How to Detect Altruists: Experiments Using a Zero-Acquaintance Video Presentation Paradigm

Ryo Oda¹ · Tomomi Tainaka² · Kosuke Morishima² · Nobuho Kanematsu² · Noriko Yamagata-Nakashima² · Kai Hiraishi³

Accepted: 24 December 2020 / Published online: 7 January 2021
© The Author(s) 2021

Abstract

In this study, we investigated the cognitive processes and nonverbal cues used to detect altruism in three experiments based on a zero-acquaintance video presentation paradigm. Cognitive mechanisms of altruism detection are thought to have evolved in humans to prevent subtle cheating. Several studies have demonstrated that people can correctly estimate levels of altruism in others. In this study, we asked participants to distinguish altruists from non-altruists in video clips using the Faith game. Participants decided whether they could trust allocation of money to the targets who were videotaped while talking to the experimenter. In our first experiment, we asked the participants to play the Faith game under cognitive load. The accuracy of altruism detection was not reduced when participants simultaneously performed a cognitive task, suggesting that altruist detection is rapid and effortless. In the second experiment, we investigated the effects of affective status on the accuracy of altruism detection. Compared with participants in a positive mood, those in a negative mood were more hesitant to trust videotaped targets. However, the accuracy with which altruism levels were detected did not change when we manipulated participants' moods. In the third experiment, we investigated the facial cues by which participants detected altruists. Participants could not detect altruists when the upper half of the target's face was hidden, suggesting that judgment cues exist around the eyes. We also conducted a meta-analysis on the effect size in each experimental condition to verify the robustness of altruism detection ability.

Keywords Altruist detection · Subtle cheating · Facial expression · Cognitive load · Affective state · Dual process theories

✉ Ryo Oda
oda.ryo@nitech.ac.jp

¹ Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

² Department of Computer Science, Nagoya Institute of Technology, Nagoya, Japan

³ Faculty of Letters, Keio University, Tokyo, Japan

Introduction

Cognitive mechanisms for detecting altruists are thought to have evolved in humans because those who cheat (i.e., receive benefits without contributing) pose a serious problem for the maintenance of reciprocal exchanges. According to Trivers (1971), cheating manifests in two forms: gross and subtle. Subtle cheating occurs in reciprocal relationships when one party attempts to give less than the other, whereas gross cheating occurs when a cheater does not reciprocate at all. Subtle cheating poses a major problem for the maintenance of reciprocal, altruistic relationships in division-of-labor situations that characterize complex human societies. One can escape the exploitation of subtle cheating by selectively engaging in exchanges with genuine altruists who behave altruistically irrespective of their own benefits and costs. Thus, cognitive mechanisms for detecting altruists can confer an evolutionary advantage. Indeed, some studies have shown that people can discern altruists from non-altruists. For example, using a zero-acquaintance video presentation paradigm, Brown et al. (2003) found that participants could detect altruists based on nonverbal cues. Oda et al. (2009b) reported similar results using 30-s video clips of natural conversations between Japanese individuals. After viewing the video clips without sound, participants correctly estimated the altruism levels of target individuals. Oda et al. (2009a) found that people tended to trust altruists more than non-altruists in an economic game using real money. The detailed mechanisms of altruist detection, however, were not extensively studied, prompting us to explore the factors that could enhance or hinder the detection of altruism. Specifically, we examined the effects of cognitive load (Experiment 1), mood induction (Experiment 2), and hiding upper and lower parts of facial stimuli (Experiment 3) on altruist detection.

Several points should be addressed before we explain the details of each experiment. First, it should be noted that genuine altruism is somehow different from other prosocial characteristics, such as cooperativeness and trustworthiness. Quite often, prosociality is defined by behaviors in economic games like the one-shot prisoner's dilemma game (which is related to cooperativeness) or trust game (which is related to trustworthiness). Those behaviors, however, are not necessarily products of genuine altruism because a player's outcome rests on the behavior of their counterpart. For example, a manipulative player may behave cooperatively to elicit cooperation from their counterpart. Thus, cooperativeness and trustworthiness in these games entail strategic altruism, which is different from genuine altruism (e.g., blood donation) expressed spontaneously and habitually. Genuine altruism could reflect the actors' combination of underlying personality traits.

How, then, can we assess genuine altruism? Two major methods have been proposed. One uses questionnaires that ask participants to self-report the frequency of helpful behaviors they have performed in the past (Johnson et al. 1989; Oda et al. 2013). The other utilizes the dictator game (Camerer 2003) in which one of the players, "the dictator", determines the distribution of "pie" (e.g., \$10) to their counterpart (recipient) without any influence from the latter. The altruistic behavior of the dictator, such as even distribution of the pie among the two, can thus be considered a manifestation of genuine altruism. In this study, we employed the questionnaire method to classify target individuals into altruist and non-altruist categories (see the "Method" section for more detail).

There is then the issue of how to measure the ability to detect altruism in participants. We employed the Faith game (Kiyonari and Yamagishi 1999; Mifune and Li 2018), which is an extension of the dictator game. In the dictator game, the recipient can do nothing but accept the decision made by the dictator. In the Faith game, recipients have two options.

Option A is to accept the decision made by the allocator, as in the dictator game (trust choice). Option B is to have a fixed amount of money provided by the experimenter (sure choice). There are three tricks. First, the fixed amount of money is always less than half of the pie. Second, the allocator believes that the recipient has no option but to accept their decision, and the recipient is informed of the dictator's misbelief. Third, the recipient has to decide before they know the decision by the allocator. If a recipient thinks the allocator is genuinely altruistic and will distribute the pie evenly (or a large share to the recipient), then they should choose Option A. Otherwise, they should take Option B to get the sure reward. That is, the trust choice is selected solely based on the recipient's estimation of the allocator's altruism. In this sense, the trust choice in the Faith game is different to the trust choice in the trust game. We asked our participants to play the Faith game with the videotaped individuals as allocators. This gave us the participants' dichotomous decisions as to whether they trusted each allocator. There could be four cases. First, the participant entrusts an altruistic allocator (Hit). Second, the participant entrusts a non-altruist (false alarm or type I error). Third, the participant fails to entrust an altruist (miss or type II error). Fourth, the participant does not entrust a non-altruist (correct rejection). The relative frequencies of the four cases enabled us to employ the signal detection theory to evaluate the accuracy of altruist detection.

Finally, there is the concern of using videos as stimuli. Recent studies using still pictures have shown that people attribute social and personality characteristics to facial appearances at a cost of little time and effort (Olivola et al. 2014). However, the accuracy of trustworthiness scores attributed to still pictures is thought to be only slightly higher than that of random guessing (Bonnefon et al. 2015; Todorov et al. 2015). In addition, different pictures of the same individual can induce different guesses (Todorov and Porter 2014). The detection of trustworthy partners from still pictures may also depend on the moment when the picture was taken (Verplaetse et al. 2007; Yamagishi et al. 2003). Given that portrait photographs are evolutionarily novel stimuli for humans, it is perhaps unsurprising that people cannot make reliable trustworthiness judgments based on them. Videos of target individuals, on the other hand, would contain more ecologically valid stimuli. However, only a few studies have employed video stimuli (e.g., Centorrino et al. 2015) and the factors that work to enhance or hinder the ability to detect altruism are not well known. Therefore, we explored three factors that may influence altruist detection.

In the first experiment, we focused on the process by which altruists are detected. If the cognitive mechanisms for detecting altruists are adaptations that prevent subtle cheating, it is plausible that these mechanisms may be domain-specific, working rapidly and implicitly like the cheater detection mechanism (Cosmides and Tooby 1992). We hypothesized that people make fast and frugal decisions using specific cues when they try to detect altruists. Evolved, domain-specific modules, such as cheater detection, are expected to exclude the contribution of common domain-general resources, such as working memory, attention, and inhibition. Indeed, Van Lier et al. (2013) tested the independency of the cheater detection module from general cognitive capacity and reported that a secondary task requiring working memory resources had no impact on the performance of the cheater detection task (Van Lier et al. 2013). We conjectured that an examination of the effects of cognitive load on altruism detection would clarify the process. If participants under cognitive load detect altruists as accurately as those under controlled conditions, then altruism detection via a domain-specific module would be supported. It should be noted, however, that the impact of cognitive load on performance depends on the complexity and difficulty of the task in hand (i.e., the focal task and the load task; Schmid 2016). This point will be addressed in the “Discussion” section.

In Experiment 1, we asked participants to decide whether they trusted targets in a video clip. Following the methods of Oda et al. (2009a), participants played the Faith game, in which they were recipients of videotaped altruist and non-altruist allocators. During the game, participants were asked to perform mental calculations (thus, dual tasks were required). Single-digit numbers (1–9) were orally presented in a pseudo-random sequence during the video, and participants were requested to provide the sum of these numbers at the end of each video clip. Similar methods have been applied in studies on the effects of cell phone use and conversation on driver performance (e.g., Caird et al. 2008). Signal detection analysis was then performed to compare detection performance between conditions with and without calculation.

In Experiment 2, we investigated mood-congruent effects on altruism detection by manipulating the affective state of the detector. Several studies have suggested that people are inclined to perceive faces in a mood-congruent manner (e.g., Bouhuys et al. 1995; Terwogt et al. 1991). Forgas and East (2008) reported that positive affect increased, and negative affect decreased, the perceived genuineness of videotaped facial expressions. Experiment 2 explored the effects of participants' moods on altruism detection and their skepticism about facial expressions. We experimentally manipulated the affective state of participants by asking them to listen to music that induced a positive or negative mood during the Faith game. Signal detection analysis was conducted to compare detection performance under the two mood conditions. There were four possible outcomes of this experiment. (1) If positive music increased, and negative music decreased, the perceived genuineness of facial expressions, then "false alarms" (i.e., trusting non-altruists) would increase under positive mood conditions and "misses" (i.e., failing to trust altruists) would increase under negative mood conditions, resulting in reduced detection accuracy under both mood conditions; (2) Misses would occur less frequently under positive mood conditions, whereas false alarms would occur less frequently under negative mood conditions, resulting in an increase in detection accuracy under both mood conditions; (3) The effect of mood on skepticism about the genuineness of facial expressions could be asymmetrical, resulting in a significant difference in detection accuracy between the two mood conditions (4) Detection accuracy would be robust, i.e., not affected by skepticism about the genuineness of facial expressions.

In Experiment 3, we investigated which parts of the face participants examined to detect altruism. Brown et al. (2003) reported that the degree of felt smile (i.e., *orbicularis oculi* muscle activity), head nods, time per smile, and smile symmetry were correlated with the altruism score of target persons. Oda et al. (2009b) replicated Brown et al.'s (2003) study and found that only the degree of *orbicularis oculi* muscle activity showed a significantly high correlation with the altruism score. That is, *orbicularis oculi* muscle activity (AU6 of the Facial Action Coding System) is a candidate cue employed to detect altruists. Voluntary control of the *orbicularis oculi* muscle was said to be much more difficult, making it a better cue for the genuineness of a smile because any cues should be difficult to mimic, to function as an honest signal (Brown et al. 2003; Zahavi 1975). Recent studies, however, have reported that a substantial number of people can deliberately manipulate the *orbicularis oculi* muscle (e.g., Gunnery et al. 2013), which suggests that its movement cannot be an honest signal. Despite uncertainty surrounding the genuineness of *orbicularis oculi* muscle activity, it would be a cue for detecting altruists if it correlates with the degree of altruism. Experiment 3 investigated altruism cues by restricting eye or mouth stimuli in the video clips. Participants in the "upper mosaic" condition were asked to play the Faith game with videotaped targets whose faces were blurred from the tip of the nose upward by mosaic processing. Participants in the "lower mosaic" condition were asked to play against targets

whose faces were blurred from the tip of the nose downward. Participants were expected to detect altruists more accurately in the lower mosaic condition than in the upper mosaic condition if *orbicularis oculi* muscle movement was the primary judgment cue.

In all experiments, we performed signal detection analysis (Gescheider 1997) to evaluate detection accuracy and criteria. The sensitivity parameter (d') and criteria parameter (c) were calculated using the hit rate (how often each participant trusted altruistic targets) and false alarm rate (how often each participant trusted non-altruistic targets).

Experiment 1: Cognitive Load

Participants engaged in two simultaneous tasks: the Faith game and mental arithmetic. In the Faith game, participants were required to decide whether they would entrust JPY 300 to videotaped target individuals. While the participants watched video clips of the target individuals, numbers from one to nine were orally presented in a pseudo-random sequence and participants were asked to sum the numbers after watching the clip. In the control condition, participants were not asked to calculate a sum. It was expected that participants in the control condition could trust altruist targets more than non-altruist targets (d' became significantly positive), while participants under cognitive load would detect altruists less accurately than those in the control condition if the cognitive process demanded effort. Participants were expected to be hesitant about trusting targets (c became significantly positive) under the cognitive load because of the stressful situation.

Method

Participants

We recruited 139 Japanese undergraduates (96 males, 43 females; mean $\pm$ SD age, 19.2 ± 1.2 years) from two universities. Each participant was paid JPY 500. Among the recruited participants, 42 (30 males, 12 females) were exposed to control conditions and 97 were exposed to experimental conditions. Of those in the experimental conditions, 46 (23 males, 23 females) were exposed to 2.0-s interval tasks of mental arithmetic and 51 (43 males, eight females) were exposed to 1.8-s intervals.

Stimuli

As stimuli, we used the same video clips used by Oda et al. (2009a, b). A self-reported altruism scale (Johnson et al. 1989) was used to select altruists and non-altruists for videotaping. On a scale of 1 (*never*) to 5 (*very often*), 69 male Japanese undergraduates (mean $\pm$ SD age, 18.7 ± 0.9 years) indicated how often they performed each of the 56 altruistic acts described in short statements. All participants were volunteers from a single class at a university in Japan. Participants' altruism scores were converted into percentiles; those at or above the 90th percentile were defined as altruists, and those at or below the 10th percentile were defined as non-altruists. Using these criteria, seven altruists and seven non-altruists were selected and asked to participate in videotaping. One altruist and three non-altruists declined to participate. The remaining 10 individuals (six altruists and four non-altruists) out of the 14 candidates for targets were individually escorted to a laboratory. The interviewer, who was unaware of each person's altruism score, sat beside a video camera

before the target individual. The interviewer asked trivial questions, such as targets' likes and dislikes in a general conversational style, similar to that when people meet for the first time. Targets were recorded in front of a white screen at close range (over the shoulder) in a setting completely unrelated to altruistic behavior. That is, conversations in the videos did not include anything about behavioral decisions in economic games or other situations where altruism might be involved. Videos were transformed into digital files and the first 30 s of each presentation were isolated. These 10 video clips were then edited into a sequence with deleted sound to limit the influence of verbal content (see Oda et al. 2009b for details).

Procedure

First, an experimenter explained the rules of the Faith game to the participants. Participants were told they would see video clips of interviews involving 10 Japanese men and that the videotaped targets were asked to allocate JPY 300 to recipients (this was a fabrication; targets were not actually asked to do this). Participants would then decide whether to entrust JPY 300 to each target. If the participant entrusted the target with the money, then the amount of money returned would be at the discretion of the target (trust choice); if the participant did not entrust the money to the target, then the participant would receive JPY 100 (sure choice). Although no money actually changed hands during this exercise, participants were asked to try to maximize the amount they would hypothetically receive.

The experimenter then escorted the participants to a soundproof room made of corrugated cardboard (VIBE Inc.; 110 cm width×110 cm length×164 cm height). The room contained a computer (Acer Aspire E1-571-H54D), a response sheet, a chair, a light-emitting diode (LED) desk lamp (Twinbird LE-H637W), and headphones (SONY MDR-ZX110NC). Participants were asked to wear the headphones and click a mouse to start the video file. Before the first video clip, instructional text was presented for about 1 min, indicating that participants would engage in mental arithmetic while they watched each video clip, such that a random selection of single-digit numbers would be orally presented through the headphones at intervals that would then be required to be mentally summed at the end of the clip. Fourteen to 16 single-digit numbers were orally presented. The time interval was either 2.0 s or 1.8 s, depending on the condition. The mental arithmetic was expected to be more difficult in the shorter 1.8-s interval condition. Participants were not informed as to how many numbers they would have to add or what the intervals between the numbers were. In the control condition, participants listened to the same synthetic voice at an interval of 1.8 s, but they were not asked to calculate a sum. At the end of each video clip, participants wrote their arithmetic answer and Faith game decision (entrust or not entrust) on a response sheet within 10 s, after which time the next video clip started. This study design was approved by the Institute's Ethics Committee.

Analysis

We performed signal detection analysis on each participant. Because hit and false alarm rates were sometimes equal to zero or one, we added 0.5 to all data (Hautus 1995). All errors reported with means in the text are standard deviation. The dataset is available in Online Source 1.

Results and Discussion

Manipulation Check

Significantly fewer correct answers were obtained at 1.8-s intervals than at 2.0-s intervals, indicating that mental calculation functioned as a cognitive load: 6.8 ± 2.5 for 2.0-s interval; 5.5 ± 2.8 for 1.8-s interval; $t(95) = 2.48$, $p = .014$, 95% CI for the difference of means (0.26, 2.37), Cohen's $d = 0.49$.

Altruist Detection

The mean d' in the control condition was 0.23 ± 0.63 . A one-sample t test rejected the null hypothesis that the mean of the population from which the data were drawn would equal zero, $t(41) = 2.38$, $p = .022$, 95% CI for the difference of means (0.03, 0.42), Cohen's $d = 0.37$; Fig. 1. This result supports previous findings that people can detect altruists in social exchanges (Oda et al. 2009a; Naganawa et al. 2010). Notably, the sensitivity parameter for this experiment was similar to that reported in a study using real monetary incentives ($d' = 0.27$ in Naganawa et al. 2010). This result suggests that the accuracy of altruist detection exists independently of reward.

Mean d' values obtained for 2.0-s and 1.8-s intervals were 0.20 ± 0.63 and 0.22 ± 0.58 , respectively (Fig. 1), which could not have been drawn from a population with a mean of zero: 2.0-s interval: $t(45) = 2.11$, $p = .040$, 95% CI for the difference of means (0.01, 0.38), Cohen's $d = 0.31$; 1.8-s interval: $t(50) = 2.73$, $p = .008$, 95% CI for the difference of means (0.05, 0.38), Cohen's $d = 0.38$. It should be noted, however, that when we control for the family-wise type I error rate with Bonferroni correction, mean d' values for 2.0-s intervals were no longer significant ($p = .040 > .025$), whereas those for 1.8-s intervals were significant ($p = .008 < .025$). There were no significant differences in d' among the three conditions, $F(2, 136) = 0.04$, $p = .95$. These results suggest that participants could detect altruists even while performing a cognitive task, at least in the 1.8-s interval condition. It remains unclear why performance in the apparently easier 2.0-s interval condition was worse than in the 1.8-s interval condition.

Fig. 1 Mean and SE of the score of d' divided by the condition

Correlations Between the Number of Correct Answers and Detection Accuracy

Failure to engage seriously with the calculation task could lead to apparent null effects of cognitive load on detection accuracy. However, correlations between the numbers of correct answers in the calculation task and detection accuracy were not statistically significant for either condition: 2.0-s interval: $r=0.12$, $t(44)=0.77$, $p=.443$, 95% CI for the difference of means (-0.18 , 0.39); 1.8-s interval: $r=0.12$, $t(49)=0.81$, $p=.420$, 95% CI for the difference of means (-0.16 , 0.37); see Fig. 2. This result indicates that participants who could detect altruists did not cut corners in calculation.

The mean c value for the control condition was 0.05 ± 0.50 . A one-sample t -test did not reject the null hypothesis that the mean of the population from whom the data were drawn would equal zero, $t(41)=0.70$, $p=.48$, 95% CI for the difference of means (-0.10 , 0.21). The mean c for the 2.0-s interval was 0.30 ± 0.59 , which could not have been drawn from a population with a mean of zero, even when Bonferroni correction was applied, $t(45)=3.45$, $p=.001 < .025$, 95% CI for the difference of means (0.12 , 0.47), Cohen's $d=0.51$. The mean c value for the 1.8-s interval was 0.14 ± 0.60 . The mean of the population from whom the data sample was drawn could be equal to zero, $t(50)=1.73$, $p=.08$, 95% CI for the difference of means (-0.02 , 0.31). There were no significant differences in c among the three conditions, $F(2, 136)=2.14$, $p=.121$; Fig. 3. C was significantly positive under the 2.0-s interval condition but decreased under the 1.8-s interval condition. If cognitive load made altruist detection more conservative, then c should have increased under the 1.8-s interval condition. Therefore, it remains unclear whether the detection strategy (i.e., the balance between preventing incorrect trust and promoting correct trust) depended on the degree of cognitive load. Further experiments are required to elucidate this relationship.

The result that cognitive load did not decrease detection accuracy raises the question of whether the load was appropriate. In a similar study, Phillips et al. (2007) employed the dual-task methodology to investigate the role of working memory in dynamic social cue decoding by using the Interpersonal Perception Task (IPT) and Profile of Nonverbal Sensitivity (PONS). Results revealed that decoding social cues from the IPT was not affected by the dual-task, while decoding from the PONS required working memory resources. The IPT consisted of video clips (cut to approximately 30 s) of social scenes from which the viewer had to decode an aspect of the social behavior. In the PONS, dynamic displays of face, body, and tone of voice of the same woman (who was not a professional actress) were portrayed, with each clip lasting approximately 2 s. Phillips et al. (2007) discussed whether the difference in interference could be due to the difference in the length between clips of the two tasks. In addition, it is also conceivable that when the cognitive load task is too difficult compared to the focal task, participants would ignore the load task entirely. Thus, the impact of cognitive load depends on the complexity and difficulty of the tasks (Schmid 2016). It is not clear, however, whether the detection of altruism involves the same mechanism as the decoding of dynamic social cues. It is difficult to measure objectively the effectiveness of the cognitive load task we employed because it is unknown what kind of cognitive task the detection of altruism is equivalent to. The correct answer rate for the calculation tasks was over 50%, which indicates that participants did not ignore the task but engaged seriously with it. It can be said that altruist detection is not affected by such a degree of load.

Fig. 2 Participants' sensitivity parameters d' as a function of the number of correct answers in mental calculation **a** in the 2.0 s-interval condition and **b** in the 1.8 s-interval condition

Experiment 2: Mood Manipulation

We investigated mood-congruent effects on altruist detection by manipulating the affective state of detectors. Regarding detection accuracy, the four possible outcomes described in

Fig. 3 Mean and SE of the score of c divided by the condition

the “Introduction” were expected. Regarding detection criteria, participants in a negative mood would be hesitant to trust the targets (c became significantly positive) if skepticism about them was induced by the negative mood.

Method

Participants

The study group consisted of 159 Japanese undergraduates (100 males, 59 females; mean age, 19.4 ± 1.0 years) recruited from two universities. Each participant was paid JPY 500. Among the total participants, 79 (48 males, 31 females) and 80 (52 males, 28 females) were exposed to positive and negative mood conditions, respectively.

Procedure

The procedure was similar to that in Experiment 1, except that no mental calculations were required. Instead, participants heard music through headphones throughout the video clip. Participants under the positive mood condition listened to ‘Hora staccato’ by Grigoris Dinicu, and those under the negative mood condition listened to ‘Chaconne for violin and continuo in G minor’ by Tomaso Antonio Vitali. Taniguchi (1995) investigated the mood induced by these pieces using the Affective Value Scale of Music and Multiple Mood Scale short form (Terasaki et al. 1991). Compared with the Chaconne, Hora staccato induced more “enhancement” and “liveliness” and less “depression (anxiety)”. We also manipulated the light conditions in the soundproof room. Under the negative mood condition, the brightness of the desk lamp was reduced to 16% that of the positive mood condition. We assessed participants’ affective state using the Multiple Mood Scale short form (MMS). This scale consisted of eight five-item subscales (depression, hostility, boredom, liveliness, well-being, friendliness, concentration, and startle), where each item was rated from 1 (*do not feel at all*) to 4 (*feel clearly*). We asked participants to complete the form immediately

after watching each video clip. This study design was approved by the Institute's Ethics Committee. The dataset is available in Online Source 1.

Results and Discussion

Manipulation Check

We compared subscale scores between the two mood conditions. As there were eight subscales, we set the alpha as $.0062 = .05/8$ (Bonferroni correction). Because we had directed prediction that the negative mood condition would induce negative mood and vice versa, we employed one-tailed tests. Among the eight subscales, only the hostility subscale (hostile, aggressive, hateful, repentant, and huffy) scores were significantly higher under the negative mood: positive, 6.8 ± 2.1 ; negative, 7.8 ± 2.6 ; $t(151.76) = 2.70$, $p = .0038 < .0062$, 95% CI for the difference of means (0.27, 1.77), Cohen's $d = 0.42$. There were no significant differences among the remaining seven subscales. Unlike Taniguchi (1995), we found that Hora staccato did not induce more "liveliness" or less "depression", perhaps because participants in this study completed the affective scale after the tiresome tasks, which may have led to a hostile mood.

Altruist Detection

The mean d' value for the positive mood condition was 0.43 ± 1.11 , whereas for the negative mood condition it was 0.36 ± 0.83 (Fig. 4). A one-sample t -test rejected the null hypothesis that the mean of the population from whom the data sample was drawn would equal zero: positive mood, $t(78) = 3.47$, $p < .001$, 95% CI for the difference of means (0.18, 0.68), Cohen's $d = 0.39$; negative mood, $t(79) = 3.87$, $p < .001$, 95% CI for the difference of means (0.17, 0.54), Cohen's $d = 0.43$. Both were significant even after Bonferroni correction was applied ($ps < .025 = .05/2$). There was no significant difference in d' between the positive and negative mood conditions, $t(144.03) = -0.49$, $p = .625$, 95% CI for the

Fig. 4 Mean and SE of the score of d' in the positive and negative mood conditions

difference of means ($-0.38, 0.23$), therefore, participants were able to detect altruists under both conditions.

The mean of the criteria parameter was negative when participants were in a positive mood (-0.06 ± 0.52), though a one-sample t -test did not reject the null hypothesis that the mean of the population from whom the data sample was drawn would equal zero, $t(78) = -1.02, p = .312$, 95% CI for the difference of means ($-0.17, 0.05$). The mean was positive when participants were in a negative mood (0.14 ± 0.57). Although a one-sample t -test did not reject the null hypothesis when we applied Bonferroni correction and set $\alpha = .025$, the 95% CI for the difference of means was above zero, $t(79) = 2.23, p = .028$, 95% CI for the difference of means ($0.01, 0.26$), Cohen's $d = 0.25$; Fig. 5. Moreover, the criteria parameter was significantly larger in the negative mood condition than in the positive mood condition, $t(155.96) = 2.33, p = .02$, 95% CI for the difference of means ($0.03, 0.37$), Cohen's $d = 0.37$. In sum, participants in a negative mood were more hesitant to trust targets than those in a positive mood, although the effect size was not particularly large.

The miss and false alarm rates were compared between the two mood conditions. The mean numbers of misses under positive and negative mood conditions were 2.4 ± 1.7 and 2.9 ± 1.5 , respectively; their difference was not statistically significant, $t(153.95) = 1.76, p = .08$, 95% CI for the difference of means ($-0.05, 0.97$), Cohen's $d = 0.31$. The mean number of false alarms for positive and negative moods were 1.7 ± 1.2 and 1.5 ± 1.2 , respectively; the difference was not significant, $t(156.71) = -1.23, p = .21$, 95% CI for the difference of means ($-0.60, 0.14$), Cohen's $d = 0.17$.

Although we found no evidence that a positive mood was induced in this study, we successfully induced negative moods, as shown by the significant difference in MMS hostility scores. The experimental booth was relatively small and dark, which may have been uncomfortable for some participants. The significant positive value of c observed in the negative mood condition suggests that the induction of a negative mood increased the hesitation of participants to trust targets. Nevertheless, detection accuracy was not affected by skepticism, and participants were able to detect altruists. The number of misses and false alarms were not significantly different between the two mood conditions, seeming to support the prediction that detection accuracy would be robust and unaffected by skepticism

Fig. 5 Mean and SE of the score of c in the positive and negative mood conditions

about the genuineness of facial expressions. It is necessary, however, to be careful about interpretation because the magnitude of the effect of mood manipulation is unclear.

Meta-Analysis

We performed a meta-analysis on sensitivity parameter effect sizes among the five conditions of Experiments 1 and 2, which involved 298 participants, to see if participants were able to detect altruists at all. A forest plot of sample effect estimates, their respective confidence intervals, and average effect sizes are shown in Fig. 6. Because the sample size (i.e., five studies) was small, we employed a fixed-effects model. In fact, heterogeneity (I^2) was 0.0, which supported the notion that neither cognitive load nor mood induction had substantial effects on detection accuracy. The estimated aggregate effect size was 0.38, 95% CI (0.27, 0.50), $z=6.36$, $p<.001$. Although the effect size was small, these meta-analysis results indicate that participants were able to detect altruists despite the obstructions of mental calculations and mood inducing music.

Experiment 3: Mosaic

Next, we investigated possible cues of altruism. We asked participants to play the Faith game with the upper or lower half of targets' faces blurred. It was expected that only d' in the lower mosaic condition would be significantly positive if *orbicularis oculi* muscle movement was the primary judgment cue. With regard to detection criteria, participants in both conditions would be hesitant about trusting the targets (c became significantly positive) because the mosaic would evoke anonymity, as half of the detection cues were lost.

Method

Participants

The study group consisted of 165 Japanese undergraduates (99 males, 66 females; mean age, 19.3 ± 0.9 years) recruited from two universities. Each participant was paid JPY 500. Among the total participants, 83 (50 males, 33 females) were exposed to upper mosaic blurring conditions and 82 (49 males, 33 females) were exposed to lower mosaic blurring conditions.

Fig. 6 Forest plot of experimental conditions in Experiments 1 and 2

Stimuli

We used the same video clips in this experiment that were used Experiments 1 and 2. The upper mosaic blurring condition consisted of targets with faces blurred from the tip of the nose upward using a mosaic processing software (Wondershare Filmora9). The lower mosaic blurring condition consisted of targets whose faces were blurred from the tip of the nose downward.

Procedure

The procedure followed that of Experiment 2, except that participants were not asked to listen to music. Thus, in the upper mosaic condition, participants were asked to decide whether to trust the target by relying on cues from only the lower part of the face, whereas in the lower mosaic condition, participants were asked to decide whether to trust the target by relying on cues from only the upper part of the face. This study design was approved by the Institute's Ethics Committee. The dataset is available in Online Source 1.

Results and Discussion

Altruist Detection

The mean d' value for the upper mosaic condition was 0.06 ± 0.80 . A one-sample t -test did not reject the null hypothesis that the mean of the population from whom the data were drawn would equal zero, $t(82)=0.66$, $p=.511$, 95% CI for the difference of means (-0.11 , 0.23). The mean d' value for the lower mosaic condition was 0.19 ± 0.81 , which could not have been drawn from a population with a mean of zero, $t(81)=2.13$, $p=.036$, 95% CI for the difference of means (0.01 , 0.37), Cohen's $d=0.23$; Fig. 7. There was no significant difference in d' between the upper and lower mosaic conditions, $t(162.84)=1.06$, $p=.289$, 95% CI for the difference of means (-0.11 , 0.38). It should

Fig. 7 Mean and SE of the score of d' in the upper-mosaic and lower-mosaic conditions

be noted that when we control for the family-wise type I error rate with Bonferroni correction, mean d' values for the lower mosaic condition were no longer significantly positive ($p = .036 > .025$). However, the mean d' value for the lower mosaic condition was not significantly different from that of the control condition ($d' = 0.23$) in Experiment 1 ($t(102.53) = 0.31$, $p = .758$, 95% CI for the difference of means $(-0.22, 0.30)$). In addition, the 95% CI for the difference of means was above zero, which suggests that participants were not entirely unable to detect altruists.

The mean c value for the upper mosaic condition was 0.23 ± 0.67 , whereas for the lower mosaic condition it was 0.26 ± 0.58 (Fig. 8). A one-sample t -test rejected the null hypothesis that the mean of the population from whom the data were drawn would equal zero: upper mosaic, $t(82) = 3.14$, $p = .002$, 95% CI for the difference of means $(0.08, 0.37)$, Cohen's $d = 0.34$; lower mosaic, $t(81) = 4.02$, $p < .001$, 95% CI for the difference of means $(0.13, 0.38)$, Cohen's $d = 0.45$. The criteria parameters were significantly positive in both conditions, even when we controlled for the family-wise type I error rate with Bonferroni correction, indicating that trust was generally conservative. In fact, the mean number of targets trusted in this experiment (4.1 ± 2.4) was significantly lower than in the control condition (Experiment 1), 4.9 ± 2.0 ; $t(73.1) = 2.17$, $p = .033$, 95% CI for the difference of means $(0.06, 1.51)$, Cohen's $d = 0.36$.

It was shown that hiding half of the faces significantly reduced the accuracy of altruist detection. The effect appeared to be stronger in the upper mosaic condition where the accuracy (d') was almost zero, suggesting that crucial cues about altruism appear in the upper part of a face. The criterion parameters were larger in the two experimental conditions in comparison to the control condition in Experiment 1, which indicates that participants were more conservative and less likely to entrust allocators when half of the targets' faces were blurred.

Fig. 8 Mean and SE of the score of c in the upper-mosaic and lower-mosaic conditions

General Discussion

Neither cognitive load nor mood manipulation greatly reduced the accuracy of altruism detection. A meta-analysis of the results for the five conditions revealed that participants were able to detect altruists despite interferences. However, participants could not recognize altruists when half of the targets' faces were blurred. The effect was more apparent when the upper half was blurred, suggesting there is a crucial judgment cue around the eyes.

Apart from the upper and lower mosaic conditions in Experiment 3, our meta-analysis results revealed that people were able to detect altruists through nonverbal cues that were available through video clips of target individuals. Even though the effects were not very large in size, they were observed and robust even when participants were distracted by cognitive tasks (Experiment 1) or emotion-inducing music (Experiment 2). This finding is remarkable given that the information available to participants lacked several important factors: the stimuli were two-dimensional, only the shoulders and above could be seen; there was no sound and the image was presented for only 30 s. In addition, there was no prior interaction between the participants and the target individuals.

There may be some possible limitations to this study. One is that it is not necessarily clear what the participants "detected" in the current experiments. The target individuals were classified either as an altruist or a non-altruist based on their responses on an altruism scale. Put differently, "altruism" was operationally defined as the scores of the scale. Therefore, the possibility remains that what the participants detected was not the altruistic traits per se but the "the scores of altruism scale." Although we cannot entirely disregard the possibility, it is notable that the behavioral validity of the scale has been guaranteed to a certain degree by Brown et al. (2003) who reported positive association between the scale scores and altruistic behavior on a dictator game. Another related point is that participants might not have found an altruistic target as "altruistic" but agreeable or nice. We would argue that, from a functional point of view, what counts is whether participants could discriminate trust those who were more likely to behave generously. The psychological labels attached to the person such as "altruistic", "agreeable", or "nice" have little relevance. For instance, there is an argument that male humans could detect mate value of a female human based on her physical appearances (Buss and Schmitt 2019). The same phenomenon can be described as that the men find the woman attractive, beautiful, or charming. It is apparent that those labels have little importance from a functional point of view. Therefore, we should emphasize that we defined "altruism" operationally and functionally in this study and that the concept should not be taken as a psychological construct.

Second, our sample was composed of participants from a single cultural background (i.e., Japanese culture). The generalizability of the results should be considered carefully, in the same way as knowledge from the US undergraduates or the Western, Educated, Industrialized, Rich, and Democratic (WEIRD) sample should not be naively applied to samples from other cultures (Cheon et al. 2020; Henrich et al. 2010). In fact, preceding studies have indicated that Japanese participants were more likely to rely on upper facial features, such as the eyes, while participants from other cultures (e.g., the U.S.) relied on lower facial features (Ozono et al. 2010; Yuki et al. 2007). Yuki et al. (2007) argued that this was because people in Japanese culture tend to restrict emotional expression. This means that Japanese people need to pay more attention to subtle, difficult-to-control movements around the eyes. Indeed, one study revealed that French participants were unable to correctly estimate the altruism level of videotaped Japanese targets using the same stimuli

in this study (Tognetti et al. 2018). This may be explained by the French participants, who relied more on lower facial parts, not having enough altruistic cues in the Japanese targets. To be precise, Tognetti et al. (2018) did not use the Faith game to measure participants' evaluation of altruism in the targets. Therefore, it is possible that it was methodological, not cultural, differences that brought about the differences in results. Still, we believe caution should be used in generalizing the results of our study.

A third limitation is that we videotaped only natural and spontaneous facial appearances of altruists. Even if people can detect altruism from the upper part of the face, it does not necessarily mean that movement of the *orbicularis oculi* muscle is the sole, reliable signal of altruism. In fact, recent studies suggest that a substantial number of people can manipulate the *orbicularis oculi* muscle deliberately (e.g., Gunnery et al. 2013). To escape the exploitation of such deceptive “smiles”, one has to use other signals. For example, Namba et al. (2017b) reported that dynamic sequences of facial movements were different between spontaneous and posed smiles and that observers were sensitive to such differences (Namba et al. 2017a). Unfortunately, our study did not address whether people can discriminate between those who pretended to be an altruist and those who are genuine. Future studies are needed to address how people evaluate intentional deceivers.

Acknowledgements This work was supported by JSPS KAKENHI Grant Nos. 15K04042 and 16K13462.

Authors' contribution All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ryo Oda, Tomomi Tainaka, Kosuke Morishima, Nobuho Kanematsu, Noriko Yamagata-Nakashima, and Kai Hiraishi. The first draft of the manuscript was written by Ryo Oda, Tomomi Tainaka, Kosuke Morishima, Nobuho Kanematsu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences*, 19, 421–422. <https://doi.org/10.1016/j.tics.2015.05.002>.
- Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, 33, 215–226. [https://doi.org/10.1016/0165-0327\(94\)00092-N](https://doi.org/10.1016/0165-0327(94)00092-N).
- Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, 1, 42–69. <https://doi.org/10.1177/147470490300100104>.
- Buss, D. M., & Schmitt, D. P. (2019). Mate preferences and their behavioral manifestations. *Annual Review of Psychology*, 70(1), 77–110. <https://doi.org/10.1146/annurev-psych-010418-103408>.
- Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis and Prevention*, 40(4), 1282–1293. <https://doi.org/10.1016/j.aap.2008.01.009>.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

- Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. *Evolution and Human Behavior*, 36(1), 8–16. <https://doi.org/10.1016/j.evolhumbehav.2014.08.001>.
- Cheon, B. K., Melani, I., & Hong, Y.-Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*. <https://doi.org/10.1177/1948550620927269>.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford: Oxford University Press.
- Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, 44, 1362–1367. <https://doi.org/10.1016/j.jesp.2008.04.010>.
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, 37, 29–41.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 27, 46–51. <https://doi.org/10.3758/BF03203619>.
- Henrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29. <https://doi.org/10.1038/466029a>.
- Johnson, R. C., Danko, G. P., Darvill, T. J., Bochner, S., Bowers, J. K., Huang, Y.-H., et al. (1989). Crosscultural assessment of altruism and its correlates. *Personality and Individual Differences*, 10, 855–868. [https://doi.org/10.1016/0191-8869\(89\)90021-4](https://doi.org/10.1016/0191-8869(89)90021-4).
- Kiyonari, T., & Yamagishi, T. (1999). A comparative study of trust and trustworthiness using the game of enthrone. *Japanese Journal of Social Psychology*, 15, 100–109. **(in Japanese with an English abstract)**.
- Mifune, N., & Li, Y. (2018). Trust in the faith game. *Psychologia*, 61, 70–88. <https://doi.org/10.2117/psysoc.2019-B008>.
- Naganawa, T., Yamauchi, S., Yamagata, N., Matsumoto-Oda, A., & Oda, R. (2010). Do altruists detect altruists easier than non-altruists? *Letters on Evolutionary Behavioral Science*, 1, 2–5. <https://doi.org/10.5178/lebs.2010.1>.
- Namba, S., Kagamihara, T., Miyatani, M., & Nakao, T. (2017a). The differences between true smile and false smile: The sequential differences affected emotion recognition. *Taijin Shyakai Sinrigaku Kenkyu*, 7, 45–51. **(in Japanese with an English abstract)**.
- Namba, S., Makihara, S., Kabir, R., Miyatani, M., & Nakao, T. (2017b). Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Current Psychology*, 36, 593–605. <https://doi.org/10.1007/s12144-016-9448-9>.
- Oda, R., Dai, M., Niwa, Y., Ihobe, H., Kiyonari, T., Takeda, M., et al. (2013). Self-report altruism scale distinguished by the recipient (SRAS-DR): Validity and reliability. *Sinrigaku Kenkyu: The Japanese Journal of Psychology*, 84, 28–36. <https://doi.org/10.4992/jjpsy.84.28>. **(in Japanese with an English abstract)**.
- Oda, R., Naganawa, T., Yamauchi, S., Yamagata, N., & Matsumoto-Oda, A. (2009a). Altruists are trusted based on non-verbal cues. *Biology Letters*, 5, 752–754. <https://doi.org/10.1098/rsbl.2009.0332>.
- Oda, R., Yamagata, N., Yabiku, Y., & Matsumoto-Oda, A. (2009b). Altruism can be assessed correctly based on impression. *Human Nature*, 20, 331–341. <https://doi.org/10.1007/s12110-009-9070-8>.
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18, 566–570. <https://doi.org/10.1016/j.tics.2014.09.007>.
- Ozono, H., Watabe, M., Yoshikawa, S., Nakashima, S., Rule, N. O., Ambady, N., et al. (2010). What's in a smile? Cultural differences in the effects of smiling on judgments of trustworthiness. *Letters on Evolutionary Behavioral Science*, 1, 15–18. <https://doi.org/10.5178/lebs.2010.4>.
- Phillips, L. H., Tunstall, M., & Channon, S. (2007). Exploring the role of working memory in dynamic social cue decoding using dual task methodology. *Journal of Nonverbal Behavior*, 31, 137–152. <https://doi.org/10.1007/s10919-007-0026-6>.
- Schmid, P. C. (2016). Situational influences in interpersonal accuracy. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 230–252). Cambridge: Cambridge University Press.
- Taniguchi, T. (1995). Construction of an affective value scale of music and examination of relations between the scale and a multiple mood scale. *The Japanese Journal of Psychology*, 65, 463–470. **(in Japanese with an English abstract)**.

- Terasaki, M., Koga, A., & Kishimoto, Y. (1991). Construction of the multiple mood scale-short form. In *Proceedings of the 55th annual convention of Japanese Psychological Association* 1991. (in Japanese).
- Terwogt, M. M., Kremer, H. H., & Stegge, H. (1991). Effects of children's emotional state on their reactions to emotional expressions: A search for congruency effects. *Cognition and Emotion*, 5, 109–121. <https://doi.org/10.1080/02699939108411028>.
- Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth' in facial inferences. *Trends in Cognitive Sciences*, 19, 422–423. <https://doi.org/10.1016/j.tics.2015.05.013>.
- Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different images of the same person. *Psychological Science*, 25, 1404–1417. <https://doi.org/10.1177/0956797614532474>.
- Tognetti, A., Yamagata-Nakashima, N., Faurie, C., & Oda, R. (2018). Are non-verbal facial cues of altruism cross-culturally readable? *Personality and Individual Differences*, 127, 139–143. <https://doi.org/10.1016/j.paid.2018.02.007>.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–55. <https://doi.org/10.1086/406755>.
- Van Lier, J., Revlin, R., & De Neys, W. (2013). Detecting cheaters without thinking: Testing the automaticity of the cheater detection module. *PLoS ONE*, 8, e53827. <https://doi.org/10.1371/journal.pone.0053827>.
- Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predicting cheating detection. *Evolution and Human Behavior*, 28, 260–271. <https://doi.org/10.1016/j.evolhumbehav.2007.04.006>.
- Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior*, 24, 290–301. [https://doi.org/10.1016/S1090-5138\(03\)00035-7](https://doi.org/10.1016/S1090-5138(03)00035-7).
- Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43(2), 303–311. <https://doi.org/10.1016/j.jesp.2006.02.004>.
- Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, 53, 205–214. [https://doi.org/10.1016/0022-5193\(75\)90111-3](https://doi.org/10.1016/0022-5193(75)90111-3).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.