendency not to mentalize after forming the initial impressions. These results are consistent with Moran et al. (2012), who demonstrated lower dmPFC activity in OAs than in YAs across a variety of social cognitive tasks and discussed associations between their findings and age-related differences in the default mode network (DMN). The DMN is a set of functionally interconnected neural structures showing greater activity at rest than during cognitive tasks, and its major nodes include mPFC, TPJ, and precuneus, that is, the mentalizing-related regions (Mars et al., 2012). Age-related disturbances in DMN activity and connectivity are well documented (Mevel et al., 2011), and such neural deterioration may explain the age-related hypoactivation observed in this study. However, the exclusive use of younger faces as stimuli might be responsible for the main effect of age on brain activation as well. That is, lower motivation of OAs to learn about other-age (i.e., younger) individuals (Ebner et al., 2013) can result in less

engagement of the mPFC and precuneus during impression formation (Ames and Fiske, 2013). Likewise, smaller ACC responses in older participants might reflect the lack of self-referential processing of other-age people (Ebner et al., 2011, 2013). Although the ACC was designated as reward-related region in this study, its greater activity for self- versus other-relevant information has also been reported (Ebner et al., 2011; Hu et al., 2016). However, behavioral data, including both the accuracy and latency of trustworthiness judgments, demonstrated essentially comparable performances between the 2 age groups, suggesting that they were similarly engaged in the task; in fact, accuracy was significantly higher in older than in younger participants. Therefore, the potential role of the DMN disturbances and other-age effects must be clarified in future studies.

In addition, a range of brain regions associated with the mentalizing and reward processing showed greater responses to impression-congruent as opposed to incongruent feedback in both age groups. These findings seem to contradict prior reports describing an increased activity in the mentalizing-related regions during the processing of impression-incongruent information (Cassidy and Gutchess, 2015; Hackel et al., 2015; Ma et al., 2012; Mende-Siedlecki et al., 2013; Stanley, 2016). The present study, however, was unique in that feedback was given according to the actual trustworthiness of stimulus persons. Thus, due to a self-serving bias, impression-congruent information that confirms one's own competence (i.e., "good news") may tend to undergo greater processing (Eil and Rao, 2011), which in turn may result in greater recruitment of the mentalizing circuits. Indeed, such a motivational modulation of the neural activity during impression formation has been reported (Ames and Fiske, 2013).

The main limitation of this study is that trustworthiness learning was not clearly observed in the behavioral data. That is, although the accuracy of the trustworthiness judgment improved from chance to above-chance levels between face-based and memory-based sessions in the younger participants, suggesting a learning effect, an ANOVA revealed that the between-session difference was not significant in either younger or older participants. This obscurity of the learning effects may be due to difficulty in

remembering the trustworthiness of 64 unfamiliar people from a single study session; in comparison, for example, Suzuki (2018) had participants remember the trustworthiness of 16 individuals from 4 study sessions, to probe trustworthiness learning. While we found that the reduced striatal activity for incongruent feedback was associated with an impression updating failure in older participants, more direct evidence is necessary to establish whether such neural differences result in age-related difficulty in the learning of trustworthiness (Bailey et al., 2016; Bell et al., 2013; Rasmussen and Gutchess, 2018; Suzuki, 2018).

5. Conclusions

This fMRI study measured the brain activity of older and younger participants during their processing of feedback on their initial face-based trustworthiness impressions of real cooperators and real cheaters. Reduced striatal responses to impression-incongruent relative to congruent feedback was greater in older than in younger participants, and such a reduction was associated with memory failure. This suggests a lack of motivated processing of impression-incongruent information by OAs, which may constitute a neural basis for the age-related differences in trustworthiness learning that have been reported in previous studies.

Disclosure statement

The authors have no conflicts of interest to disclose.

Acknowledgements

References

Ames, D.L., Fiske, S.T., 2013. Outcome dependency alters the neural substrates of impression formation. Neuroimage 83, 599–608.

Bailey, P.E., Petridis, K., McLennan, S.N., Ruffman, T., Rendell, P.G., 2016. Age-related preservation of trust following minor transgressions. J. Gerontol. B Psychol. Sci. Soc. Sci. https://doi.org/10.1093/geronb/gbw141. [Epub ahead of print].

Bell, R., Giang, T., Mund, I., Buchner, A., 2013. Memory for reputational trait information: is social-emotional information processing less flexible in old age? Psychol. Aging 28, 984–995.

Bhanji, J.P., Delgado, M.R., 2014. The social brain and reward: social information processing in the human striatum. WIRES Cogn. Sci. 5, 61–73.

Cassidy, B.S., Gutchess, A.H., 2015. Neural responses to appearance-behavior congruity. Soc. Cogn. 33, 211–226.

Cassidy, B.S., Leshikar, E.D., Shih, J.Y., Aizenman, A., Gutchess, A.H., 2013. Valence-based age differences in medial prefrontal activity during impression formation. Soc. Neurosci. 8, 462–473.

Cassidy, B.S., Shih, J.Y., Gutchess, A.H., 2012. Age-related changes to the neural correlates of social evaluation. Soc. Neurosci. 7, 552–564.

Castle, E., Eisenberger, N.I., Seeman, T.E., Moons, W.G., Boggero, I.A., Grinblatt, M.S., Taylor, S.E., 2012. Neural and behavioral bases of age differences in perceptions of trust. Proc. Natl. Acad. Sci. U S A 109, 20848–20852.

Chowdhury, R., Guitart-Masip, M., Bunzeck, N., Dolan, R.J., Düzel, E., 2012. Dopamine modulates episodic memory persistence in old age. J. Neurosci. 32, 14193–14204.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Duzel, E., Dolan, R.J., 2013. Dopamine restores reward prediction errors in old age. Nat. Neurosci. 16, 648–653.

Cohen, M.S., Rissman, J., Suthana, N.A., Castel, A.D., Knowlton, B.J., 2014. Value-based modulation of memory encoding involves strategic engagement of fronto-temporal semantic processing regions. Cogn. Affect Behav. Neurosci. 14, 578–592.

Cosmides, L., Tooby, J., 1989. Evolutionary psychology and the generation of culture, part II: Case study: a computational theory of social exchange. Ethol Sociobiol. 10, 51–97.

DePasque, S., Tricomi, E., 2015. Effects of intrinsic motivation on feedback processing during learning. Neuroimage 119, 175–186.

Deveney, C.M., Connolly, M.E., Haring, C.T., Bones, B.L., Reynolds, R.C., Kim, P., Pine, D.S., Leibenluft, E., 2013. Neural mechanisms of frustration in chronically irritable children. Am. J. Psychiatry 170, 1186–1194.

Drueke, B., Weichert, L., Forkmann, T., Mainz, V., Gauggel, S., Boecker, M., 2015. Neural correlates of positive and negative performance feedback in younger and older adults. Behav. Brain Funct. 11, 17.

Ebner, N.C., Gluth, S., Johnson, M.R., Raye, C.L., Mitchell, K.J., Johnson, M.K., 2011. Medial prefrontal cortex activity when thinking about others depends on their age. Neurocase 17, 260–269.

Ebner, N.C., Johnson, M.R., Rieckmann, A., Durbin, K.A., Johnson, M.K., Fischer, H., 2013. Processing own-age vs. other-age faces: neuro-behavioral correlates and effects of emotion. Neuroimage 78, 363–371.

Eil, D., Rao, J.M., 2011. The good news-bad news effect: asymmetric processing of objective information about yourself. Am. Econ. J. 3, 114–138.

Eppinger, B., Hämmerer, D., Li, S.-C., 2011. Neuromodulation of reward-based learning and decision making in human aging. Ann. N. Y Acad. Sci. 1235, 1–17.

Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T., Todorov, A., Cattaneo, Z., 2016. The dorsomedial prefrontal cortex plays a causal role in integrating social impressions from faces and verbal descriptions. Cereb. Cortex 26, 156–165.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J. Psychiat Res. 12, 189–198.

Fung, H.H., Ho, Y.W., Zhang, R., Zhang, X., Noels, K.A., Tam, K.-P., 2016. Age differences in personal values: Universal or cultural specific? Psychol. Aging 31, 274–286.

Hackel, L.M., Doll, B.B., Amodio, D.M., 2015. Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on choice. Nat. Neurosci. 18, 1233–1235.

Hu, C., Di, X., Eickhoff, S.B., Zhang, M., Peng, K., Guo, H., Sui, J., 2016. Distinct and common aspects of physical and psychological self-representation in the brain: a meta-analysis of self-bias in facial and self-referential judgements. Neurosci. Biobehav Rev. 61, 197–207.

Kensinger, E.A., Gutchess, A.H., 2017. Cognitive aging in a social and affective context: Advances over the past 50 years. J. Gerontol. B Psychol. Sci. Soc. Sci. 72, 61–70.

Koster-Hale, J., Saxe, R., 2013. Functional neuroimaging of theory of mind. In: Baron-Cohen, S., Lombardo, M., Tager-Flusberg, H. (Eds.), Understanding Other Minds: Perspectives from Developmental Social Neuroscience. Oxford University Press, Oxford, pp. 132–163.

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., Fias, W., 2012. Inconsistencies in spontaneous and intentional trait inferences. Soc. Cogn. Affect Neurosci. 7, 937–950.

Mars, R., Neubert, F.-X., Noonan, M., Sallet, J., Toni, I., Rushworth, M., 2012. On the relationship between the "default mode network" and the "social brain". Front Hum. Neurosci. 6, 189.

Mende-Siedlecki, P., Cai, Y., Todorov, A., 2013. The neural dynamics of updating person impressions. Soc. Cogn. Affect Neurosci. 8, 623–631.

Mevel, K., Chételat, G., Eustache, F., Desgranges, B., 2011. The default mode network in healthy aging and Alzheimer's disease. Int. J. Alzheimers Dis. 2011, 535816.

Moran, J.M., Jolly, E., Mitchell, J.P., 2012. Social-cognitive deficits in normal aging. J. Neurosci. 32, 5553–5561.

Okubo, M., Kobayashi, A., Ishikawa, K., 2012. A fake smile thwarts cheater detection. J. Nonverbal Behav. 36, 217–225.

Rasmussen, E.C., Gutchess, A., 2018. Can't read my broker face: learning about trustworthiness with age. J. Gerontol. B Psychol. Sci. Soc. Sci. https://doi.org/10.1093/geronb/gby012. [Epub ahead of print].

Ross, M., Grossmann, I., Schryer, E., 2014. Contrary to psychological and popular opinion, there is no compelling evidence that older adults are disproportionately victimized by consumer fraud. Perspect. Psychol. Sci. 9, 427–442.

Salthouse, T.A., 2004. What and when of cognitive aging. Curr. Dir. Psychol. Sci. 13, 140–144.

Samanez-Larkin, G.R., Knutson, B., 2015. Decision making in the ageing brain: changes in affective and motivational circuits. Nat. Rev. Neurosci. 16, 278–289.

Samanez-Larkin, G.R., Kuhnen, C.M., Yoo, D.J., Knutson, B., 2010. Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. J. Neurosci. 30, 1426–1434.

Seymour, B., Maruyama, M., De Martino, B., 2015. When is a loss a loss? Excitatory and inhibitory processes in loss-related decision-making. Curr. Opin. Behav. Sci. 5, 122–127.

Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., Pessiglione, M., 2012. Neural mechanisms underlying motivation of mental versus physical effort. PLoS Biol. 10, e1001266.

Shohamy, D., Adcock, R.A., 2010. Dopamine and adaptive memory. Trends Cogn. Sci. 14, 464–472.

Spreng, R.N., Cassidy, B.N., Darboh, B.S., DuPre, E., Lockrow, A.W., Setton, R., Turner, G.R., 2017. Financial exploitation is associated with structural and functional brain differences in healthy older adults. J. Gerontol. A. Biol. Sci. Med. Sci. 72, 1365–1368.

Stanley, D.A., 2016. Getting to know you: general and specific neural computations for learning about people. Soc. Cogn. Affect Neur 11, 525–536.

Suzuki, A., 2018. Persistent reliance on facial appearance among older adults when judging someone's trustworthiness. J. Gerontol. B Psychol. Sci. Soc. Sci. 73, 573–583.

Suzuki, A., Suga, S., 2010. Enhanced memory for the wolf in sheep's clothing: facial trustworthiness modulates face-trait associative memory. Cognition 117, 224–229.

Todorov, A., 2017. Face Value: The Irresistible Influence of First Impressions. Princeton University Press, Princeton, NJ.

Tong, Y., Chen, Q., Nichols, T.E., Rasetti, R., Callicott, J.H., Berman, K.F., Weinberger, D.R., Mattay, V.S., 2016. Seeking optimal region-of-interest (ROI) single-value summary measures for fMRI studies in imaging genetics. PLoS One 11, e0151391.

Tricomi, E., DePasque, S., 2016. The role of feedback in learning and motivation. In: Kim, S., Reeve, J., Bong, M. (Eds.), Recent Developments in Neuroscience Research on Human Motivation: Advances in Motivation and Achievement. Emerald Group Publishing Limited, pp. 175–202.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Wyer, N.A., 2014. When I think of you: memory for persons and groups. In: Perfect, T.J., Lindsay, D.S. (Eds.), The SAGE Handbook of Applied Memory. Sage Publications, London, pp. 292–311.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670.

Zebrowitz, L.A., Ward, N., Boshyan, J., Gutchess, A., Hadjikhani, N., 2018. Older adults' neural activation in the reward circuit is sensitive to face trustworthiness. Cogn. Affect Behav. Neurosci. 18, 21–34.