Psychological Science

http://pss.sagepub.com/

Facial Self-Imitation : Objective Measurement Reveals No Improvement Without Visual Feedback

Richard Cook, Alan Johnston and Cecilia Heyes

Psychological Science 2013 24: 93 originally published online 29 November 2012

DOI: 10.1177/0956797612452568

Research Report

ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Psychological Science
24(1) 93–98
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797612452568
http://pss.sagepub.com
SAGE

Facial Self-Imitation: Objective Measurement Reveals No Improvement Without Visual Feedback

Richard Cook[1], Alan Johnston[2,3], and Cecilia Heyes[4,5]
[1]Department of Psychology, City University London; [2]Division of Psychology and Language Sciences, University College London; [3]CoMPLEX, University College London; [4]All Souls College, University of Oxford; and [5]Department of Experimental Psychology, University of Oxford

Abstract

Imitation of facial gestures requires the cognitive system to equate the seen-but-unfelt with the felt-but-unseen. Rival accounts propose that this "correspondence problem" is solved either by an innate supramodal mechanism (the active intermodal-mapping, or AIM, model) or by learned, direct links between the corresponding visual and proprioceptive representations of actions (the associative sequence-learning, or ASL, model). Two experiments tested these alternative models using a new technology that permits, for the first time, the automated objective measurement of imitative accuracy. Euclidean distances, measured in image-derived principal component space, were used to quantify the accuracy of adult participants' attempts to replicate their own facial expressions before, during, and after training. Results supported the ASL model. In Experiment 1, participants reliant solely on proprioceptive feedback got progressively worse at self-imitation. In Experiment 2, participants who received visual feedback that did not match their execution of facial gestures also failed to improve. However, in both experiments, groups that received visual feedback contingent on their execution of facial gestures showed progressive improvement.

Keywords

facial expressions, motor processes, associative processes, cognitive development, social cognition

Received 10/10/11; Revision accepted 5/15/12

Studies of the *chameleon effect* confirm what salespeople, confidence tricksters, and Lotharios have long known: Imitating another person's postures and expressions is an important social lubricant; it secures positive evaluation, rapport, and cooperation (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003). But a fundamental question about imitation remains unresolved: How is it accomplished? The problem, often described as the "correspondence problem" (Brass & Heyes, 2005), is most acute in the case of facial expressions. Unless a person is looking in a mirror, his or her own facial expressions are felt but unseen, whereas the facial expressions of others, the targets for imitation, are seen but unfelt (Meltzoff & Decety, 2003).

Two alternative solutions to the correspondence problem currently have empirical support. One solution, embodied in the associative sequence-learning model (ASL), suggests that third-person visual representations of actions are linked directly with corresponding proprioceptive representations, and that these links are acquired associatively during correlated sensorimotor experience (Heyes, 2001, 2010). For example, when infants perform facial gestures repetitively while looking in a mirror, or while being imitated by an adult, they experience contingencies between the gestures they observe and their own performance. Such experience establishes excitatory links between corresponding visual and proprioceptive representations of actions. The alternative solution, embodied in the active intermodal-mapping model (AIM), suggests that facial imitation is mediated by an innate mechanism that allows visual descriptions of facial gestures to be matched to their proprioceptive consequences (Meltzoff & Moore, 1997).[1]

Crucially, the AIM and ASL models make different predictions about the conditions in which imitative performance will show progressive improvement. Both models allow that improvement could result from independent perceptual and motor learning. However, the models make different predictions

Corresponding Author:
Richard Cook, City University London, Department of Psychology, Whiskin Street, London EC1R 0JD, United Kingdom
E-mail: Richard.Cook.1@city.ac.uk

about the kind of experience necessary to enhance the functioning of the core mechanisms thought to solve the correspondence problem. The ASL model suggests that these core mechanisms are direct associations between visual and proprioceptive representations, and that they can be refined only through experience in which performance of actions is paired contingently with accurate third-person visual feedback. In contrast, the AIM model suggests that the core mechanism is a system that explicitly compares vision-derived and proprioception-derived supramodal representations of actions, and that its operation can be enhanced by the execution of actions in the absence of visual feedback (Meltzoff & Moore, 1997).

The present study investigated improvement in facial imitation to test these rival accounts. The accuracy of adults' imitation of facial gestures was measured for the first time using a precise, objective procedure based on automated calculation of euclidean distances. Using these methods, we examined imitative performance over successive attempts. In each of two experiments, one group of participants was given contingent, third-person visual feedback following training attempts, whereas the other group was not. Both models predicted improvement in performance of the visual-feedback groups—because visual feedback provides the opportunity to learn associations between visual and proprioceptive representations of action (ASL), or because visual feedback can be used to refine action plans generated by an innate supramodal mechanism (AIM). However, the AIM model assumes that proprioceptive feedback is sufficient to improve imitation, and therefore predicted improvement in the no-visual-feedback groups as well. In contrast, the ASL model suggests that correlated visual-proprioceptive experience is necessary to improve imitation, and therefore predicted no improvement in the absence of visual feedback.

General Method

First, participants were filmed for 16 s while they recited jokes. The variation present within each sequence was used to construct a measurement space within which the accuracy of self-imitation was subsequently assessed. Next, participants attempted to imitate four target expressions that the experimental program selected from their joke sequence. They were free to take as long as they wished to pose their head and face. When they reached what they perceived to be the optimal pose, they recorded their attempt with the click of a computer mouse. Participants made 18 successive attempts to imitate each target: 4 pretest attempts, 10 training attempts, and 4 posttest attempts. No visual feedback was given after pre- and posttest attempts. The visual-feedback groups received contingent visual feedback after each training attempt, whereas the control groups received an alternative type of visual stimulation after these attempts.

The accuracy of self-imitation was assessed by projecting targets and attempts into a multidimensional space derived from principal component analysis (PCA; Fig. 1). First, an algorithm was applied to each frame from a given participant's joke sequence and to the participant's self-imitation attempts; this algorithm recovered the displacement vector field and texture variation between each frame and the average facial image (Berisha, Johnston, & McOwan, 2010). Next, PCA was used to extract the 50-dimensional structure that most efficiently described the image variance present in the joke sequence. The application of PCA permitted each frame from the joke sequence to be represented as a vector defined by its loadings on the 50 principal components (PCs). Through the application of the same algorithm, it was possible to represent each self-imitation attempt as a vector within the same space. The accuracy of an attempt was quantified by calculating its euclidean distance to the target in PC space. The smaller the distance, the better the attempt.

Experiment 1

Method

In Experiment 1, the performance of the visual-feedback group was compared with that of a no-visual-feedback group who were reliant solely on proprioception. Twenty healthy adults (6 male, 14 female; mean age = 23.9 years) participated in the experiment in return for a small honorarium. Participants were assigned randomly to the two groups in equal numbers. Following the capture of each training attempt, participants in the visual-feedback group were immediately presented with a 5-s acknowledgment display depicting a still image of their attempt alongside the target. To ensure that the two groups received equal exposure to the targets, we presented participants in the no-visual-feedback group with the same acknowledgment displays except that the attempt images were not included. Participants were instructed to replicate the target expressions as closely as possible on each attempt. They were not explicitly informed that the experiment was about imitative improvement, and remained naive to the three phases (pretest, training, posttest) of the design.

Results and discussion

The mean euclidean distances between targets and training attempts (Fig. 2a) were analyzed in an analysis of variance (ANOVA) with training attempt (1–10) as a within-subjects factor and group (visual feedback, no visual feedback) as a between-subjects factor. No main effects of attempt or group were observed (both *p*s > .50). However, the analysis revealed a highly significant interaction between the linear trend observed across training attempts and group, $F(1, 18) = 17.01$, $p < .001$, $\eta^2 = .49$. The visual-feedback group showed a significant linear improvement across attempts, $F(1, 9) = 6.02$, $p < .05$, $\eta^2 = .40$. The mean distance from the target fell from 1,320 on the first training attempt to 1,211 on the final training attempt. In contrast, the no-visual-feedback group showed a significant linear deterioration across training attempts, $F(1, 9) = 14.82$, $p < .01$, $\eta^2 = .62$. The mean distance from the target increased from 1,336 on the first training attempt to 1,397 on the final training attempt.

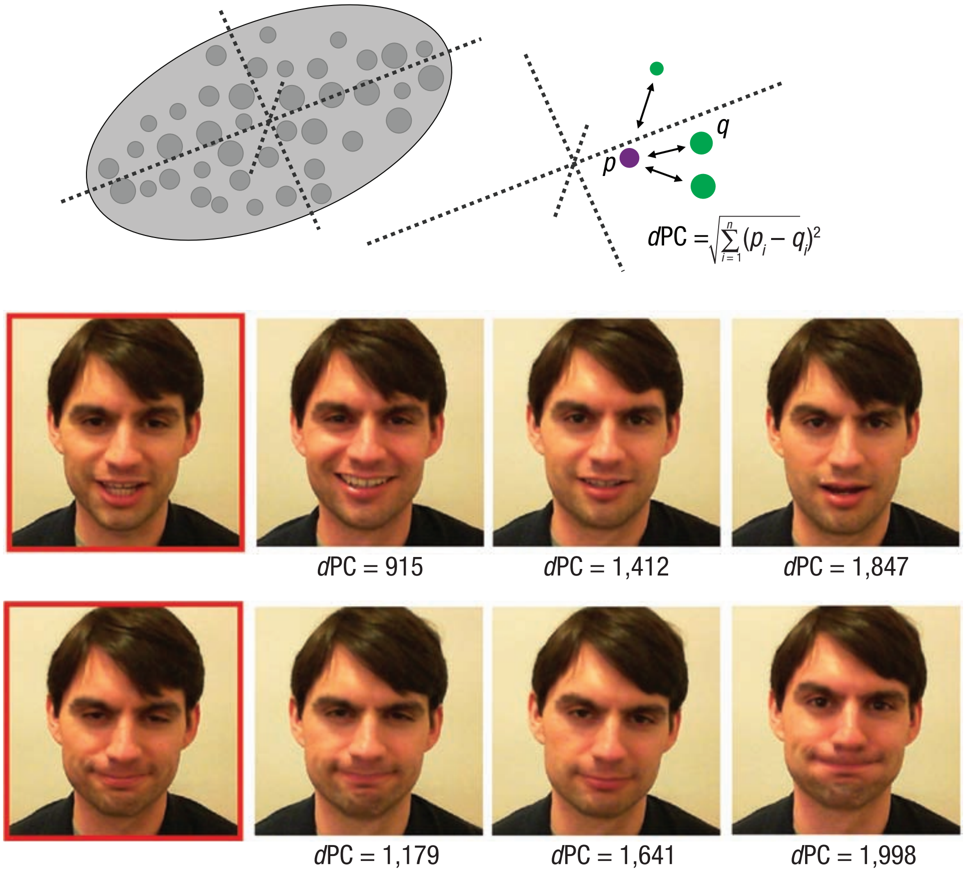

Fig. 1. Illustration of the scoring of accuracy in the two experiments. Principal component analysis was used to extract the orthogonal dimensions that represented the variation (i.e., spatial and texture deviation from the mean) in the facial images in each video sequence most efficiently (top left panel). This allowed a target expression (*p*; purple circle) and attempts to imitate it (*q*; green circles) to be represented as points within a common multidimensional space (top right panel). Within this framework, the accuracy of an attempt is indicated by the euclidean distance between the attempt and the target (denoted by *d*PC): The closer the attempt vector is to the target, the better the attempt. The application of this scoring is demonstrated in the bottom panel. Two targets are presented at the left within red borders; corresponding imitative attempts are shown to the right. The euclidean distance to the target is provided under each attempt.

The mean distances between targets and pre- and posttest attempts (Fig. 2a) were analyzed in an ANOVA with test (pretest, posttest) as a within-subjects factor and group (visual feedback, no visual feedback) as a between-subjects factor. A highly significant group-by-test interaction was observed, $F(1, 18) = 18.02$, $p < .001$, $\eta^2 = .50$. Whereas the attempts of the visual-feedback group were closer to the targets at posttest ($M = 1{,}254$) than at pretest ($M = 1{,}331$), $t(9) = 3.02$, $p < .025$, the attempts of the no-visual-feedback group were further from the targets at posttest ($M = 1{,}416$) than at pretest ($M = 1{,}353$), $t(9) = 3.02$, $p < .025$. No main effects of test or group were observed (both *p*s > .40).

Participants in the visual-feedback group improved the accuracy of their self-imitation during training, whereas the performance of participants reliant solely on proprioceptive feedback deteriorated. This result is consistent with the ASL model, which suggests that the functioning of the core mechanisms of imitation can be enhanced only through the provision of contingent visual feedback.

Experiment 2

On each training trial in Experiment 1, the no-visual-feedback group observed the target expression alone, whereas the visual-feedback group observed the target alongside the feedback image. The ASL model suggests that the feedback image improved performance in the visual-feedback group by virtue of its contingent relationship with the executed expression; this contingency strengthened links between specific visual and proprioceptive representations of task-relevant facial gestures. However, it is possible that the visual feedback improved performance by a different route: through perceptual learning. The comparison between the adjacent attempt and target images may have enhanced participants' visual representations of the target expressions.

To test this perceptual-learning account, in Experiment 2 we compared the performance of a visual-feedback group with that of a noncontingent-feedback group, which on each trial received visual feedback derived from a different attempt. If

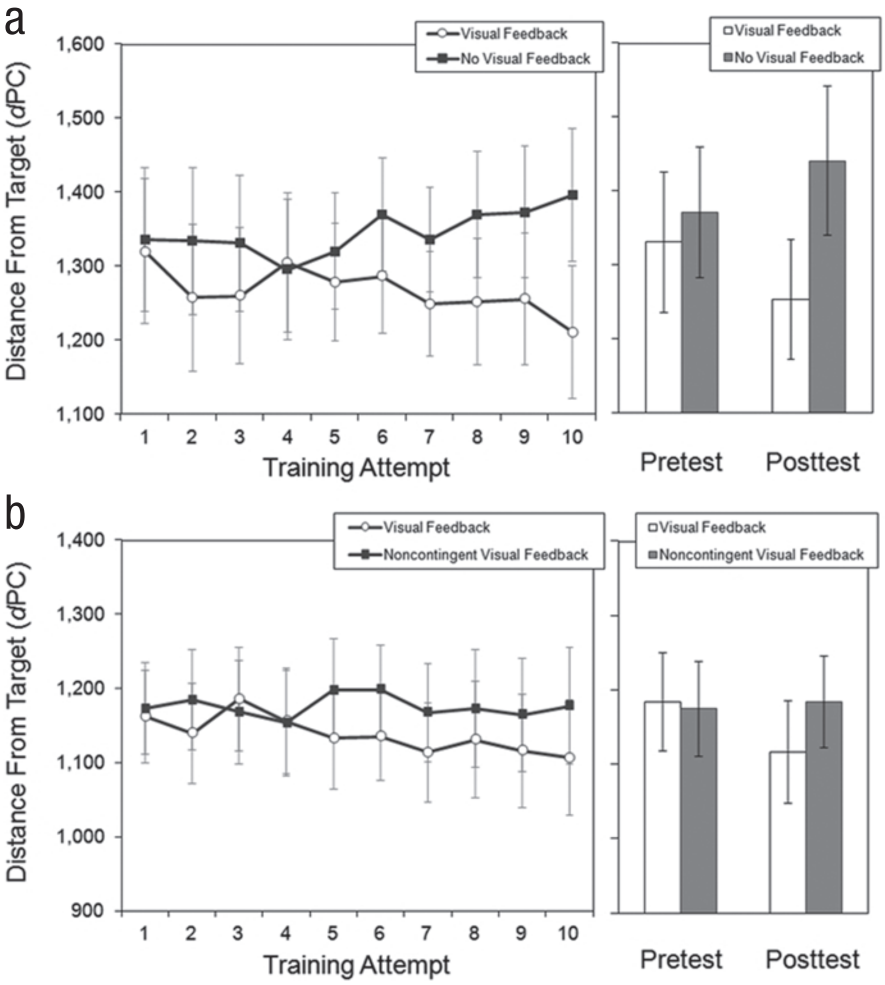

Fig. 2. Mean distances of imitative attempts from the target for (a) the visual-feedback and no-visual-feedback groups in Experiment 1 and (b) the visual-feedback and noncontingent-feedback groups in Experiment 2. The panels on the left depict mean distances across the 10 training attempts. The panels on the right depict mean distances during the pre- and posttest attempts. Smaller distances reflect better self-imitation. Error bars denote ±1 *SEM*.

the group difference observed in Experiment 1 was due to perceptual learning, this difference would disappear in Experiment 2 because both groups saw the target alongside a novel attempt, and therefore the two groups had equal opportunity for perceptual learning. However, if the group difference reflected the strengthening of associations between specific visual and proprioceptive representations of facial expressions, it would be sustained in Experiment 2.

Method

A new group of 20 healthy adults (6 male, 14 female; mean age = 24.2 years) participated in this experiment in return for a small honorarium. Participants were assigned randomly to the two groups in equal numbers. The visual-feedback condition was identical to that used in Experiment 1. Participants in the noncontingent-feedback group also received visual feedback, but from a previous trial, not the current training attempt. On the first training attempt (the fifth attempt overall), the feedback presented was from the fourth attempt ($n - 1$). The feedback presented on the second training attempt (the sixth attempt overall) was from the third attempt ($n - 3$). Thereafter, the feedback presented alternated between the nth – 1 and nth – 3 attempts. This manipulation served to degrade the sensorimotor contingency because execution of the expression was no longer predictive of the visual feedback. Both groups were informed that they would receive feedback “from a recent trial,” and they received identical instructions in all respects.

Results and discussion

The mean euclidean distances between targets and training attempts (Fig. 2b) were analyzed in an ANOVA with training attempt (1–10) as a within-subjects factor and group (visual feedback, noncontingent feedback) as a between-subjects factor.[2] No main effects of attempt or group were observed (both *p*s > .25). However, the analysis revealed a significant interaction between the linear trend observed across training attempts and group, $F(1, 18) = 4.59$, $p < .05$, $\eta^2 = .20$. The visual-feedback group showed a significant linear improvement in their attempts, $F(1, 9) = 15.76$, $p < .01$, $\eta^2 = .64$. The mean distance from the target fell from 1,162 on the first training attempt to 1,107 on the final training attempt. In contrast, the noncontingent-feedback group showed no linear trend across training attempts ($p > .65$).

The mean distances between targets and pre- and posttest attempts (Fig. 2b) were analyzed in an ANOVA with test (pretest, posttest) as a within-subjects factor and group (visual feedback, noncontingent feedback) as a between-subjects factor. A significant group-by-test interaction was observed, $F(1, 18) = 6.59$, $p < .025$, $\eta^2 = .27$. Simple-effects analysis revealed that the attempts of the visual-feedback group were closer at posttest (M = 1,117) than at pretest (M = 1,184), $t(9) = 3.99$, $p < .01$. In contrast, for the noncontingent-feedback group, there was no significant difference between the mean distance from the target at pretest (M = 1,175) and the mean distance from the target at posttest (M = 1,184; $p > .70$). No main effects of test or group were observed (both *p*s > .20).

Because the task was self-paced, we were able to assess participants' motivation to succeed by examining attempt duration. No main effect of group or group-by phase-interaction was observed in either Experiment 1 or Experiment 2 (all *p*s > .35). The absence of group effects suggests that participants who did and did not receive contingent visual feedback were equally motivated.

In summary, participants in the visual-feedback group of Experiment 2 improved the accuracy of their self-imitation during training, whereas participants in the noncontingent-feedback group showed no evidence of improvement. The absence of improvement in the no-visual-feedback group in Experiment 1 could in principle have been due to lack of opportunity for perceptual learning. However, such an account cannot explain the differential learning effects observed in Experiment 2, because the two groups in that experiment viewed the same number of displays containing both a target and an attempt.

General Discussion

The AIM model suggests that the correspondence problem is solved by an innate supramodal mechanism that matches vision-derived and proprioception-derived representations of action. The present findings provide no support for this hypothesis. Rather, the results are consistent with the ASL model. In two experiments, we found incremental improvement in the accuracy of self-imitation in adults who received visual feedback, but no improvement in participants who were reliant solely on proprioception.

According to the ASL model, a person's ability to imitate accurately depends on the sensorimotor experience to which he or she has been exposed. Noisy sensorimotor environments, where the execution of an action predicts the observation of several similar actions, promote associations between broadly matching visual and proprioceptive representations—associations that will detract from precise imitation. Accurate visual feedback, such as that provided to participants in the visual-feedback condition, may refine these associations by strengthening links between precisely matching visual and proprioceptive representations, and by weakening links between approximately matching representations.

The present results do not accord with previous reports of progressive improvement in infants' imitation of midline (Abravanel & Sigafoos, 1984; Johansson, 1973; Soussignan, Courtial, Canet, Danon-Apter, & Nadel, 2011) and sideways (Meltzoff & Moore, 1994) tongue protrusion in the absence of visual feedback. The contrast between these reports and the present findings could be due to several factors. First, the studies of tongue protrusion failed to use objective methods to score performance. For example, Meltzoff and Moore (1994) arbitrarily deemed a large midline tongue protrusion a closer match of sideways tongue protrusion than small sideways protrusions were. Second, these studies confounded accuracy of imitation with response vigor. When repeated execution causes infants' gestures to become more pronounced, scorers may interpret changes over time as increasing accuracy (Ray & Heyes, 2011).

The present study was the first to use an automated objective method, rather than a subjective scoring procedure, to measure facial imitation, and the results favor one of two theories that have previously proved difficult to distinguish empirically. Thus, our results have substantive theoretical, methodological, and practical implications. They suggest that the correspondence problem is solved by direct visual-proprioceptive associations, acquired through correlated sensorimotor experience, rather than an innate supramodal matching mechanism. In addition, our results indicate that euclidean distances, calculated in image-derived PCA space, have significant potential as a means of quantifying the accuracy of facial imitation, and thereby providing insight into the mechanisms that control this behavior. Finally, our results indicate that rehabilitation and skill-training programs designed to improve the quality of the imitation of facial gestures should include contingent visual feedback.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. The ASL model states that links between visual and "motor" representations solve the correspondence problem, and assumes that motor representations have both proprioceptive—i.e., proximal sensory—and executive components. The proprioceptive component is emphasized in the current article because it is the primary focus of

the AIM model, and therefore this emphasis facilitates comparison between the two models.

2. Euclidean distances tended to be smaller in Experiment 2 than in Experiment 1. This difference likely reflects the fact that participants were seated slightly further away from the camera in Experiment 2, to accommodate tall participants more easily. Consequently, expression variation accounted for a greater proportion of the total image variance in Experiment 1.

References

Abravanel, E., & Sigafoos, A. D. (1984). Exploring the presence of imitation during early infancy. *Child Development*, *55*, 381–392.

Berisha, F., Johnston, A., & McOwan, P. W. (2010). Identifying regions that carry the best information about global facial configurations. *Journal of Vision*, *10*(11), Article 27. Retrieved from http://www.journalofvision.org/content/10/11/27.full

Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*, 489–495.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*, 253–261.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*, 575–583.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*, 334–339.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*, 491–500.

Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior & Development*, *17*, 83–99.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, *6*, 179–192.

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, *14*, 92–105.

Soussignan, R., Courtial, A., Canet, P., Danon-Apter, G., & Nadel, J. (2011). Human newborns match tongue protrusion of disembodied human and robotic mouths. *Developmental Science*, *14*, 385–394.