

## Intersubjective action-effect binding 1

*Cognition*

Volume 127, Issue 3, June 2013, Pages 383–390

**Intersubjective action-effect binding: Eye contact modulates acquisition of bidirectional association between our and others' actions**Atsushi Sato<sup>a</sup> and Shoji Itakura<sup>b</sup><sup>a</sup> Faculty of Human Development, University of Toyama, 3190 Gofuku, Toyama 930-8555, Japan<sup>b</sup> Graduate School of Letters, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Correspondence concerning this article should be addressed to Atsushi Sato, Faculty of Human Development, University of Toyama, 3190 Gofuku, Toyama 930-8555, Japan

E-mail: [a\\_sato@edu.u-toyama.ac.jp](mailto:a_sato@edu.u-toyama.ac.jp)

**Abstract**

In everyday social life, we predict others' actions in response to our own actions. Subsequently, on the basis of these predictions, we control our actions to attain desired social outcomes and/or adjust our actions to accommodate the anticipated actions of the others. Representation of the bidirectional association between our and others' actions, that is, intersubjective action-effect binding, could make such intersubjective action control easier and smoother. The present study investigated not only whether or not intersubjective action-effect binding was acquired but also whether or not eye contact modulated it. Experiment 1 showed that after a repeated experience during which participants' finger movements triggered a target female individual's mouth gesture, observing the target's mouth gestures came to automatically trigger the participants' finger movements. Experiments 2 and 3 revealed that this effect was not observed when the target's gaze direction was averted (Experiment 2) or when the target's eyes were closed (Experiment 3) throughout the acquisition phase. These results indicate that intersubjective action-effect binding occurs and that an ostensive signal, that is, eye contact modulates it.

**Key Words:**

Intersubjective action-effect binding, Action control, Eye contact, Sensorimotor learning, Counter-mirror activation, Joint action

## 1. Introduction

How do we become intentional agents? Inspired by Lotze (1852) and Harleß (1861), James (1890) indicated that we cannot perform an act voluntarily unless we can foresee its effects. Voluntary action is taken according to our anticipation of the movements' sensible effects. We can voluntarily act only by anticipating the effects of the movement. In James' words, "when a particular movement, having once occurred in a random, reflex, or involuntary way, has left an image of itself in the memory, then the movement can be desired again, proposed as an end, deliberately willed" (James, 1890, Vol. 2, p. 487). Later, James' idea was further elaborated by Elsner and Hommel (2001), who proposed a two-stage model of acquisition of voluntary action control. According to their model, in the first stage, randomly produced movements lead to specific, perceivable changes in the environment. After repeated co-occurrences between movements and their effects, the motor pattern of that action becomes associated with that effect in a bidirectional manner. In the second stage, once such a bidirectional action-effect association has been acquired, actions are automatically activated by anticipation of their effects. Thus, movements come to be intentionally executed by activating the perceptual codes that represent the desired goal (i.e., expected effect).

Once bidirectional association between movement and effect is acquired, it has several functions other than action control. James (1890) noted a report by Lotze (1852) that the thrust of a sword triggered slight movements of spectators' arms. Later theory has suggested that such automatic mimicry allows us to directly understand the meanings of actions by internally replicating them without any explicit reflective mediation (e.g., Rizzolatti, Fogassi, & Gallese, 2001). In addition, it also contributes to the self/other distinction. My observed leg movement (and/or the feeling of my leg moving) is perfectly contingent upon my motor commands (or predictions based on motor commands), but another person's is not (e.g., Bahrck & Watson, 1985; Daprati, et al., 1997; Morgan & Rochat, 1997; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005). Moreover, self-produced sensations or distal effects can be correctly predicted on the basis of motor commands. This predictable component is removed from incoming sensory signals, thereby attenuating the sensory effect of self-generated movement. Such a mechanism enables differentiation between self-produced and externally generated sensations, thereby producing a sense of self-agency (i.e., the sense that I am the one who causes the action) (Bays, Flanagan, & Wolpert, 2006; Blakemore, Frith, & Wolpert, 1999; Sato & Yasuda, 2005).

To test their model, Elsner and Hommel (2001) conducted several influential experiments. Their experiments consisted of two phases: an acquisition phase and a test

phase. In the acquisition phase, participants experienced co-occurrences between left and right key presses and low- and high-pitched tones. In the subsequent test phase, participants were instructed to respond to the tones now used as imperative stimuli as quickly and correctly as possible according to a fixed stimulus-response mapping. Participants were randomly divided into two groups: the non-reversal and reversal groups. For the non-reversal group, the stimulus-response mapping in the test phase was consistent with the response-stimulus mapping acquired in the preceding acquisition phase. For the reversal group, the stimulus-response mapping was inconsistent with the response-stimulus mapping acquired in the preceding acquisition phase. The logic was as follows: if participants had acquired bidirectional associations between responses and tones, then presenting the tones should activate the associated responses. Therefore, response-compatible tones should allow for faster responses than response-incompatible tones. Their results supported this assumption. In the next experiment, they showed that acquired bidirectional action-effect associations also biased participants' freely made choices according to the learned associations. The acquisition of action-effect representations depends on both the temporal contiguity and the contingency between an action and an effect (contingency is the extent to which an action reliably predicts an effect), thus suggesting that action-effect representations are acquired by associative learning mechanisms (Elsner & Hommel, 2004). Action-effect binding can occur for voluntary nonactions and their effects (Kühn, Elsner, Prinz, & Brass, 2009). Moreover, action-effect associations can be acquired not only through active experience, but also through observing the actions of others (Paulus, van Dam, Hunnius, Lindemann, & Bekkering, 2011). A recent study showed that even 9-month-olds can acquire bidirectional action-effect associations through active experience (Verschoor, Weidema, Biro, & Hommel, 2010).

Although a number of studies have investigated action control in the physical world, it remains unclear whether or not bidirectional associations can be acquired even between one's own and another person's actions. Such intersubjective action-effect binding is important in the following respect: in daily life, we predict others' actions in response to our actions, and we select and control our actions to obtain desired responses from others on the basis of these predictions. Moreover, successful joint action in which co-actors cooperate to attain shared goals depends on the ability to integrate the predicted effects of one's own and others' actions (e.g., Sebanz, Bekkering, & Knoblich, 2006). In particular, when attainment of shared goals requires opposing or complementary actions between co-actors, we must adjust our actions to accommodate the anticipated actions of others. In that case, intersubjective action-effect binding

would enable smooth cooperation between co-actors. In addition to intersubjective action control and action understanding, intersubjective action-effect binding would contribute to the development of the sense of social agency (i.e., the sense that I am the one who causes and/or controls the other's action) or "interpersonal self" (Neisser, 1988), just like action-effect binding does to the development of the sense of self-agency in the physical world. In this study, by using a similar paradigm to Elsner and Hommel (2001), we investigated whether or not intersubjective action-effect binding was acquired (Experiment 1) and whether or not eye contact modulated it (Experiments 2 and 3).

Each experiment consisted of two phases: an acquisition phase and a test phase. In the acquisition phase, participants learned that each button press triggered a mouth gesture of a target female individual. In the test phase, previous effect stimuli were used as primes. If participants have acquired bidirectional associations between their own actions and the target individual's actions, presenting the effect-prime (mouth gesture) should activate the associated responses (finger movement). Therefore, responses should be facilitated when primed (observed) and required responses are congruent and should be interfered when they are incongruent. Moreover, given that eye contact modulates various aspects of the cognitive processing and/or behavioral responses which take place concurrent to or immediately following it (e.g., Csibra & Gergely, 2009; Farroni, Mansfield, Lai, & Johnson, 2003; Macrae, Hood, Milne, Rowe, & Mason, 2002; Mason, Hood, & Macrae, 2004; Senju & Csibra, 2008; Senju & Johnson, 2009), eye contact should also modulate an intersubjective action-effect binding.

In Experiment 1, the target individual's gaze direction was always direct throughout the acquisition phase, whereas in Experiment 2, the target's gaze direction was always averted. In Experiment 3, the target individual's eyes were always closed. Even if evidence is obtained showing that presenting an effect-prime activates the associated responses, it is possible that participants have simply acquired bidirectional associations between their actions and the change in photographs. In that case, the observed effect might be simply caused by the well-established action-visual event integration and priming rather than intersubjective action-effect binding, in which case it would have nothing to do with intersubjectivity. The findings that the presence or absence of eye contact modulates this effect would exclude this possibility and clarify the conditions under which intersubjective action-effect binding is facilitated.

## **2. Experiment 1**

### **2.1. Methods**

### 2.1.1. Participants

Twenty-two healthy, right-handed volunteers participated in this study. The participants were 8 male and 14 female individuals ranging in age from 19 to 25 years, with a mean age of 21.26 years. None had a history of neurological or psychiatric disease. All participants gave written informed consent but were naïve as to the purpose of the experiment.

### 2.1.2. Stimuli.

The four photographs (7.63° high and 4.77° wide) of the same female individual whose face was depicted in front view were used as stimuli with directed eyes. These stimuli were identical except for the mouth gestures depicted therein: mouth closing, lip protrusion, tongue protrusion, or cheeks puffing. A pilot study ( $N = 32$ ) showed that the following four gestures were judged as having neutral affect on a 9-point scale ranging from negative (-4) to positive (4): mouth closing ( $M = -0.47$ ,  $SD = 0.67$ ), lip protrusion ( $M = 0.47$ ,  $SD = 0.98$ ), tongue protrusion ( $M = -0.19$ ,  $SD = 0.99$ ), and cheeks puffing ( $M = -0.62$ ,  $SD = 0.97$ ).

### 2.1.3. Procedure

Upon entering the laboratory, participants were seated 60 cm from a 17-in. monitor. Experiment 1 consisted of two phases: an acquisition phase and a test phase.

#### 2.1.3.1. Acquisition phase

Participants were told to make a self-paced left button press with left index finger or a right button press with the right index finger. They were instructed to choose freely which button to press, but they were instructed to press the buttons in a random order and about equally often. Each trial started with the presentation of the photograph of the female face with mouth closed. This stimulus remained on the center of the monitor until the participants pressed the left or right button. After 50 ms had elapsed following each button press, a certain mouth gesture was presented for 300 ms: lip protrusion, tongue protrusion, or cheeks puffing (Figure 1a). In one example condition, the left button press triggered lip protrusion, while the right button press triggered tongue protrusion. The assignment of mouth gestures to buttons was consistent for each participant and counterbalanced across participants. The participants were not informed about the response-effect mapping but were told that the mouth gestures were completely irrelevant to the task (Elsner & Hommel, 2001). Each participant performed 300 acquisition trials.

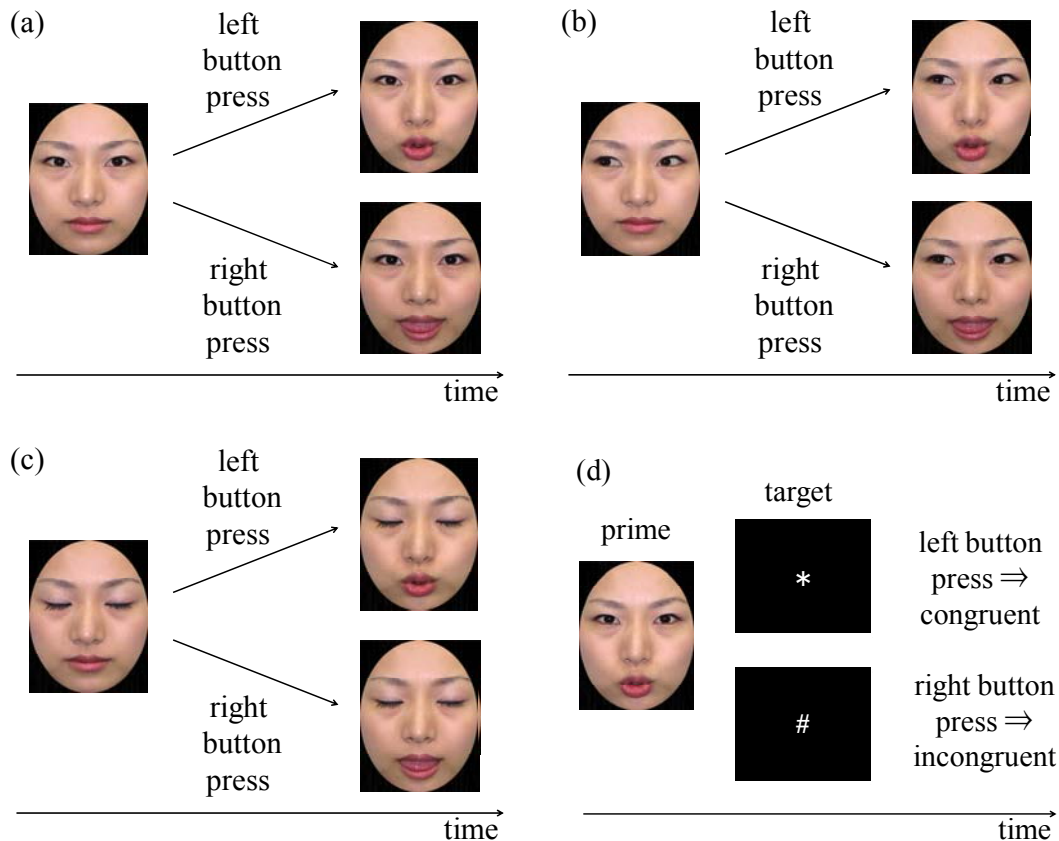


Figure 1 Schematic representations of the acquisition phase in Experiments 1(a), 2 (b) and 3 (c) and a schematic representation of test phase (d). In the case of (a), if a requested response to “\*” is pressing the left button, the requested response is congruent with the primed response.

### 2.1.3.2. Test phase

In the test phase, the above effect stimuli were used as primes. If participants had acquired bidirectional associations between actions and effects, presenting an effect-prime should activate the associated responses. Therefore, reaction times (RTs) should be slower when primed and required responses are incongruent than when they are congruent. Each trial started with the presentation of an effect-prime (300 ms duration). The primes consisted of the photographs of faces with lip protrusion, tongue protrusion, and cheeks puffing. Two of these faces had been used as effect stimuli in the preceding acquisition phase, whereas the remaining one had not been presented. After the presentation of the prime, one of two target stimuli was presented (“\*” (sized  $0.6 \times 0.6$  cm) or “#” (sized  $0.6 \times 0.6$  cm)). Participants were instructed to ignore the prime

stimuli and to press the left button with the left index finger whenever a “\*” appeared on the center of the monitor and the right button with the right index finger whenever a “#” appeared on the monitor as quickly and accurately as possible (Figure 1 (d)). Stimulus-response mapping was counterbalanced across participants. The next trial started after a 1-s delay.

The design of this experiment had one within-participants factor: congruency between primed and required responses (congruent or incongruent). In the congruent and incongruent conditions, the above effect stimuli were used as primes. The test phase consisted of 120 trials (40 per condition). In the remaining 40 trials, the non-presented stimuli from the acquisition phase were used as the primes. The trials were presented in random order.

## 2.2. Results and discussion

### 2.2.1. Acquisition phase.

The participants rendered button presses at a rate of about one per second ( $M = 1059$  ms,  $SD = 212$  ms). The proportions of left and right button presses were calculated in order to confirm whether or not the participants pressed each button equally as often. A t-test revealed that participants pressed the left and right buttons equally often (49.85% vs. 50.15%, respectively;  $t(21) = .30$ ,  $p = .766$ ).

### 2.2.2. Test phase

Error rates ( $M = 2.43\%$  of trials) were submitted to a paired t test according to the design in the previous section. There was no significant difference between congruent and incongruent conditions,  $t(21) = -.97$ ,  $p = .342$ . The mean RTs for correct responses were analyzed using a paired t test with the same design. As shown in Figure 2, the RTs were significantly faster in the congruent condition ( $M = 447$  ms,  $SEM = 10.16$ ) than in the incongruent condition ( $M = 463$  ms,  $SEM = 9.72$ ),  $t(21) = -4.70$ ,  $p < .001$ ,  $r = .72$ .

As noted above, if participants have acquired bidirectional associations between actions and effects, then presenting an effect-prime should activate the associated responses. Therefore, responses should be facilitated when the primed (observed) and required responses are congruent and should be interfered when they are incongruent. The present results showed this pattern. In this study, participants were told that the mouth gestures were completely irrelevant to the task in the acquisition phase and instructed to ignore the task-irrelevant prime stimuli in the test phase. Nonetheless, their behaviors were influenced by the task-irrelevant prime stimuli. These results were consistent with previous studies that showed that bidirectional associations can be automatically acquired and further demonstrated that presenting effect stimuli can

automatically activate associated responses (Elsner & Hommel, 2001). The present results extend previous findings by showing that action-effect binding can occur even between actions of the self and those of another.

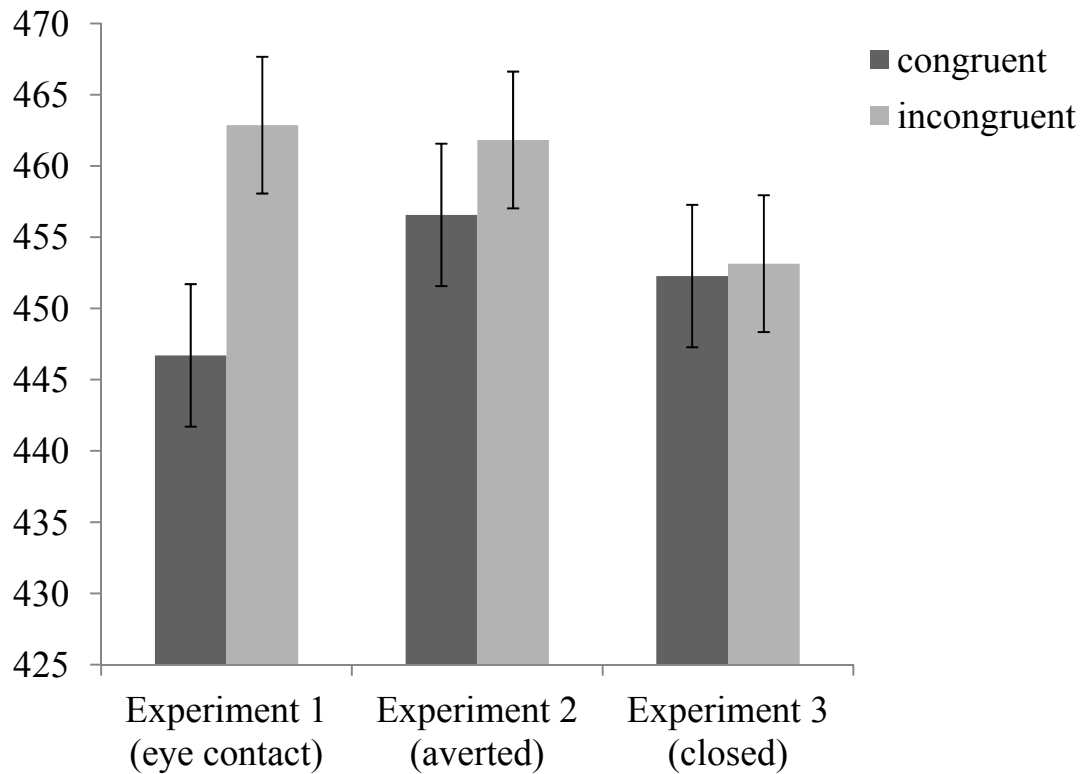


Figure 2 Mean reaction times for congruent and incongruent trials in Experiment 1 (eye contact), Experiment 2 (averted gaze), and Experiment 3 (closed eyes). Error bars represent standard error.

However, is the interaction measured by this experiment truly social? There is a possibility that participants had only acquired bidirectional associations between their actions and the change in photographs. If so, these results might have shown nothing more than the well-established concept of automatic action-visual event integration and priming. To exclude this possibility, in Experiment 2, face photographs of the female individual with averted gazes were used as the stimuli. Previous social cognitive studies have revealed that perceived eye contact with another human face modulates various aspects of the cognitive processing and/or behavioral response which take place concurrent to or immediately following it (e.g., Csibra & Gergely, 2009; Farroni, et al., 2003; Macrae, et al., 2002; Mason, et al., 2004; Senju & Csibra, 2008; Senju & Johnson,



2009). For example, eye contact triggers subsequent gaze following and enhances automatic mimicry of observed hand movements (Farroni, et al., 2003; Wang, Newport, & Hamilton, 2011). In addition, direct gaze captures visuospatial attention and facilitates face recognition (Guellai & Streri, 2011; Macrae, et al., 2002; Mason, et al., 2004; von Grünau, & Anston, 1995). Given these results, if the results of Experiment 1 are truly social in nature, eye contact should modulate the observed effect. Experiment 2 tested this hypothesis.

### 3. Experiment 2

#### 3.1. Methods

##### 3.1.1. Participants

Another 22 healthy, right-handed volunteers participated in this study; the participants were 7 male and 15 female individuals ranging in age from 18 to 23 years ( $M = 20.78$  years). None had a history of neurological or psychiatric disease. All participants gave written informed consent but were naïve as to the purpose of the experiment.

##### 3.1.2. Stimuli

The stimuli were the same as those in Experiment 1, except that all gazes depicted in the photographic stimuli were averted (Figure 1b). Averted gazes to the left were used throughout the experiment for one-half of the participants, and averted gazes to the right were used for the other half. Thus, gaze directions were constant through the experiment as in Experiment 1.

##### 3.1.3. Procedure

The procedure was identical to that in Experiment 1.

#### 3.2. Results and discussion

##### 3.2.1. Acquisition phase

Participants made button presses at the rate of about one button per second ( $M = 992$  ms,  $SD = 222$  ms). This pace did not significantly differ from that observed in Experiment 1 ( $t(42) = -1.01$ ,  $p = .315$ ). Participants pressed the left and right buttons about equally as often (49.18% vs. 50.82%, respectively,  $t(21) = -1.66$ ,  $p = .111$ ). These proportions did not significantly differ from those observed in Experiment 1 ( $t(42) = -1.01$ ,  $p = .315$ ).

##### 3.2.2. Test phase

Error rates ( $M = 2.84\%$  of trials) were submitted to a paired t test. There was no significant difference between congruent and incongruent conditions,  $t(21) = -1.03$ ,  $p$

= .313. The mean RTs for correct responses were analyzed using a paired t test with the same design. As shown in Figure 2, There was no significant difference between congruent and incongruent conditions,  $t(21) = -1.44$ ,  $p = .165$ .

In Experiment 2, face photographs with averted gazes were used as the primes in the test session. Thus, there is a possibility that gaze direction (left or right) has primed spatially compatible responses similar to the Simon task (e.g., Ansorge, 2003; Zorzi, Mapelli, Rusconi, & Umiltà, 2003). To examine this possibility, a repeated-measures ANOVA for RTs with two within-participants factors was conducted: one factor was congruency with the primed and required responses (congruent, incongruent, or non-presented stimuli from the acquisition phase); another factor was spatial correspondence between gaze direction and required response side (corresponding or non-corresponding). This analysis revealed no significant main effects or interaction,  $F_s = .60$ ,  $p > .446$ , partial  $\eta^2 < .028$ . These results suggest that the spatial correspondence between gaze direction and response side did not affect the results observed in Experiment 2. Unlike previous studies that found the “gaze-direction Simon effect” (in which the left- or right-gazing stimulus was presented in a random order), gaze direction was always constant throughout the test session in Experiment 2. This may be why the gaze-direction Simon effect was not observed in Experiment 2.

To confirm whether the presence or absence of eye contact modulated intersubjective action-effect binding, ANOVAs were conducted for the combined data from Experiments 1 and 2 with one within-participants factor (congruency) and one between-participants factor (gaze direction). Using error rate as the dependent variable, there were no significant main effects (congruency,  $F(1, 42) = 1.94$ ,  $p = .171$ ; gaze direction,  $F(1, 42) = .76$ ,  $p = .387$ ) or interaction ( $F(1, 42) = .12$ ,  $p = .729$ ). Using the mean RTs as dependent variables, there was a significant main effect of congruency,  $F(1, 42) = 18.44$ ,  $p < .001$ , partial  $\eta^2 = .31$ . Post-hoc analysis (Bonferroni,  $p < .05$ ) revealed that RTs were significantly faster in the congruent condition ( $M = 452$  ms,  $SEM = 6.88$ ) than in the incongruent condition ( $M = 462$  ms,  $SEM = 6.61$ ). There was no significant main effect of gaze direction,  $F(1, 42) = .11$ ,  $p = .741$ . More importantly, there was a significant interaction between congruency and gaze direction,  $F(1, 42) = 4.77$ ,  $p < .05$ , partial  $\eta^2 = .102$ . This interaction is explained by the fact that RTs were significantly faster in the congruent condition than in the incongruent condition in Experiment 1 whereas there was no significant difference between the two in Experiment 2.

These results indicated that eye contact did modulate intersubjective action-effect binding. This observed effect of eye contact was not due to a general arousal effect, as there was no significant main effect of gaze direction. Nevertheless, the present results

do not imply that action-effect binding would never occur without eye contact. In the present study, the effect-stimuli were more complex than those previously used (e.g., an 800-Hz tone), but they were presented for only 300 ms. Thus, there is a possibility that action-effect binding might have occurred only under an elevated number of acquisition trials or a lengthened presentation duration of effect-stimuli. Nevertheless, even under such restricted circumstances, evidence of action-effect binding was obtained in the presence of eye contact (Experiment 1). These results may suggest that human beings have an input bias towards socially significant and/or rewarding events, like mutual gaze. Consistent with this suggestion, previous studies have shown that ostensive signals of being addressed by communication, such as direct gazes, significantly modulate behavioral responses and learning which take place concurrently or immediately afterwards (e.g., Senju, & Csibra, 2008; Topál, Gergely, Miklósi, Erdohegyi, & Csibra, 2008; Wang, et al., 2011).

However, in Experiment 2, the target individual's gaze was always averted to the left or right side through the acquisition phase. Given that observing another person's averted gaze automatically shifts spatial attention (e.g., Langton, Watt, & Bruce, 2000), the present results may have been caused by participants' visual attention being distracted away from the center of the monitor. To exclude this possibility, in Experiment 3, face photographs of the female individual with closed eyes were used as the stimuli. If eye contact truly modulates intersubjective action-effect binding, a significant interaction between congruency and the presence or absence of eye contact should be observed even when face photographs with closed eyes were used as the stimuli. Experiment 3 tested this hypothesis.

## 4. Experiment 3

### 4.1. Methods

#### 4.1.1. Participants

Another 22 healthy, right-handed volunteers participated in this study; the participants were 6 male and 16 female individuals ranging in age from 19 to 23 years ( $M = 21.00$  years). None had a history of neurological or psychiatric disease. All participants gave written informed consent but were naïve as to the purpose of the experiment.

#### 4.1.2. Stimuli

The stimuli were the same as those in Experiment 1, except that the female individual's eyes were closed (Figure 1c).

#### 4.1.3. Procedure

The procedure was identical to that in Experiment 1.

## 4.2. Results and discussion

### 4.2.1. Acquisition phase

Participants made button presses at the rate of about one button per second ( $M = 1058$  ms,  $SD = 263$  ms). This pace did not significantly differ from that observed in Experiment 1 ( $t(42) = .01$ ,  $p = .996$ ). Participants pressed the left and right buttons about equally as often (49.76% vs. 50.24%, respectively,  $t(21) = .48$ ,  $p = .635$ ). These proportions did not significantly differ from those observed in Experiment 1 ( $t(42) = -.13$ ,  $p = .899$ ).

### 4.2.2. Test phase

Error rates ( $M = 1.59\%$  of trials) were submitted to a paired t test. There was no significant difference between congruent and incongruent conditions,  $t(21) = -.90$ ,  $p = .378$ . The mean RTs for correct responses were analyzed using a paired t test with the same design. As shown in Figure 2, There was no significant difference between congruent and incongruent conditions,  $t(21) = -.24$ ,  $p = .810$ .

To confirm whether the results were influenced by the presence or absence of eye contact, ANOVAs were conducted for the combined data from Experiments 1 and 3 with one within-participants factor (congruency) and one between-participants factor (eye contact). Using error rate as the dependent variable, there were no significant main effects (congruency,  $F(1, 42) = 1.75$ ,  $p = .194$ ; eye contact,  $F(1, 42) = 1.25$ ,  $p = .271$ ) or interaction ( $F(1, 42) = .00$ ,  $p = 1.00$ ). Using the mean RTs as dependent variables, there was a significant main effect of congruency,  $F(1, 42) = 12.05$ ,  $p < .005$ , partial  $\eta^2 = .22$ . Post-hoc analysis (Bonferroni,  $p < .05$ ) revealed that RTs were significantly faster in the congruent condition ( $M = 450$  ms,  $SEM = 6.78$ ) than in the incongruent condition ( $M = 458$  ms,  $SEM = 6.78$ ). There was no significant main effect of eye contact,  $F(1, 42) = .02$ ,  $p = .877$ . More importantly, there was a significant interaction between congruency and eye contact,  $F(1, 42) = 9.72$ ,  $p < .005$ , partial  $\eta^2 = .188$ . This interaction is explained by the fact that RTs were significantly faster in the congruent condition than in the incongruent condition in Experiment 1, whereas there was no significant difference between the two in Experiment 3.

To exclude the potentially confounding factor of gaze following, face photographs with closed eyes were used as the stimuli in Experiment 3. Nevertheless, the presence or absence of eye contact significantly modulated the results. As in Experiment 2, given that there was no significant main effect of eye contact, this observed effect of eye contact was not due to a general arousal effect. Taken together

with the results of Experiment 2, these results strongly indicated that eye contact did modulate intersubjective action-effect binding.

### 5. General discussion

The present study revealed two important findings. The first was that bidirectional associations between motor patterns and events could be acquired even between one's own actions and another person's actions. The second finding was that eye contact modulated such intersubjective action-effect binding.

Previous studies have shown that social contingency plays a crucial role in social interaction and learning (e.g., Kuhl, Tsao, & Liu, 2003; Murray & Trevarthen, 1985; Okanda & Itakura, 2008; Nadel, Carchon, Kervella, Marcelli, & Réserbat-Plantey, 1999). In infant-mother interactions, for example, Murray and Trevarthen (1985) showed that by 6 weeks of age, infants already displayed more positive affect during live (variably contingent) interactions with their mothers and more negative affect while viewing replayed interactions. Indeed, contingent interactions with a live person activate the reward, attention, and social-cognitive systems (e.g., Redcay, et al., 2010), thus facilitating early human learning (e.g., Kuhl, et al., 2003). However, in the present study, both temporal contiguity (50 ms delay) and contingency (each button press was a perfect predictor of the mouth gesture of the target female) were constant in all experiments. Thus, the present results cannot be attributed entirely to social contingency.

Naturally, even if observing mouth gestures triggers an observers' corresponding mouth gestures (Meltzoff & Moore, 1977), it does not trigger the observers' finger movements. In most cases, these phenomena are explained by the direct matching hypothesis, which postulates that imitation is based on an innate mechanism—such as mirror neurons—that directly maps the observed action to an internal motor representation of that action (e.g., Rizzolatti, et al., 2001). Indeed, a number of studies have demonstrated that observing an index-finger movement automatically triggers an index-finger movement rather than a middle-finger movement (e.g., Brass, Bekkering, Wohlschläger, & Prinz, 2000). It has also been suggested that such a system contributes not only to resolution of the corresponding problem (how another's action can be translated into one's own action, or vice versa) but also to action understanding (e.g., Rizzolatti, et al., 2001). However, through a very simple associative learning process, the target female's mouth gesture in the present study did activate participants' corresponding finger movements. Recently, Heyes (2001) and Heyes (2010) has proposed that mirror neurons are a product of associative learning: that is, each mirror

neuron is forged through the correlated experience of observing and performing the same action. In support of her associative hypothesis, Catmur, Gillmeister, Bird, Liepelt, Brass and Heyes (2008) demonstrated that after repeatedly observing a foot lift while a hand was concurrently lifted, and vice versa, the cortical area that responds more strongly to observation of hand actions also showed greater responses to observation of foot actions. The present results are in line with her proposal and extend it by showing that through the correlated experience of executing a finger movement and observing another's mouth gesture, observing another's mouth gesture can activate one's own finger movement.

As noted above, the present findings suggest that intersubjective action-effect binding proceeds according to general-purpose learning mechanisms, which are common in other species. So, why are human beings so good at intersubjective action-effect binding? The finding that eye contact modulated intersubjective action-effect binding may provide a clue. Although all experiments in the present study used the same procedure, the evidence of intersubjective action-effect binding was not obtained when there was no eye contact (Experiment 2 and 3). These results suggest that human beings are biased towards socially significant stimuli, as conveyed by eye contact. In humans, even newborns prefer to look at faces with direct gazes over faces with averted gazes (Farroni, Csibra, Simion, & Johnson, 2002). In addition, even in newborns, eye contact facilitates face recognition (Guellai & Streri, 2011). Senju and Johnson (2009) propose the neural basis of such an effect of eye contact: they postulate that the subcortical pathway including the superior colliculus, pulvinar, and amygdale—which is relatively mature even in infants—rapidly detects eye contact (or socially significant stimuli), and then it regulates subsequent cortical processing in the areas to which it projects in a top-down fashion. In addition, they propose that although detection of eye contact initially activates a widespread cortical structure, postnatal social experience interacts with the innate architectural bias to narrow down widespread activation to form specialized connections between subcortical and cortical structures during the course of development. As Farroni et al. (2002) suggest, human infants are equipped with a bias to detect and orient towards faces that make eye contact with them. Moreover, in our social lives, important information is often conveyed through eye contact. Such a cultural habitus might further shape our input bias towards socially significant and/or rewarding events by making such stimuli more salient. Although intersubjective action-effect binding depends on general-purpose learning mechanisms, our innate and/or culturally inherited input bias modulates or guides our learning, thus allowing for intersubjective action-effect binding. If this is true, then individuals with

autism spectrum disorder, in whom the development of eye contact might be disrupted (e.g., Senju & Johnson, 2009), might display difficulties in the development of intersubjective action-effect binding even if they easily acquire bidirectional associations between their actions and effects in the physical world. Consequently, they might display difficulties in the development of the sense of social agency, despite an intact sense of self-agency in the physical world. Future research is required to test this hypothesis.

The present findings have several implications. In everyday life, we predict other people's actions in response to our actions, and then, based on these predictions, we control our actions to attain desired social outcomes and/or adjust our actions to accommodate the anticipated actions of others. The present results suggest that intersubjective action-effect binding contributes to such intersubjective action control. Moreover, in joint action settings, co-actors are often required to select opposing or complementary rather than identical actions in order to achieve shared goals. In such cases, automatic imitation through direct matching may even be an obstacle to successful cooperation between persons. In this regard, intersubjective action-effect binding will enable smooth cooperation between persons. In addition, George Herbert Mead (1934) suggested that we know ourselves through the actions others take in response to our actions. In his theory, to take the attitude of another is to evoke within oneself another's response that is triggered by one's own actions. We act in anticipation of the responses of other people. These responses in anticipation of responses make for a dialectical self: the attitudes of others constitute the organized "me," and then one reacts toward that construct as "I." Beyond dyadic interactions, in games like baseball or football, we are required to learn not only the responses of specific others but also the actions associated with every position on the field. That is, we must take on the organized and generalized attitude of a social group (i.e., the generalized other). According to Mead, a mature sense of self can be achieved by learning to respond to and take on the attitude of the generalized other. Thus, in different terms, intersubjective action-effect binding constitutes the very basis of Mead's theory. Intersubjective action-effect binding has the potential to clarify not only how representations of society or social norms are formed (e.g., it may be a hierarchically organized system of multiple intersubjective action-effect bindings), but also how self-consciousness emerges.

The present study has several limitations. The effect stimuli from the acquisition phase were also used as the prime stimuli in the test phase. Thus, in some conditions (Experiment 1), the prime stimuli were faces with direct gazes, while in other cases (Experiment 2 and 3), the prime stimuli were faces with averted gazes or closed eyes.

This procedural constraint made it difficult to exclude completely the possibility that the observed effect was simply caused by the gaze directions of the prime stimuli. Moreover, although Elsner and Hommel (2004) showed that action-effect representations are acquired by associative learning mechanisms, it remains unresolved whether or not the same is true for intersubjective action-effect binding. Regarding temporal contiguity, parental responses in early infant-caregiver interactions generally occur within 1–2 seconds after the infant’s behavior (e.g., Keller, Lohaus, Völker, Cappenberg, & Chasiotis, 1999). Such prompt responses to communicative signals allow the infant to easily associate his/her action with subsequent parental action. In the case of an action-effect in the physical world, such as a tone, the evidence of action-effect binding was obtained in one study only if the effect of the action was delayed for no more than 1 s (Elsner & Hommel, 2004). Can social interaction lengthen this time window? If the answer is yes, which factors (e.g., ostensive signals) enlarge it? Regarding contingency, Cook, Press, Dickinson, and Heyes (2010) have demonstrated that the development of the mirror system is sensitive to sensorimotor contingency. If the development solely depended on Hebbian learning (i.e., temporal contiguity), then any observed action would become associated with any performed action only if they occurred together. However, that is not the case. The finding that the development of the mirror system relies on contingency as well as contiguity can explain why most mirror neurons respond to the observation and performance of the same action. Is the same true of intersubjective action-effect binding? Future research is required to answer these questions.



## References

- Ansorge, U. (2003). Spatial Simon effects and compatibility effects induced by observed gaze direction. *Visual Cognition*, **10**, 363-383.
- Bahrick, L. E., & Watson, J. S. (1985). Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology*, **21**, 963-973.
- Bays, P.M., Flanagan, J.R., & Wolpert, D.M. (2006). Attenuation of self-generated tactile sensations is predictive not postdictive. *PLoS Biology*, **4**, e28.
- Blakemore, S-J., Frith, C.D., & Wolpert, D.W. (1999). Spatiotemporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, **11**, 551-559.
- Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial and imitative cues. *Brain and Cognition*, **44**, 124-143.
- Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M., & Heyes, C. (2008). Through the looking glass: counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience*, **28**, 1208-1215.
- Cook, R., Press, C., Dickinson, A., & Heyes, C. M. (2010) The acquisition of automatic imitation is sensitive to sensorimotor contingency? *Journal of Experimental Psychology: Human perception and Performance*, **36**, 840-852.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, **13**, 148-153.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., & Jeannerod, M. (1997). Looking for agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, **65**, 71-86.
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 229-240.
- Elsner, B., & Hommel, B. (2004). Contiguity and contingency in the acquisition of action effects. *Psychological Research*, **68**, 138-154.
- Farroni, T., Csibra, G., Simion, F., & Johnson, M.H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 9602-9605.
- Farroni, T, Mansfield, E. M. Lai, C., & Johnson M. H., (2003). Motion and mutual gaze in directing infants' spatial attention. *Journal Experimental Child Psychology*, **85**, 199-212.
- Guellai, B., & Streri, A. (2011) Cues for early social skills: direct gaze modulates

- newborns' recognition of talking faces. *PLoS ONE*, **6**, e18610.
- Harleß, E. (1861). Der Apparat des Willens [The apparatus of the will]. *Zeitschrift für Philosophie und philosophische Kritik*, **38**, 50-73.
- Heyes, C. M. (2001) Causes and consequences of imitation. *Trends in Cognitive Sciences*, **5**, 253-261.
- Heyes, C. M. (2010) Where do mirror neurons come from? *Neuroscience and Biobehavioural Reviews*, **34**, 575-583.
- James, W. (1890). *The principles of psychology (Vol. 2)*. New York: Dover Publications.
- Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development*, **70**, 474-485.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9096-9101.
- Kühn, S., Elsner, B., Prinz, W., & Brass, M. (2009) Busy doing nothing: Evidence for nonaction-effect binding. *Psychonomic Bulletin and Review*, **16**, 542-549.
- Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, **4**, 50-59.
- Lotze, R.H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.
- Macrae, C.N., Hood, B.M., Milne, A.B., Rowe, A.C., & Mason, M.F. (2002). Are you looking at me? Eye gaze and person perception. *Psychological Science*, **13**, 460-464.
- Mason, M.F., Hood, B.M., & Macrae, C.N. (2004) Look into my eyes: gaze direction and person memory. *Memory*, **12**, 637-643.
- Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, **198**, 75-78.
- Morgan, R., & Rochat, P. (1997). Intermodal Calibration of the body in early infancy. *Ecological Psychology*, **9**, 1-23.
- Murray, L., & Trevarthen, C. (1985). Emotional regulations of interactions between two-month-olds and their mothers. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 177-197). Norwood, NJ: Ablex.
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, **2**, 164-174.

- Neisser, U. (1988). Five Kinds of Self-Knowledge. *Philosophical Psychology*, **1**, 35-59.
- Okanda, M., & Itakura, S. (2008). One-month-old infants' sensitivity to social contingency from mothers and strangers: a pilot study. *Psychological Reports*, **102**, 293-298.
- Paulus, M., van Dam, W., Hunnius, S., Lindemann, O., & Bekkering, H. (2011). Action-effect binding by observational learning. *Psychonomic Bulletin and Review*, **18**, 1022-1028.
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, **50**, 1639-1647.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, **2**, 661-670.
- Sato, A., & Yasuda, A. (2005). Illusion of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, **94**, 241-255.
- Sebanz, N., Bekkering, H., & Knoblich G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, **10**, 70-76.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, **18**, 668-671.
- Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, **13**, 127-134.
- Topál, J., Gergely, G., Miklósi, A., Erdohegyi, A., & Csibra, G. (2008). Infant perseverative errors are induced by pragmatic misinterpretation. *Science*, **321**, 1831-1834
- Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, **96**, 215-231.
- Verschoor, S.A., Weidema, M., Biro, S., & Hommel, B. (2010). Where do action goals come from? Evidence for spontaneous action-effect binding in infants. *Frontiers in Psychology*, **1**: 201.
- von Grünau, M., & Anston, C. (1995). The detection of gaze direction: a stare-in-the-crowd effect. *Perception*, **24**, 1297-1313.
- Wang, Y., Newport, R., & Hamilton, A.F.D.C. (2011). Eye contact enhances mimicry of intransitive hand movements. *Biology Letters*, **7**, 7-10.
- Zorzi, M., Mapelli, D., Rusconi, E., & Umiltà, C. (2003). Automatic spatial coding of perceived gaze direction is revealed by the Simon effect. *Psychonomic Bulletin and Review*, **10**, 1022-1028.

*Review*, **10**, 423-429.

### **Figure Captions**

Figure 1 Schematic representations of the acquisition phase in Experiments 1(a), 2 (b) and 3 (c) and a schematic representation of test phase (d). In the case of (a), if a requested response to “\*” is pressing the left button, the requested response is congruent with the primed response.

Figure 2 Mean reaction times for congruent and incongruent trials in Experiment 1 (eye contact), Experiment 2 (averted gaze), and Experiment 3 (closed eyes). Error bars represent standard error.