JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2009, **91,** 355–376 NUMBER 3 (MAY)

DO INFANTS SHOW GENERALIZED IMITATION OF GESTURES? II. THE EFFECTS OF SKILLS TRAINING AND MULTIPLE EXEMPLAR MATCHING TRAINING

MIHELA ERJAVEC, VICTORIA E. LOVETT, AND PAULINE J. HORNE

BANGOR UNIVERSITY

The determinants of generalized imitation of manual gestures were investigated in 1- to 2-year-old infants. Eleven infants were first trained eight baseline matching relations; then, four novel gestures that the infants did not match in probe trials were selected as target behaviors. Next, in a generalized imitation test in which matching responses to baseline models were intermittently reinforced, but matching responses to target models were not eligible for reinforcement, the infants matched baseline models but not the majority of their target behaviors. To ensure their failure to match the target behaviors was not due to motor constraints, the infants were trained, in a multiple-baseline procedure, to produce the target responses under stimulus control that did not include an antecedent model of the target behavior. There was no evidence of generalized imitation in subsequent tests. When the infants were next trained to match each target behavior to criterion (tested in extinction) in a multiple-baseline-across-behaviors procedure, only 2 infants continued to match all their targets in subsequent tests; the remaining infants matched only some of them. Seven infants were next given mixed matching training with the target behaviors to criterion (tested in extinction); they subsequently matched these targets without reinforcement when interspersed with trials on which matching responses to baseline models were intermittently reinforced. In repeat tests, administered at 3-week intervals, these 7 children (and 2 that did not take part in mixed matching training) continued to match most of their target behaviors. The results support a trained matching account, but provide no evidence of generalized imitation, in 1- to 2-year-old infants.

Key words: imitation, generalized imitation, infants, trained matching, multiple exemplars, manual gestures

---

Imitation is thought to be one of the driving forces in child development, but there are differing opinions as to its provenance, an issue that is very much at the focus of current developmental, cognitive neuroscience, and behavioral research (for recent reviews of imitation studies in these respective research domains see Hurley & Chater, 2005; Lepage & Theoret, 2007; and Zentall, 2006). A distinctive feature of the behavioral approach is the recognition that calling a behavior imitative does no more than describe the observation that one organism, under certain circumstances, has replicated a behavior of another—it does not explain how both organisms came to perform the same behavior, one after the other. Behavioral accounts of imitation also make the distinction between (i) a repertoire that consists of discrete matching relations each of which is *directly trained* (established through discriminative reinforcement; see Skinner, 1953, pp. 119–120) and (ii) a generative repertoire termed *generalized imitation*, in which new matching relations seem to emerge without training (Catania, 1998, p. 228). Clearly, only the latter kind of imitation repertoire could enable a child to learn new behaviors rapidly and without the need for direct training; therefore, the determinants of generalized imitation as opposed to trained imitation require careful study.

In the typical study of generalized imitation, children are presented with modeling of several actions, a different model on each trial; their matching responses to some of these models receive intermittent reinforcement whereas responses to the remaining probe models do not. Matching of the unreinforced probes is taken as evidence of generalized imitation. In order to determine whether or not the reinforced and nonreinforced responses are members of the same overarching response class, the effects of

We are grateful to the children, their parents, and the staff at the Daycare Nursery and Center for Child Development at Bangor University for their participation in the study. The research reported here was supported by a PhD studentship awarded by the U.K. Economic and Social Research Council to Victoria Lovett.

Correspondence and requests for reprints should be addressed to Mihela Erjavec, School of Psychology, Bangor University, Brigantia Building, Penrallt Road, Bangor, Gwynedd, LL57 2AS, United Kingdom (e-mail: m.erjavec@bangor.ac.uk; web: www.psychology.bangor.ac.uk/playlab).

doi: 10.1901/jeab.2009.91-355

further testing under extinction for matching responses to all modeled behaviors have also been evaluated. Children's responses to the probes appear to be sensitive to the reinforcement contingencies scheduled for matching responses to the remaining target models. When the latter responses are no longer reinforced, matching of all models decreases, including the probes. When reinforcement is reinstated for the subset of matching responses that were previously eligible for reinforcement, matching once again occurs for all target behaviors. Early research suggests that the imitation repertoires of normally developing children exhibit the properties of higher-order behavior classes (Baer & Sherman, 1964; Catania, 1998; Sherman, Clark, & Kelly, 1977).

Over the past four decades, generalized imitation has been studied in normally developing preschool and school age children (e.g., Baer & Sherman, 1964; Baer & Deguchi, 1985; Erjavec & Horne, 2008; Steinman, 1970; Waxler & Yarrow, 1970), children from special populations (e.g., Baer, Peterson, & Sherman, 1967; Garcia, Baer, & Firestone, 1971; Peterson, 1968), and with normally developing infants (Horne & Erjavec, 2007; Poulson & Kymmissis, 1998; Poulson, Kymmissis, Reeve, Andreatos, & Reeve, 1991; Poulson, Kyparissos, Andreatos, Kymmissis, & Parnes, 2002). However, because they do not already have extensive imitation repertoires, it is research with normally developing infants that is best placed to investigate the determinants of generalized imitation in humans.

Infants have been presented with generalized imitation tests in only a few studies. Poulson and her colleagues first reported generalized imitation of object-directed actions, vocal responses, and empty-handed gestures in infants who were between 9 and 18 months old at the start of their experiments (Poulson & Kymmissis, 1998; Poulson et al., 1991, 2002). More recently, however, Horne and Erjavec (2007) found no evidence of generalized imitation of empty-handed gestures in infants who were between 11 and 18 months old. There are a number of differences between the procedures employed in the Horne and Erjavec (2007) study and those employed by Poulson et al. (2002), but the key difference is that Horne and Erjavec first established that the probe target behaviors modeled to the infants did not already feature in their trained matching repertoires, whereas Poulson et al. did not do so. If the criterion for generalized imitation is to be met, we consider it necessary to establish, at the outset of generalized imitation testing, that the target behaviors are novel for the infants in the sense that these behaviors do not feature already in the infants' trained matching relations. Infants' subsequent matching of these demonstrably novel behaviors in the generalized imitation tests would provide evidence of generalized imitation. Conversely, without such a novelty pretest, it is possible that the infants' matching responses to target behavior probes in the generalized imitation tests occur because they feature in matching relations that have already been trained prior to the experiment. Therefore, they ought not to be taken as evidence of generalized imitation. We have presented a thorough discussion of this issue in Horne and Erjavec (2007, p. 65).

When Horne and Erjavec (2007) employed behaviors that did not yet feature in the infants' trained matching repertoires, they found no evidence of generalized imitation: The infants' responses to target models consisted of behaviors that bore only minimal resemblance to the novel modeled gestures. Following the negative results in their initial generalized imitation test, Horne and Erjavec presented a motor skills training condition, in which the infants performed the target behaviors in response to a variety of stimuli that did not include modeling of the corresponding gestures. The infants performed all the relevant target action sequences under these nonmodeling conditions, showing that their generalized imitation test performances were not hampered by motor constraints. Thus the Horne and Erjavec (2007) procedure incorporated safeguards against false positives (accepting previously established trained matching relations as novel imitative responses) and against false negatives (employing target behaviors that were outside of the infants' motor competencies).

Poulson et al. (2002) initially presented models of all behaviors—those destined for matching training in the intervention phase and those that would continue to serve as unreinforced probes—in a baseline phase in which there was no reinforcement delivered for matching responses to any of the modeled behaviors. Effectively this meant that the class

of imitation was under extinction during the baseline trials. Because the novelty of the probes was evaluated under extinction and not in the context of reinforcement for matching of the other interspersed behaviors, it was not possible to determine reliably whether or not the probes already featured in the infants' trained matching repertoires. Therefore, a more parsimonious explanation of the probe matching that Poulson et al. (2002) observed in their intervention phase, when reinforcement was delivered for matches to behaviors other than the probes, is that the infants' matching responses to the probes had already been trained by caregivers prior to the study. This is likely given that the probes consisted of well-practiced behaviors such as clapping, frequently emitted vocalizations, and conventional toy manipulations (and see Horne & Erjavec, 2007, pp. 80–81). If this interpretation is correct, the probe matching that occurred in the intervention phase of their study cannot provide reliable evidence of generalized imitation. Indeed, Poulson et al. (2002) have also mentioned this possibility in their discussion.

We conclude that there is as yet no evidence that infants can imitate novel behaviors. Therefore, the present study was designed to further explore the contingencies that may be effective in establishing generalized imitation in this age group. In our earlier study (Horne & Erjavec, 2007) we trained the infants to match four baseline behaviors before testing the infants' untrained matching of four novel target behaviors. In the generalized imitation tests, the infants did not match the novel target behaviors, but continued to match the intermittently reinforced baseline behaviors. Next, we demonstrated that this failure to match the novel target behaviors was not due to limitations in the infants' motor repertoires. Last, we demonstrated that even after their performance of the target behaviors during the motor skills training the infants still failed to match the target behaviors in further generalized imitation tests. However, it is possible that development of the higher-order class of generalized imitation requires more extensive exemplar training than the four baseline matching relations we established in the 2007 study. For example, it has been reported that extensive multiple exemplar training was necessary to establish generalized imitation in nonimitative and nonverbal children from special populations (Baer et al., 1967; Garcia et al., 1971). In the course of this exemplar training, the children may have learned to discriminate the topographical similarity between each behavior they saw modeled and their own matching responses. This similarity or parity may have become a secondary conditioned reinforcer, which in turn may have established children's matching of novel behaviors without their being explicitly trained to do so (Baer & Deguchi, 1985; Palmer, 1996).

The conditioned reinforcement hypothesis predicts that generalized imitation should emerge only after a child learns a sufficient repertoire of trained matching relations, but the necessary extent and complexity of the latter repertoire remains to be empirically determined. Therefore, in the present study, infants were first trained to match eight baseline gestures each—double the matching training conducted in Horne and Erjavec (2007)—before the initial generalized imitation tests were presented. If training a total of eight baseline matching relations did not result in untrained matching of the four novel target behaviors in the subsequent generalized imitation test then, after checking systematically for any motor constraints as in the 2007 study, we examined the effects of yet further multiple exemplar matching training. This was achieved by successively training matching responses to each of the novel target behaviors, with a generalized imitation test given as each new matching relation was trained. Once the infants matched all four target behaviors, together with the eight baseline behaviors, we investigated whether and for how long the infants would continue to match all 12 behaviors in repeat imitation tests given at 3-week intervals.

As in our previous studies on imitation in infants, we employed a multiple baseline design. Potential confounding sources of control over participants' responses in the generalized imitation tests were minimized by our use of empty-handed gestures as target behaviors, and by keeping the parents unaware of the experimental task and contingencies until the end of the procedures (for further discussion of these procedural issues see Horne & Erjavec, 2007, pp. 65–66). For all generalized imitation tests, infants' perfor-

Table 1

Infants' gender, age at start of each experimental phase and at the end of study, and total number of sessions administered in all phases.

		Age in months/days							
Subject	Gender	Baseline	Probe sessions	Skills training	Staggered matching training	Mixed matching training	Follow-up	End of study	Sessions
Haf	F	13/30	21/28	22/06	23/14	26/10	28/28	28/28	148
Mai	F	14/18	19/24	20/02	22/16	24/01	26/15	28/19	128
Iolo	M	15/15	22/20	22/24	23/27	24/25	27/14	28/06	151
Cat	F	16/02	18/09	20/24	–	–	–	25/03	78
Aled	M	17/08	19/22	19/24	21/14	24/20	26/22	26/22	97
Elin	F	19/01	21/26	22/18	23/21	–	–	24/27	70
Alaw	F	19/06	20/05	20/25	23/02	26/09	29/24	32/24	104
Eleri	F	21/03	22/03	22/25	24/28	–	27/6	27/27	68
Ceri	F	21/22	22/13	22/26	24/30	27/20	31/02	35/09	105
Rhun	M	23/02	23/24	24/07	25/16	–	28/17	33/28	75
Caid	M	24/12	24/20	25/10	26/27	28/29	33/20	36/19	103

mances were evaluated in terms of the relative frequencies of their matching responses and mismatching responses to each of the modeled target behaviors.

METHOD

Subjects

Eleven infants, 7 girls and 4 boys, were recruited from the Daycare Nursery and Centre for Child Development at Bangor University. All infants were judged to be developing normally by the nursery staff and the experimenter; each child attended the Nursery at least three days a week. The infants were aged from 13 months and 30 days to 24 months and 12 days at the start of training; they were assigned brief Welsh code names to preserve anonymity. Table 1 shows the infants' ages at the start of each experimental phase and the number of sessions conducted during the entire study. This research complied with British Psychological Society guidelines for psychological research and was approved by the School of Psychology Ethics Committee at Bangor University.

Setting, Apparatus, and Stimuli

All sessions took place in a purpose-built testing room in the Nursery. To avoid distraction, the room was furnished with only the necessary equipment; it contained a safe infant chair and a small chair for the experimenter, who was seated facing and at the same eye level as the infant. A lidded chest containing age-appropriate toys (e.g., musical books, puzzles, drawing materials, play-dough, seasonal crafts, balloons) and stickers used in play, training, and as reinforcers, was positioned to the side of the experimenter. The toys and stickers employed as reinforcers were items shown to be effective in establishing and maintaining matching relations, over many months of testing, in our previous studies on imitation in infants (Erjavec, 2002; Erjavec & Horne, 2008; Horne & Erjavec, 2007). Two digital wall-mounted cameras recorded the behaviors of the infants and the experimenter, respectively; output from the cameras, together with the sound supplied by a hidden microphone, was stored. The resulting split-screen video recordings could be examined in slow- and stop-motion, as required, for coding.

Modeled target behaviors. The visual stimuli employed were manual gestures performed by the experimenter. Figure 1 (top panel) shows the set of 10 baseline gestures from among which, for each child, 8 were selected for baseline matching training; the same figure (bottom panel) also shows the 8 target gestures. These gestures were selected as targets on the basis of our previous research, which showed that infants in their first and second year of development match few of them in probe generalized imitation tests. Our research has also shown that the behaviors from this target set that are already matched in such probes vary from child to child. For each child, four target behaviors that she or he did not initially match or approximate were employed in the subsequent generalized-imitation tests (see Procedure). The description

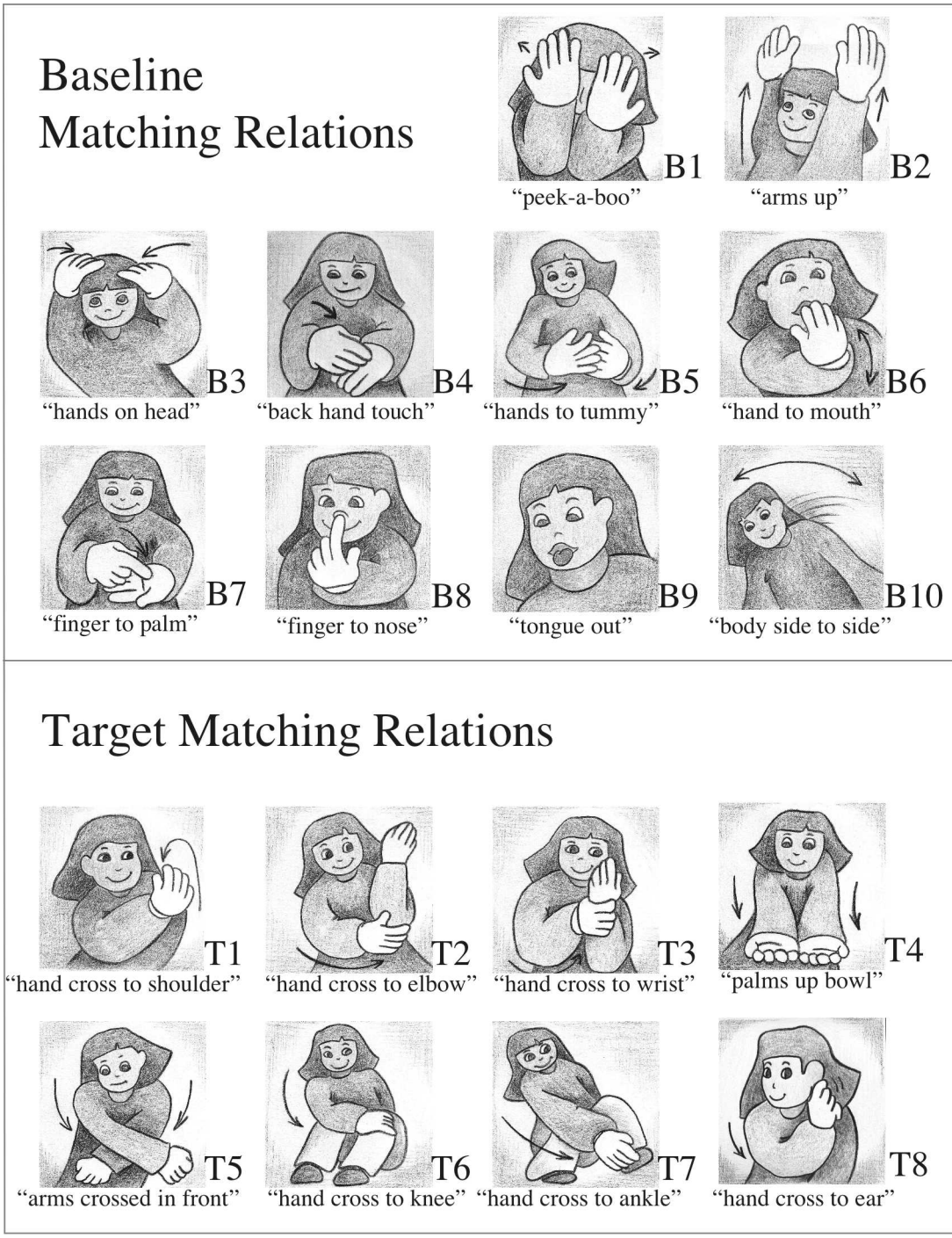

Fig. 1. Range of baseline gestures (B1–B10) that featured in the participants' trained baseline matching relations, and of target gestures (T1–T8) in the target matching relations.

of movements modeled by the experimenter and response variations that met the matching (correct response) criteria are shown in Table 2.

Procedure

A flowchart that summarises the successive phases of the study is presented in Figure 2. Participation in these phases is shown in Table 1 for each subject.

Familiarization. The experimenter established a good rapport with the infants during unstructured daily play, first over several weeks in the Nursery playroom, and next in the test room. During play, the experimenter conducted probe trials in which she modeled eight baseline gestures (B1–B8; see Figure 1), one at a time, and as each was presented, asked the infant, ''Can you do this?'' The experimenter delivered social praise following each matching response. If the infant produced an approximate match to any one of these behaviors, at least once over four trials, then that behavior was selected for inclusion in the baseline matching training phase (see below). If the infant's matching responses to any one of the behaviors did not meet this criterion, it was replaced with a new behavior, selected at random from those shown in Figure 1, and so on, until eight potential baseline matching relations were identified.

In this and all subsequent experimental phases, each session lasted approximately 15 min; infants were tested daily (with inevitable breaks for illnesses and holidays), and all sessions ended with play.

Baseline matching training. For each infant, matching responses to the eight baseline gestures were next trained to criterion. At the start of each session the experimenter asked the infant, ''Shall we play our game?'' At the start of each modeling trial, she established eye contact with the infant, then asked, ''Can you do this?'' as she modeled one of the baseline gestures then looked at the child expectantly. If the infant produced a matching response the experimenter clapped enthusiastically exclaiming, ''Yeah!'' or, ''Well done!'' and immediately delivered a sticker or a toy. If the infant did not respond within 3 s, then the experimenter again prompted her or him to do so by saying, ''You do it!'' or, ''Show me!'' If the infant still produced no response, the model and prompt sequence was repeated; if no response was emitted over three such prompts, the experimenter guided the infant's production of the target response by gently moving the infant's hands into the target matching response configuration, then delivering social praise followed by a sticker or a toy. This ''putting through'' procedure is used routinely to train behavior in applied settings (Striefel, 1981). After each ''putting through'' trial, the infant usually produces some correct, but also some incorrect, components of the reinforced target behavior on the subsequent trial. Consequently, the number of components that have to be guided tends to decrease on subsequent ''putting through'' trials and eventually the infants produce all components of the target behavior without guidance. In this sense, there is an element of shaping over such trials. Nonmatching responses were corrected in a similar manner; the experimenter said, ''Not quite; this is how we do it!'' as she manually guided the infant's correct response, then delivered the reinforcers.

In each training session there were two trials of each of the eight baseline gestures (16 trials per session), with up to three models per trial (as necessary). The gestures were presented in a predetermined randomized order. The criterion was 30 out of 32 unguided, correct responses, across all baseline gesture trials over two consecutive sessions, with the added constraint that there could be no more than one incorrect response for any of the eight baseline gestures or in any one session. When performance met the 100% reinforcement criterion, the reinforcement rate was reduced to 50%. The eight baseline gestures scheduled for reinforcement of matching responses were selected at random prior to each session, with the constraint that the gestures selected should not be the same across consecutive sessions. The intermittent reinforcement criterion was 15 out of 16 correct responses across two trials per gesture within a single session.

Probe sessions and generalized imitation test. In each session, trials with the target gestures (see Figure 1, bottom panel, and *Apparatus, Setting and Stimuli, Modeled target behaviors*) were interspersed, in a prerandomized sequence, with baseline gesture trials. There were 16 trials in total per session, 1 for each of the eight baseline gestures and 2 for each of the four target gestures. The frequency of rein-

Table 2

For each baseline and target gesture, description of movements modeled by experimenter and response variations that met the matching criteria.

Baseline/target gestures	Behavior modeled by experimenter	Accepted response variations
B1 Peek-a-boo	Both hands covering eyes, opening with a ''boo'' sound	Both hands on any part of face
B2 Arms up	Both arms raised above head, stretching and looking up at hands	Both hands at head level or above
B3 Hands on head	Both hands placed on top of head	Both hands on head, touching hair
B4 Back hand touch	Left hand tapping back of right hand	Left/right hand touching back of left/right hand or wrist
B5 Hands to tummy	Both hands tapping tummy	Both hands touching tummy/chest
B6 Hand to mouth	Right hand tapping open mouth to make ''ahh'' sound	Right/left hand touching mouth
B7 Finger to palm	Index finger of right hand touching up-turned palm of left hand	Index finger of one hand touching palm of other hand
B8 Finger to nose	Index finger of right hand touching tip of nose	Index finger of one hand touching nose
B9 Tongue out	Tongue extended fully out of mouth	Tongue extended at least part way out of mouth
B10 Body side to side	Upper part of body swaying from side to side	Marked side to side movement
B11 Hands on chair frame	Both hands tapping infant chair frame	Both hands tapping chair frame
T1 Hand cross to shoulder	Hand touching top of contralateral shoulder	Hand touching top, side, or front of contralateral shoulder
T2 Hand cross to elbow	Hand touching bottom of contralateral elbow of bent and raised arm	Hand touching bottom, front, or side of contalateral elbow with other arm bent at elbow
T3 Hand cross to wrist	Hand touching contralateral wrist of bent arm	Hand touching front or side of contralateral wrist with other arm bent
T4 Palms up bowl	Palms turned up, joined together to form a bowl, extended to front	Both palms turned up, hands in front of body
T5 Arms crossed in front	Arms crossed in front, forearms overlapping, hands fisted	Arms clearly crossed, forearms or wrists overlapping
T6 Hand cross to knee	Hand touching contralateral knee	Hand touching contralateral knee
T7 Hand cross to ankle	Hand touching tip of contralateral ankle	Hand touching contralateral lower leg area between shin and top of foot
T8 Hand cross to ear	Hand touching contralateral ear	Hand touching contralateral ear

forcement for correct responses to trained baseline models was 50%; there were no scheduled consequences (reinforcement or correction) for responses to target models. The target behaviors initially modeled for all infants were as follows: T1, ''hand cross to shoulder'', T2, ''hand cross to elbow'', T4, ''palms up bowl'', and T5, ''crossed arms''. Trials in the initial sessions in this phase served as probes, with the aim to identify and replace the behaviors that already featured in each individual infant's matching repertoire. If the infant responded correctly more than once to one of the target models in the first four trials (i.e., over two sessions), this gesture was replaced in the next session with another selected at random from the target set (see Figure 1, bottom panel). The resulting sets of four target behaviors continued to be presented in generalized imitation test sessions, which were conducted as for the probe sessions, except that none of the target behaviors were replaced in the event that the infant produced the corresponding matching responses. A minimum of five such generalized imitation test sessions was conducted for each subject in the first test phase; testing was subsequently readministered following each training intervention for each target behavior (see below and Figure 2). In all test sessions, the criterion for performance of the baseline gestures was 13 out of 16 (81%) correct over two consecutive sessions. If this criterion was not met, baseline responding was re-established, as in the baseline matching training phase, before the next test session was conducted. Likewise, in the event that an infant was absent for more than two weeks, at least one additional

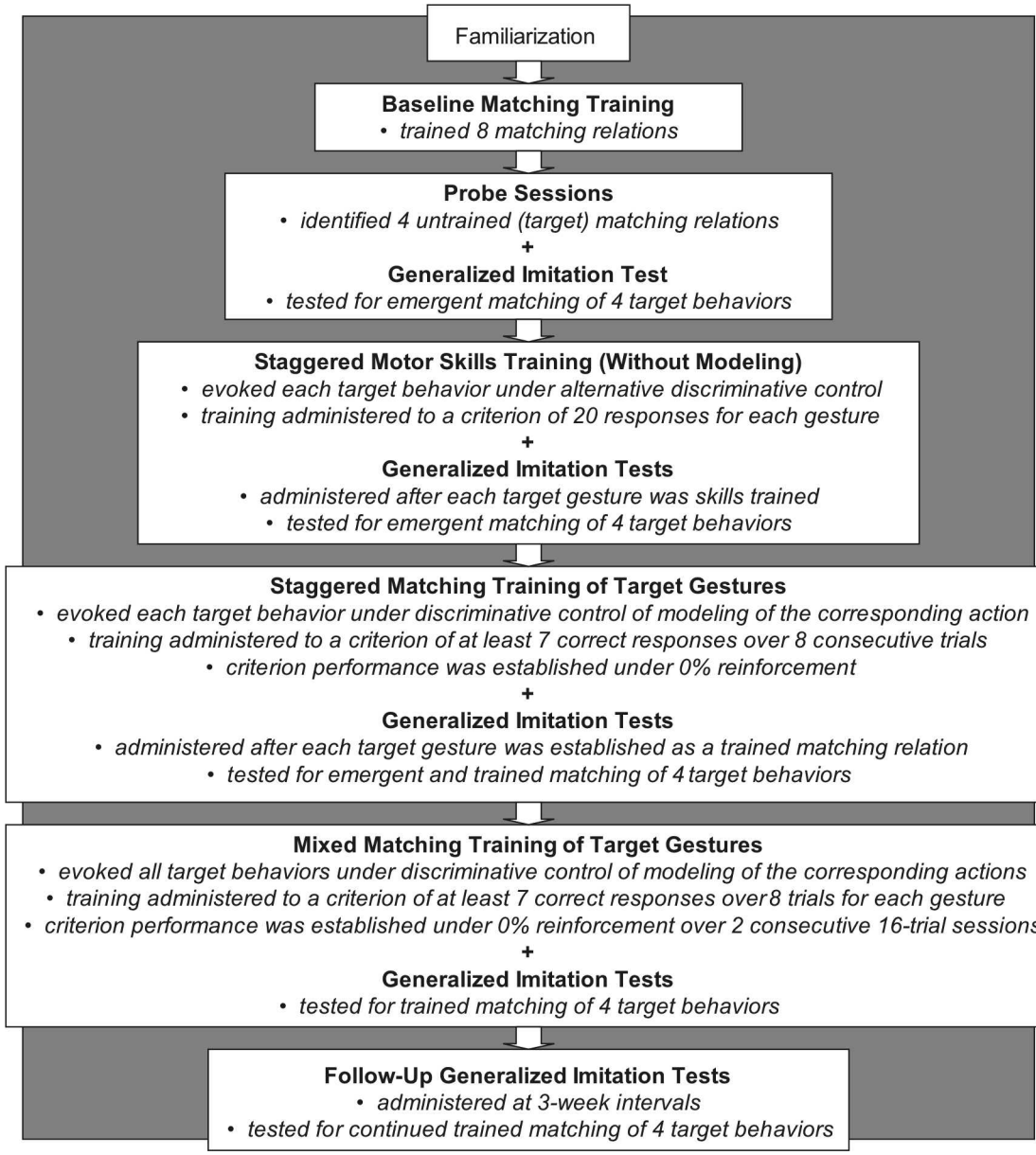

Fig. 2. Flowchart of the experimental conditions.

generalized imitation test session was conducted to determine whether the baselines were maintained or retraining was required before testing could continue. The *overall* scheduled reinforcement percentage in the 16-trial generalized imitation test sessions was 25%.

Staggered target behavior skills training without modeling. The aim of this phase was to demonstrate that the infants were capable of performing each target gesture under stimulus control that did not include modeling of that behavior. During skills training, the experimenter evoked approximations to each target gesture in various ways—through gentle manipulation of infants' limbs (using the ''putting through'' procedure employed in baseline matching training), placing of stickers on relevant body parts, and so on—until

they were reliably produced with no assistance. For example, to occasion T1 (hand cross to shoulder), the experimenter would place a sticker on the child's shoulder and deliver praise when the child removed it. Next, if the experimenter simply touched the child's shoulder the child often followed suit and received praise for doing so. On other trials, the experimenter simply gave the sticker to the child and delivered praise when the child placed the sticker on his or her own shoulder without any prompts to do so. This strategy was also employed for T2, T3, T6, T7 and T8. The cupped hands gesture, T4, was evoked by the experimenter blowing soap bubbles near the child and asking the child to catch them, or on later trials, when the experimenter simply showed the child the bubble bottle. Similarly, T4 was occasioned by throwing a ball for the child to catch, then only showing the ball, or by placing two stickers on the edge of each of the child's upturned palms and asking the child to make them ''kiss''. In the event that the child still did not produce the full target behavior, then ''putting through'' was employed to successively improve the child's performance of the target behaviors. A more detailed description of this part of the procedure can be found in Horne and Erjavec (2007, p. 70; and see Appendix 3 and Appendix 4, where it can be seen that some children eventually produced some of the target behaviors spontaneously, presumably because the experimental context itself had acquired some control over their performance).

To meet criterion in each of the staggered training blocks, the infant was required to perform a target gesture at least 20 times, under alternative stimulus control and without any form of manual guidance by the experimenter. A repeat of the generalized imitation test for all target behaviors was given after each target behavior met the skills training criterion. The tests consisted of either three or five consecutive sessions (following the skills training of the first three target behaviours, and after the last target had been trained, respectively). For 2 infants, Alaw and Ceri, who were absent from the Nursery at the end of this phase, three additional generalized imitation test sessions were administered prior to progression to the next phase. One infant dropped out of the study in this phase: Cat refused to play with the experimenter after completing the skills training of her final gesture, and before the final block of generalized imitation tests could be administered.

Staggered matching training of target behaviors. In this phase, infants were trained to produce each target behavior in response to the experimenter's modeling of that gesture. In multiple baseline training sessions, the experimenter modeled the target action scheduled for matching training, accompanied by a prompt, ''Do this!'' and then in the same way as for previous training phases immediately guided the infant's performance of the target behavior; this response was then reinforced. The experimenter's manual guidance of successive approximations was faded out over trials until the modeled target action alone was sufficient to evoke the infant's performance of the target behavior. The reinforcers were as in the previous phases, and the procedure was the same as in baseline matching training, except that only one of the target responses was trained at a time. Once the infant reliably produced the gesture under continuous reinforcement, the reinforcement rate was reduced to 50% on a variable ratio (VR) 2 schedule, then to 25% on a VR4 schedule. Next, the matching test was administered under extinction conditions. If the infant failed this test the experimenter would increase the reinforcement rate, provide additional training, and re-administer the test. If an infant showed reliable matching of a target behavior in generalized imitation tests administered before this target was scheduled for matching training, the experimenter conducted the matching test immediately.

To meet criterion in each of the staggered matching training blocks, the infant was required to perform a target gesture on seven out of eight consecutive modeling trials, under extinction. A repeat of the generalized imitation test for all target behaviors was given after each target behavior met the matching training criterion. The tests consisted of three consecutive sessions following matching training for each of the first three target behaviors, and five consecutive sessions after the last target had been trained (as in the previous phase). Additional five-session tests were administered to 2 children, Eleri and Rhun, whose matching performance at this stage was stable across all target gestures (see below and Results). Elin completed this phase but left the

Nursery before the next one could be administered.

Mixed matching training of target behaviors. Subjects' matching performances in the last imitation test in the preceding phase determined whether mixed matching training was administered. Those children who emitted less than 90% correct responses over all target trials in the last five-session imitation test were presented with this training (Haf, Mai, Iolo, Aled, Alaw, Ceri, and Caid); the remaining children (Eleri and Rhun), whose performance met criterion, were not. The procedure used in the mixed matching training sessions was similar to baseline matching training; each session consisted of 16 modeling trials, 4 for each target gesture, presented in a randomised order with the added constraint that no more than 2 trials of the same gesture could be presented in succession. The shaping and prompting procedures described earlier were used if necessary. All correct responses were reinforced until the infant's performance met the criterion of at least three correct unaided matching responses over four trials in a single session, for each target gesture. Next, the reinforcement percentage was reduced to 50% (VR2), then to 25% (VR4), using the same progression rule. If the infant produced less than three correct responses on four consecutive trials for each behavior at a reduced reinforcement percentage, then the percentage was increased until the criterion was met, and so on. The matching test was administered under extinction, over two sessions, with eight trials per gesture (32 mixed trials in all). If the infant failed this test, the experimenter increased the reinforcement percentage, provided additional training, and readministered the test.

To meet criterion for completion of mixed matching training, the infant was required to perform each target gesture correctly on seven out of eight modeling trials, under extinction conditions. A repeat of the imitation test for all baseline behaviors and target behaviors was given next, over five consecutive sessions.

Additional training (2) and testing (10) sessions were administered to 1 child, Ceri, whose baseline matching performance became unstable at the end of this phase.

Three-weekly follow-up imitation tests. Long-term stability of subjects' matching performances was investigated in one-session imitation tests administered at 21-day intervals. In each follow up test, the procedure was the same as that employed previously in each generalized imitation test session (see *Probe sessions and generalized imitation test*). The number of follow-up sessions was staggered for 9 infants who continued to attend the Nursery: Rhun (8), Ceri (6), Caid and Alaw (5), Mai (4), Eleri and Iolo (2), Aled and Haf (1).

Coding

The main coding categories for responses emitted in each response period were: target gesture, baseline gesture, other gesture, and no response (i.e., the infant did not move his or her limbs during the response period). For each baseline and target gesture, the formal response criteria are given in Table 2. In addition, we also recorded (i) the number of models (1, 2, or 3) per trial required to evoke a response, (ii) the form of each incorrect gesture, and (iii) whether reinforcement was given in a particular baseline gesture-matching trial, a skills-training trial, or a matching-training trial. Table 2 describes the modeled target behaviors and the corresponding response criteria that were formulated prior to the start of the study (and see Setting, Apparatus, and Stimuli section above). Identical response criteria were employed in all phases of the study. They excluded behaviors commonly produced by infants of this age: turning away, pointing at objects in the room or at the experimenter, vocalizing, trying to remove the chair safety belt, standing up, kicking or hitting the chair, touching clothes, yawning or rubbing eyes, leaning on the chair rest and hiding face, putting fingers or hands in nose or mouth, extending arms to be picked up by the experimenter or trying to touch experimenter, and clapping.

Occasionally, the infants produced more than one response to a modeled gesture in the baseline, probing, training, or testing trials. An incorrect response immediately followed by a correct response was counted as correct (coded as ''self-correction''); conversely, a correct response immediately followed by an incorrect response was counted as incorrect (or ''correct-to-incorrect''; see Horne and Erjavec, 2007). Such multiple responses were very infrequent. In the generalized imitation tests, they occurred on 6% of baseline trials (range: 2–10%) and on 9% of target trials (range: 4–17%); in the matching-training

Table 3

Baseline and target gestures assigned to each subject.

Subject	Baseline gestures	Target gestures
Haf	B1, B2, B3, B4, B6, B7, B8, B10	T2, T5, T6, T7
Mai	B1, B3, B4, B5, B6, B7, B8, B9	T1, T2, T4, T5
Iolo	B1, B3, B4, B5, B6, B7, B8, B10	T2, T4, T5, T7
Cat	B1, B2, B3, B4, B6, B7, B8, B9	T1, T2, T4, T5
Aled	B1, B2, B3, B4, B5, B6, B7, B8	T1, T2, T4, T5
Elin	B1, B2, B3, B5, B6, B7, B8, B10	T1, T6, T7, T8
Alaw	B1, B2, B3, B4, B5, B6, B7, B8	T1, T2, T4, T5
Eleri	B1, B2, B3, B4, B5, B6, B7, B8	T2, T3, T4, T6
Ceri	B1, B2, B3, B4, B5, B6, B7, B8	T1, T2, T5, T7
Rhun	B1, B2, B3, B4, B5, B6, B7, B8	T3, T5, T6, T7
Caid	B1, B2, B3, B4, B5, B6, B7, B8	T1, T2, T5, T7

phases, they occurred on 12% of trials (range: 1–18%). Overall, self-corrections were scored three times more frequently than correct-to-incorrect responses.

Interobserver reliability. An independent rater scored a randomly selected 29% of a total of 15797 trials across the 11 infants. The percentage interobserver agreement was calculated on a point-by-point basis by dividing the number of agreements by the sum of agreements and disagreements and multiplying the product by 100; interobserver agreement was 95% overall with an interobserver agreement of 97% on generalised imitation test trials. On each of the trials selected for reliability checks, the independent raters also recorded whether or not the experimenter (i) modeled the target behavior, and (ii) delivered reinforcers in accordance with the trial schedule for that session. No experimenter errors were found.

RESULTS

Baseline matching training. The matching relations trained for each infant are given in Table 3. For each infant, number of training sessions required was as follows: Caid (3); Ceri, Eleri and Alaw (5); Aled (9); Rhun (10); Elin (13); Cat (16); Haf and Mai (32) and Iolo (39). There was a strong inverse relationship between the infants' ages at the start of baseline training and the number of sessions required to complete the training, Pearson's $r = -.814$, $p < .01$.

Probing for target (novel) behaviors. Participants' matching responses in all test sessions are shown in Figure 3 for Haf and Mai, Figure 4 for Iolo and Aled, Figure 5 for Elin, Cat and Alaw, Figure 6 for Eleri and Ceri, and in Figure 7 for Rhun and Caid. In the first two sessions in which all participants were presented with T1, T2, T4, and T5 as probes for untrained target gestures, 3 infants (Mai, Cat, and Alaw) produced no matching responses, Iolo and Eleri matched T2 once each, Iolo and Aled matched T4 once each, Iolo matched T1 once, and Haf matched T5 once; given that the few matching responses that were produced did not meet the trained matching criterion for the aforementioned infants, none of these targets was replaced. The following children produced two or more matching responses to one or more of the targets, which met the trained matching criterion and so were replaced with other gestures from the target set (see Procedure): Rhun, Haf, Iolo, and Eleri matched T1, Rhun and Elin matched T2, Ceri, Rhun, Haf, Elin, and Caid matched T4, and Eleri and Elin matched T5. In further probing sessions, two of the replacement target gestures were also matched (T3 by Elin and T7 by Eleri) and these were replaced (by T8 and T3, respectively). The target behaviors that were identified during the probe sessions for each infant are shown in Table 3.

Matching of baseline gestures in all generalized imitation tests. In Figures 3, 4, 5, 6, and 7, the infants' matching responses to the eight baseline gestures during the generalized imitation test sessions are represented as open circles, and their matching responses to the four target models are represented by filled circles.

The average baseline matching was maintained at 95% or better for 9 children: Correct responses were recorded on 99% baseline trials for Elin (over 37 sessions) and Rhun (48 sessions); 98% for Mai (43 sessions) and

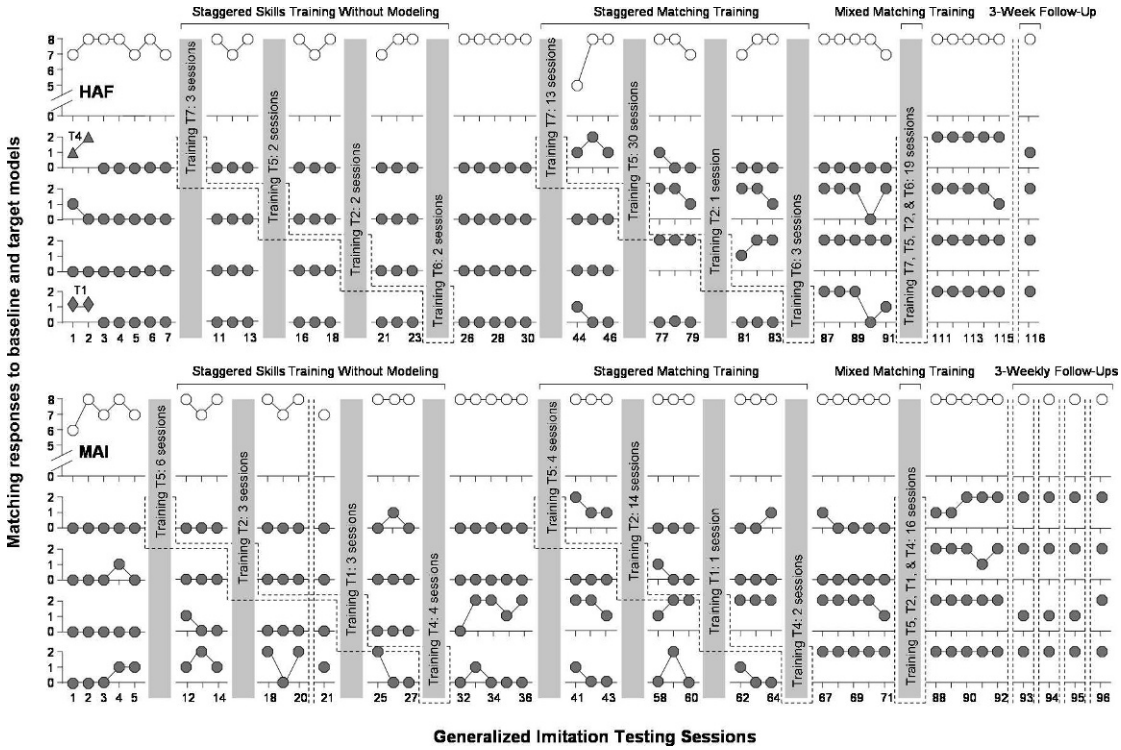

Fig. 3. Participants' matching performances, plotted for Haf and Mai. Responses on the eight baseline trials in all phases are shown as open circles. Responses on the eight target behavior trials in initial probe sessions for Haf are shown as either a filled triangle, or filled diamond, for the two target behaviors (T4 and T1, respectively) that were replaced subsequently because performance met the trained matching criterion, or as filled circles in cases where participants' matching responses to probes did not meet this criterion. Following the probe sessions, all responses in the remaining generalized imitation test sessions are shown as filled circles. Gray-shaded columns show the numbers of training sessions for each of the four untrained target behaviors in the staggered skills, staggered matching, mixed matching, and 3-week follow up phases of the procedure.

Caid (45 sessions); 97% for Alaw (46 sessions) and Eleri (44 sessions); 96% for Haf (41 sessions); 95% for Iolo (42 sessions) and Aled (39 sessions). The remaining 2 children matched the baseline models on 88% (Cat, over 17 sessions) and 80% (Ceri, 59 sessions) of their trials. Baseline retraining was required for only 1 child, Cat (see Figure 5).

Matching of target gestures in generalized imitation tests administered before and after staggered motor skills training. Staggered motor skills training took between 1 and 29 sessions to complete for each target gesture; there was no relationship between the infants' ages at the start of training and the number of sessions required to complete it, Pearson's $r = -.300$, $p > .05$.

Across the 11 infants and 44 target gestures, the mean percentage of target trials with correct matching responses emitted in all generalized imitation test sessions that were administered prior to each gesture's skills training was 12% (range: 1–34%). After the skills training was administered, but before the next training phase commenced, correct target matches were recorded on a mean of 22% of trials (range: 0–64%). Statistically, this difference was not significant, $t = -1.91$, $df = 10$, $p > .05$, as 5 children were more likely to emit some correct target responses prior to skills training (Haf, Mai, Iolo, Elin, and Ceri), whereas the remaining 6 children showed the opposite trend (Cat, Alaw, Eleri, Rhun, Aled, and Caid).

Figures 3, 4, 5, 6, and 7 show individual infants' generalized imitation test performances before and after the motor skills training intervention. In the tests administered before the skills intervention, individual children did not show matching of the target gestures that

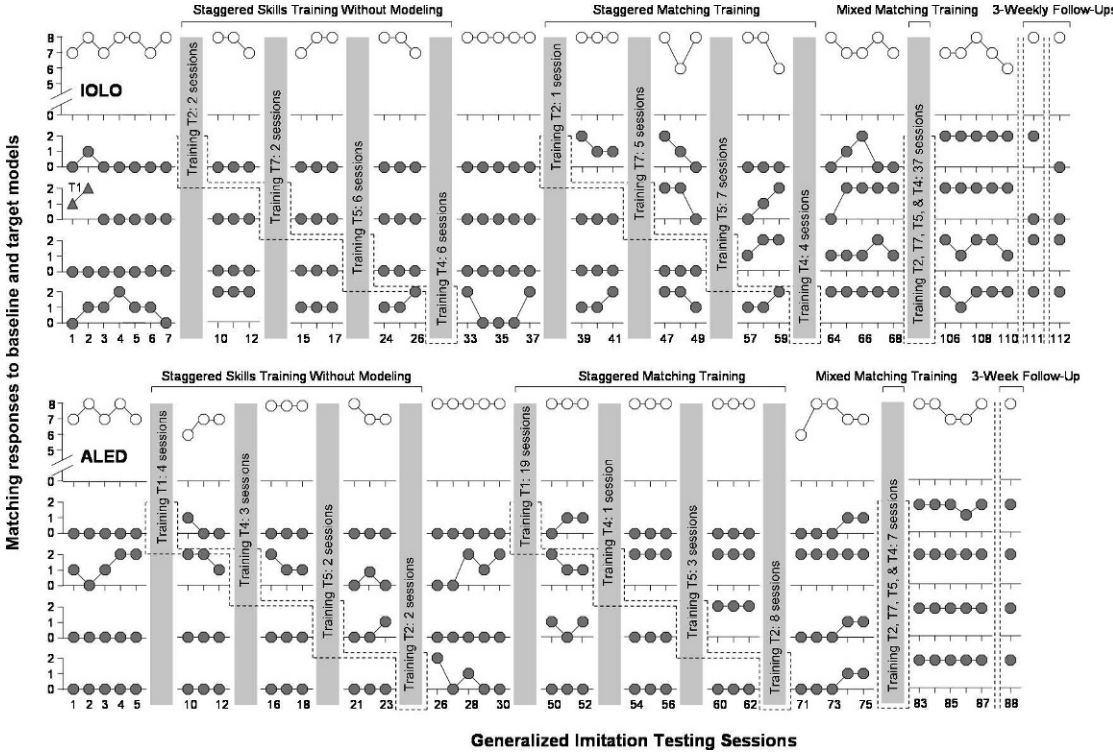

Fig. 4. As in Fig. 3, for participants Iolo and Aled. Iolo's probe responses to T1 are shown as filled triangles.

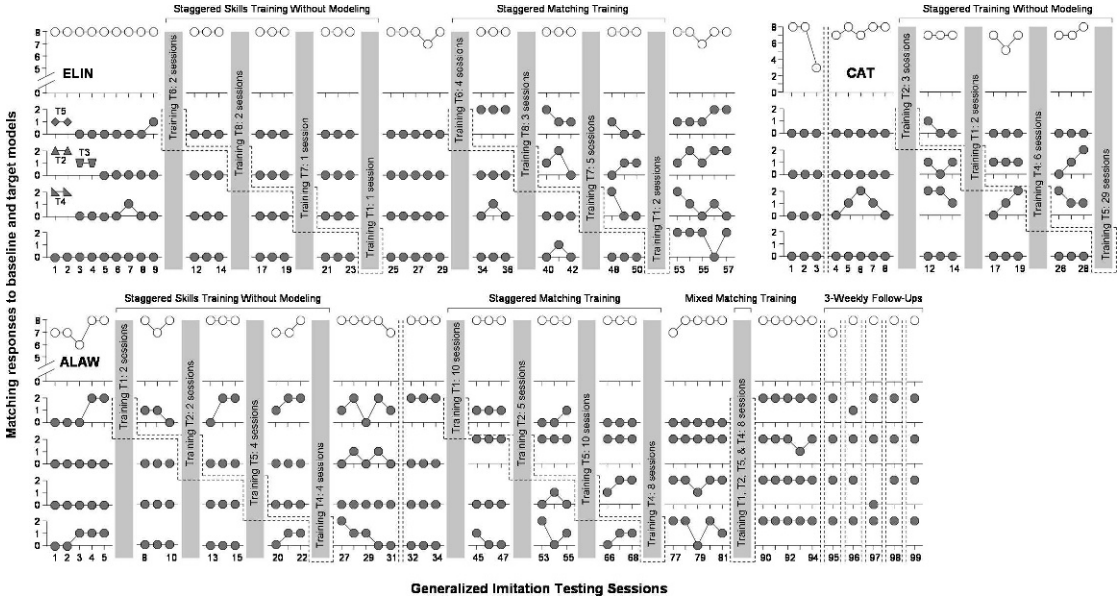

Fig. 5. As in Fig. 3, for participants Elin, Cat, and Alaw. Elin's probe responses to T5, T2, T3, and T4 are shown as filled diamonds, equilateral triangles, rhomboids, and right-angled triangles, respectively.

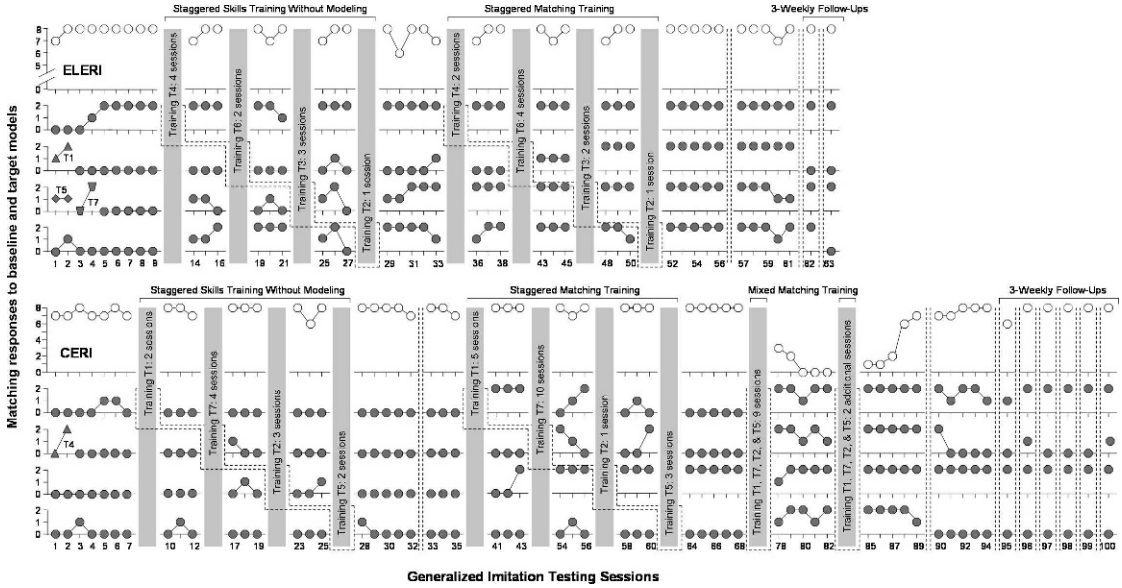

Fig. 6. As in Fig. 3, for participants Eleri and Ceri. Eleri's probe responses to T1, T5, and T7 are shown as filled triangles, diamonds, and rhomboids, respectively. Ceri's probe responses to T4 are shown by filled triangles.

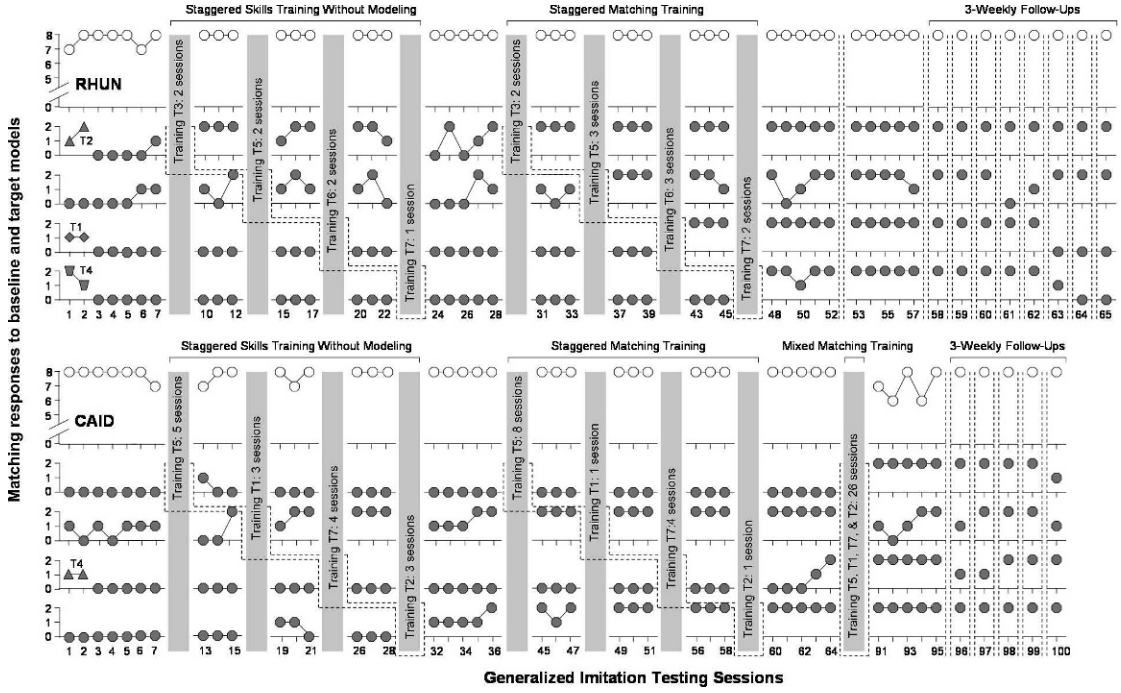

Fig. 7. As in Fig. 3, for participants Rhun and Caid. Rhun's probe responses to T2, T1, and T4 are shown as filled triangles, diamonds, and rhomboids, respectively. Caid's probe responses to T4 are shown as filled triangles.

was comparable to their matching of the baseline gestures. Overall, the children emitted no matching responses for 21 out of 44 target gestures; they matched 10 targets on less than 10% of trials, 6 on 10–25% of trials, 4 on 30–40% of trials, and 3 on 50–70% of trials. In the tests that followed the motor skills intervention, children likewise emitted no matching responses to 17 out of 43 target gestures (Cat's testing terminated at the end of the skills training, before the last generalized imitation test could be administered); 10 targets were matched on less than 10% of trials, 2 on 10–25% of trials, 6 on 30–50% of trials, and 5 on 60–75% of trials. The remaining three targets were matched on 80–93% of trials—comparable to the matching rates shown for the baseline gestures.

The analysis of infants' incorrect responses showed that these were commonly trained gestures (such as tapping hand, mid-arm, ipsilateral body locations, and hugging) and baseline responses on most of the generalized imitation test trials. This was consistent with our earlier findings (Horne & Erjavec, 2007, pp. 84–85). Most children also produced some target responses as mismatches to other target models, showing that discriminative control of such responses was not well established or exclusive—consequently, their occasional matching of these target gestures was not reliable. This was also consistent with our earlier findings (Horne & Erjavec, pp. 74 & 78); in the present generalized imitation tests, targets were mismatched on 14 trials each by Alaw and Eleri; 11 trials by Mai; 8 trials by Caid; 5 trials by Cat, and 1 trial each by Haf and Iolo.

Overall, across 11 infants and 44 target gestures, 10 gestures were matched on more than half of the generalized imitation test trials before or after the training; in all cases, matching commenced before the skills training was administered. For those children who matched some of the targets on more than half of the test trials, the performances were evaluated with reference to their other, non-matching target responses. This showed that 4 children matched one target each, albeit intermittently: Iolo, Aled, and Cat matched T4 and Rhun matched T3, with no mismatches. Only Eleri matched T4 on virtually all the generalized imitation trials. The remaining matching was not reliable: Alaw matched T1 but also emitted this response frequently as a mismatch to T2 models; Eleri matched T3 and T2 but also emitted these responses as mismatches to T2 and T3, respectively; Caid matched T1 and T2 but he also frequently emitted T1 mismatches to T2 models.

To summarise: Individual results and group statistical analysis show that the skills training intervention did not affect infants' matching in the generalized imitation tests—most of the target behaviors were either unmatched or infrequently matched, with no changes that could be attributed to the intervention. Subjects' performances before and after the skills training, considered together, also show that repeated presentation of target models—with up to 48 trials per gesture—was not sufficient to evoke reliable matching in the generalized imitation tests.

Matching of target gestures in imitation tests administered before and after staggered matching training. Ten children took part in this experimental phase. Staggered matching training took between 1 and 30 sessions to complete for each target gesture; the younger infants often took longer to complete the training than the older infants, but this trend failed to reach statistical significance, Pearson's $r = -.577$, $p = .08$.

The mean percentage of target trials on which correct matching responses were recorded prior to matching training, including the final block of imitation tests from the previous phase, was 26% (range: 3–74%). After each target gesture was trained as a match, until the end of this phase, the children matched on 61% of imitation test trials (range: 41–96%). Statistically, this difference in means was significant, $t = -4.80$, $df = 9$; $p < .001$, as 9 out of 10 children were more likely to match target models in the imitation tests after staggered matching training was administered for each behavior than before it, whereas the remaining child (Caid) showed no effect either way.

Figures 3, 4, 5, 6, and 7 show that, for most children and target gestures, matching in the imitation tests that immediately followed staggered matching training of each gesture was better than that recorded prior to the intervention. In the first three test sessions after each target was trained, across 10 children and 40 target gestures, the children matched 19 targets on 100% of trials; 5 on 83% of trials; 5 on 67%; 4 on 50%; 2 on 33%; 1

on 17% of trials; the remaining 4 targets were not matched. The effects of training were in some cases transient, as matching diminished and ceased over repeated trials, but many gestures continued to be matched in the subsequent imitation tests. The analyses of children's other responses showed continued mismatching of target responses in this phase; 29 such responses were recorded for Mai, 8 for Aled, 5 for Iolo, 4 for Elin, 3 for Haf, and 2 for Eleri.

The staggered intervention presented an opportunity to test whether the matching training administered to some of the target responses resulted in the onset of matching for the remaining targets. Individual children's data show no evidence that this happened; the statistical analyses of the percentages of trials on which correct matching responses were emitted for the untrained targets in the generalized imitation tests before the first target was trained, before the second target was trained, before the third target was trained, and before the last target was trained, confirmed that children's matching of novel targets did not change over the generalized imitation tests, $t = 0.94$ to $t = -1.95$; $df = 9$; all ps $> .05$. The statistical analysis also confirmed that, across these tests, correct responses were emitted significantly more often for the just-trained targets than for the remaining, yet-to-be-trained behaviors, $t = 3.50$, $df = 9$; $p < .01$.

In the final block of five imitation test sessions, out of 40 target behaviors, the children matched 14 targets on 100% of trials; 3 on 90% of trials; 4 on 80%; 4 on 70%; 1 on 60%; 1 on 40%; 2 on 30%; 3 on 20%; 1 on 10% of trials; the remaining 7 target behaviors were not matched. Taking into consideration their mismatched target responses, reliable matching on more than half of the trials for each target gesture was now produced by Haf and Elin for 3 targets each, Iolo, Alaw, and Caid for 2 targets each, and Mai, Aled, and Ceri for 1 target each. Two children, Eleri and Rhun, matched all their target models on 100% and 90% of trials in the final imitation test block, respectively. They received additional five-session imitation tests (see Procedure), in which Eleri matched all her targets on 93% of trials, and Rhun performed correctly on 98% of trials.

To summarise: Individual results and group statistical analysis show that staggered matching training improved infants' matching of the trained target actions in the subsequent imitation tests, but did not result in consistent improvement for the remaining, yet-to-be-trained behaviors. This training was sufficient to establish matching in subsequent imitation tests of all target responses, at rates comparable to baseline matching, for 2 out of 10 children; the remaining children showed matching of some, but not all, of their targets.

Matching of target gestures in imitation tests administered before and after mixed matching training. Seven children took part in this experimental phase: Eleri and Rhun were excluded because they showed excellent matching performances at the end of the previous phase, and Elin no longer attended the Nursery. Mixed matching training, in which all target gestures were presented together in a randomised order and matching was trained with decreasing reinforcement rates to criterion tested under extinction conditions, took between 7 and 37 sessions to complete for each infant. There was no relation between the infants' ages at the start of training and the number of sessions required to complete it, Pearson's $r = -.024$, $p > .05$.

The mean percentage correct matching responses recorded prior to mixed matching training, in the final block of imitation tests from the previous phase, was 53% (range: 25–68%). After the mixed matching intervention was administered, in the following block of imitation tests, the children produced correct matching responses on 93% of trials (range: 85–98%). Statistically, this difference in means was significant, $t = -8.06$, $df = 6$, $p < .001$, as all children's matching improved after the mixed matching training intervention.

As can be seen from Figures 3, 4, 5, 6, and 7, all children now matched their target gestures on most or all the test trials. In the final block of five imitation test sessions, out of 28 target behaviors, children matched 19 targets on 100% of trials; 6 on 90% of trials; 2 on 80% of trials; and 1 on 60% of trials. The analysis of children's other responses showed that target mismatches were almost never emitted in these tests (only Ceri, Alaw, and Mai produced one mismatch each).

Ceri's responding to baseline models deteriorated in the imitation test following mixed matching training (13% correct), although her matching of interspersed unreinforced

target models was good (85% correct), so training and testing were re-administered for this child (see Figure 6). Ceri's baseline performance was still low in the second imitation test (40% correct across baseline, 90% correct across targets), although improving over trials; therefore, she was presented with a further five-session imitation test. This was administered after a week's absence from the Nursery, and showed that her performance had deteriorated, as T7 and T5 models were no longer matched (95% correct across baseline, 50% across targets).

Three-weekly follow-up imitation tests. One-session imitation tests, conducted in the same way as for the generalized imitation tests, were administered at 21-day intervals to 9 children who continued to attend the Nursery. Subjects' matching of target models was correct on 100% of trials for Aled (one session); 91% for Mai (four sessions); 88% for Haf (one session), Alaw and Caid (five sessions each); 78% for Rhun (eight sessions); 63% for Iolo and Eleri (two sessions each); and 52% for Ceri (six sessions). While all children continued to match some of the target models, their overall matching rates were usually lower in the follow-up sessions than in the imitation tests immediately preceding this final experimental phase. In their final follow-up test sessions, 5 infants (Haf, Mai, Aled, Alaw, and Caid) continued to match all four of their target gestures, whereas the remaining 4 infants (Ceri, Iolo, Eleri, and Rhun) each matched only two of their targets.

DISCUSSION

The aim of the present study was to test whether 1- to 2-year old infants would show generalized imitation of four novel target gestures that were presented on trials in which matching responses to the targets were not reinforced and these trials were interspersed with trials of eight trained and intermittently reinforced baseline behaviors—twice the number of baseline matching exemplars investigated previously in this population (see Horne & Erjavec, 2007). The behaviors that were selected for baseline matching training were chosen because they commonly feature in the trained matching repertoires of infants (Erjavec 2002; Horne & Erjavec, 2007), and—for individual participants—because at least some approximate matching responses to them were produced in play during the familiarization phase. We found that minimal training was sufficient to meet the baseline matching criteria for the older infants: For these infants, the eight baseline responses were often matched from the first trial, most likely because they already featured in the infants' extraexperimental, trained matching repertoires. By contrast, up to 39 training sessions and gradual shaping of correct responses were necessary to establish reliable baseline matching with the youngest infants, indicating that their matching repertoires were less well developed. This is in line with the Skinnerian account of imitation as a gradually learned repertoire of trained matches, and with our previous findings showing age-related improvements in the matching performances of 2- to 3-year-old children (Erjavec & Horne, 2008). Regardless of the individual differences between the imitative repertoires at the outset of the study, our baseline matching training was very effective for each participant: All infants showed excellent maintenance of the eight baseline responses, over as many as 59 sessions and 472 trials, administered over as much as 13 months of imitation testing.

In the first few probe sessions, four gestures were selected from the target behavior set and presented on trials interspersed with the baseline gestures for which matching responses were intermittently reinforced. Consistent with our earlier research, some infants (4) matched none, but others could already match (or approximate) one (3 infants), or two (2 infants), or three (2 infants) of the four target behaviors. As in Horne and Erjavec (2007), the targets that appeared to be already trained differed from infant to infant, and when these were replaced with another, the new targets in most cases were not matched. At first sight, the fact that some infants matched some of the targets in the first two probe sessions might itself be taken as evidence of generalized imitation. However, there is a logical problem with this argument. Generalized imitation is defined as the imitation of novel behaviors when these are presented on trials interspersed with models of other behaviors that the organism has already been trained to match. Therefore, according to the definition, if the infant matches one of the four target behaviors then he or she should also match

the other three. Moreover, were we to continue to replace each matched probe with another, then these replacement probes should also be matched if we are to describe the infant's performance as generalized imitation. None of the infants' performances met the criterion. One possible caveat might be that some target behaviors were too difficult for the infants to perform, thereby limiting generalized imitation to only some of the probe target behaviors. However, the fact that the targets that were matched differed across infants suggests that infants in the age range tested had the requisite motor skills to perform these particular target behaviors. And the issue of motor constraints at the individual level was addressed directly in the subsequent motor skills training phase, which is considered below.

We found that the infants did not show generalized imitation of the four novel target behaviors in the tests administered either before or after staggered skills training. This replicated our previous findings (Horne & Erjavec, 2007) and showed that exemplar training that established as many as eight baseline matching relations was not sufficient to evoke reliable matching of the four interspersed novel target behaviors, even over repeated tests. Most infants' responses bore only minimal resemblance to the modeled target behaviors; most infants produced few matching responses and some target responses as mismatches to other target models; no infant showed reliable matching of all target models comparable to her or his matching of the baseline behaviors. After as many as 48 trials per target gesture, across 11 infants and 44 targets, only 5 targets were matched, but most of them intermittently, by 1 child each. Such low and selective matching performances could have been the result of synchronous extraexperimental contingencies, such as parental training, and cannot be considered as evidence of generalized imitation (see Horne & Erjavec, 2007).

Our results also confirmed that the infants' performances in the generalized imitation tests were not poor because the target gestures were too difficult for them to perform. First, all target responses were occasionally emitted by the infants in the probe and imitation tests; second, all infants in the motor skills training condition succeeded in producing each of these gestures at least 20 times under alternative stimulus control that did not include modeling of the corresponding behaviors. As in our previous experiments, the skills intervention did not affect infants' matching in the subsequent generalized imitation tests either positively, by increasing the correct performances of target behaviors, or negatively, by setting up response competition between the skills-trained responses and the remaining target and baseline gestures (see Horne & Erjavec, 2007, p. 79). Neither were the infants' performances hampered by their discrimination between the reinforced baseline trials and unreinforced target trials: As in Horne and Erjavec, there was no evidence that any of the infants ceased responding to any of the target models as the testing continued.

Our results show that infants' performances in the generalized imitation tests were not consistent with the conditioned reinforcement hypothesis of Baer and Deguchi (1985), which predicts that infants' matching responses to novel models should gradually increase if the infants produce some approximate matches to the target models on at least some trials because those of their responses that most resemble the target models should be more reinforcing for the participants than other, topographically dissimilar responses. This did not happen in the present study. There was no evidence in any infant's matching performance of a systematic increase in matching responses to the targets; on the contrary, although most infants produced some matching responses to their target models, on subsequent trials their responses usually reverted to incorrect and dissimilar topographies. This happened with some of the targets even after the infants had been trained to match them in later phases of the study, showing that topographical similarity did not play a significant role in either the establishment or the maintenance of infants' responses in the imitation tests.

The staggered matching training intervention, in which infants' production of each target gesture was trained to criterion under discriminative control of the corresponding modeled behavior, resulted in matching of many target gestures in the subsequent imitation tests. This training presented an opportunity to test whether matching training of some target responses would result in gener-

alized imitation of the remaining, yet-to-be-trained behaviors; our results show that this did not happen. Matching training was sufficient to establish and maintain imitation of all four targets at rates comparable to their baseline matching performances in the case of only 2 of the 10 infants. For the remaining participants, matching responses in imitation tests were evoked less reliably; some of the targets were matched only intermittently, a few were no longer matched at all, and mismatched target responses continued to be emitted on some trials. Therefore, the final intervention, mixed matching training, which was designed to establish better discriminative control of each target response by training matching responses concurrently to the four target models, as in baseline matching training, was administered to these children. Following this training, all infants showed imitation of all their target gestures in the subsequent tests. This suggests that in order for them to imitate accurately in generalized imitation test conditions, the infants must be trained to discriminate reliably which response should be produced in the context of which model. Although the infants learned to match each target model to criterion (tested in extinction) during staggered matching training this was not sufficient for the infant to discriminate between the target models with the result that the trained matching performances subsequently deteriorated. The relative merits of sequential and simultaneous matching training procedures could be investigated in future studies to determine whether or not the latter reliably produces more rapid and robust learning. In the follow-up test trials, administered at 3-week intervals, the infants showed good maintenance of most of their trained target responses.

This result is entirely consistent with the account of imitation as trained matching (Skinner, 1953): Extensive matching training of target gestures, over as many as 66 sessions administered over several months, was necessary to establish reliable matching of the four target gestures in the imitation tests. This result is also consistent with our previously reported infant data (Horne & Erjavec, 2007), with our finding that imitative performances of 2- to 3-year old children can be best explained with reference to their histories of matching training (Erjavec & Horne, 2008), and with the results from research with developmentally delayed children (Baer et al., 1967; Lovaas, Berberich, Perloff, & Schaeffer, 1966) which showed that extensive matching training was necessary before imitation of unreinforced target models could be demonstrated in the generalized imitation tests.

It has been reported that children's generalized imitation repertoires may be constrained by topographical boundaries (Garcia et al., 1971; Poulson et al., 2002); it is also possible that vocal imitation, gestural imitation, and imitation of object-directed actions may develop at different rates in infancy. In our present and previous experiments we have used empty-handed manual gestures as targets; these behaviors were carefully chosen to enable reliable coding of infant behavior and to avoid confounding sources of control over infants' responses (see Erjavec & Horne, 2008; Horne & Erjavec, 2007; also see Horne, Erjavec, & Lovett, in press). Future research should determine whether our results would be replicated with other response topographies.

The results of the present study show that typically developing infants require extensive matching training of novel target behaviors before they will reliably match those behaviors in the context of an imitation test in which only a subset of matching relations are eligible for intermittent reinforcement and matching responses to the remainder are never reinforced. By the end of staggered matching training, the infants had received multiple-exemplar training on as many as 12 matching relations in the experimental context, yet this did not establish reliable matching performances. We conclude that the imitative abilities of infants and young children may hitherto have been overestimated in the behavior analytic (e.g., Poulson et al., 2002) and cognitive developmental (e.g., Hurley & Chater, 2005) literature. The conditions under which children may show truly generative imitation of novel responses need to be addressed in future research. In doing this, a careful consideration should be given to the possibility of confounding sources of control over subjects' responses; we suggest that our present experimental design would be well suited for this task. It is possible that, as their verbal repertoires develop, the self-instructional effects of children's naming of the target

behaviors and their responses to them may serve to facilitate onset of more accurate, and ''emergent'' matching repertoires (Horne & Lowe, 1996, 1997; Lowe & Horne, 1996). Indeed, one strand of our current research is focused on the potential interplay between naming and the matching repertoire in young children.

REFERENCES

Baer, D. M., & Deguchi, H. (1985). Generalized imitation from a radical-behavioral view-point. In S. Reiss, & R. Bootzin (Eds.), *Theoretical issues in behavior therapy* (pp. 179–217). New York: Academic Press.

Baer, D. M., & Sherman, J. A. (1964). Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology, 1*, 37–49.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior, 10*, 405–416.

Catania, A. C. (1998). *Learning*. New Jersey: Prentice hall.

Erjavec, M. (2002). Determinants of gestural imitation in young children. Unpublished doctoral dissertation, University of Wales, Bangor, UK.

Erjavec, M., & Horne, P. J. (2008). Determinants of imitation of hand-to-body gestures in 2- and 3-year old children. *Journal of the Experimental Analysis of Behavior, 89*, 183–207.

Garcia, E., Baer, D. M., & Firestone, I. (1971). The development of generalized imitation within topographically determined boundaries. *Journal of Applied Behavior Analysis, 4*, 101–112.

Horne, P. J., & Erjavec, M. (2007). Do infants show generalized imitation of gestures? *Journal of the Experimental Analysis of Behavior, 87*, 63–87.

Horne, P. J., Erjavec, M., & Lovett, V. E. (in press). The effects of modelling, local stimulus enhancement, and affordance demonstration on the production of object-directed actions in 6-month-old infants. *British Journal of Developmental Psychology*.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior (Target Article). *Journal of the Experimental Analysis of Behavior, 65*, 185–241.

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior, 68*, 271–296.

Hurley, S., & Chater, N. (2005). *Perspectives on imitation: From neuroscience to social science*. London: The MIT Press.

Lepage, J. F., & Theoret, H. (2007). The mirror neuron system: Grasping others' actions from birth? *Developmental Science, 10*, 513–529.

Lovaas, O. I., Berberich, J. P., Perloff, B. F., & Schaeffer, B. (1966). Acquisition of imitative speech by schizophrenic children. *Science, 151*, 705–707.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65*, 315–353.

Palmer, D. C. (1996). Achieving parity: The role of automatic reinforcement. *Journal of the Experimental Analysis of Behavior, 65*, 289–290.

Peterson, R. F. (1968). Some experiments on the organization of a class of imitative behaviours. *Journal of Applied Behavior Analysis, 1*, 225–235.

Poulson, C. L., & Kymmissis, E. (1998). Generalized imitation in infants. *Journal of Experimental Child Psychology, 46*, 324–336.

Poulson, C. L., Kymmissis, E., Reeve, K. F., Andreatos, M., & Reeve, L. (1991). Generalized vocal imitation in infants. *Journal of Experimental Child Psychology, 51*, 267–279.

Poulson, C. L., Kyparissos, N., Andreatos, M., Kymmissis, E., & Parnes, M. (2002). Generalized imitation within three response classes in typically developing infants. *Journal of Experimental Child Psychology, 81*, 341–357.

Sherman, J. A., Clark, H. B., & Kelly, K. K. (1977). Imitative behavior in preschool children: The effects of reinforcement, instructions, and response similarity. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, method, and application. In honor of Sidney W. Bijou* (pp. 503–529). Hillsdale, NJ: Erlbaum.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Steinman, W. M. (1970). The social control of generalized imitation. *Journal of Applied Behavior Analysis, 3*, 159–167.

Striefel, S. (1981). *How to teach through modeling and imitation*. Austin, Texas: Pro-Ed.

Waxler, C. Z., & Yarrow, M. R. (1970). Factors affecting imitative learning in preschool children. *Journal of Experimental Child Psychology, 9*, 115–130.

Zentall, T. R. (2006). Imitation: definitions, evidence, and mechanisms. *Animal Cognition, 9*, 335–353.

Received: July 3, 2008
Final Acceptance: January 23, 2009

APPENDIX 1

Child: Target (Freq. correct)	Antecedent events that successfully evoked each named target behavior in motor skills training
Haf: T2 (28)	(i) Stickers (*placed by experimenter [E] on a body part of child*) + VER (*Verbal request; e.g., E asking, "Where is your sticker," or, "Where shall we put this sticker," or, "Can you make these stickers kiss?"*); (ii) Stickers; (iii) Stickers + VER + E touch (*experimenter touch to a child's body part*); (iv) Stickers + VER + PT (*child repeats action after E "puts her through" the required movement sequence*)
Haf: T5 (28)	(i) Stickers + VER; (ii) Stickers + VER + E touch; (iii) Stickers + VER + PT; (iv) Stickers + E touch
Haf: T6 (26)	(i) Stickers + VER; (ii) Stickers + VER + E touch; (iii) Stickers; (iv) Stickers + VER + PT
Haf: T7 (29)	(i) Stickers + VER; (ii) Placing of stickers (*child is given a sticker by E and places it on a body part*) + VER;(iii) Stickers + VER + PT; (iii) Stickers + VER + E touch
Mai: T1 (31)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + E touch; (iv) Anticipation of stickers (*E shows sticker and child touches a body part where stickers were previously found*) + VER; (v) Spontaneous (*antecedent not determined*; *performed by child in free play*)
Mai: T2 (29)	(i) Stickers + VER + PT; (ii) Stickers; (iii) Stickers + VER; (iv) Anticipation of stickers + VER; (v) Stickers + VER +E touch
Mai: T4 (31)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers + VER + E touch; (iv) Anticipation of stickers + VER
Mai: T5 (36)	(i) Stickers + VER + PT; (ii) Stickers + VER; (iii) Stickers + VER + E touch; (iv) Stickers
Iolo: T2 (35)	(i) Stickers; (ii) Stickers + VER; (iii) Stickers + VER + E touch; (iv) Stickers + VER + PT
Iolo: T4 (29)	(i) Bubbles (*E blows bubbles for child to catch*); (ii) Stickers + VER; (iii) Stickers + VER + PT; (iv) Bubbles; (v) Anticipation of bubbles (*child performs target behavior when E shows bubble bottle*); (vi) Ball (*E throws/drops a ball for child to catch*)
Iolo: T5 (32)	(i) Stickers + VER +PT; (ii) PT + VER; (iii) Stickers + VER; (iv) VER
Iolo: T7 (29)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + E touch;(iv) Placing of stickers + VER
Cat: T1 (27)	(i) Stickers; (ii) Stickers + VER
Cat: T2 (28)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers+ VER + PT; (iv) Stickers + VER + E touch; (v) Spontaneous
Cat: T4 (30)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers + VER + E touch; (iv) Bubbles + VER; (v) Bubbles + VER + E touch
Cat: T5 (24)	(i) Stickers + VER; (ii) Stickers + VER + E touch; (iii) Stickers + VER + PT
Aled: T1 (32)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + PT; (iv) Anticipation of stickers
Aled: T2 (29)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + PT; (iv) Stickers + VER + E touch; (v) Spontaneous
Aled: T4 (32)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers + VER + E touch
Aled: T5 (28)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers; (iv) Stickers + VER + E touch
Child: Target (Freq. correct)	Antecedent events that successfully evoked each named target behavior in motor skills training
Elin: T1 (24)	(i) Stickers + VER; (ii) Stickers; (iii) Peg (*E touches child body part with a peg*); (iv) Peg + VER; (v) Stickers + VER + E touch
Elin: T6 (23)	(i) Stickers + VER; (ii) Stickers; (iii) Play dough (*E puts play dough on child's body part; child removes it)* + VER*;* (iv) Toy (*E puts toy on child's body part; child removes it*) + VER; (v) Stickers + VER + E touch; (vi) Play dough
Elin: T7 (23)	(i) Stickers + VER; (ii) Play dough; (iii) Stickers
Elin: T8 (32)	(i) Stickers + VER; (ii) Placing stickers; (iii) Stickers + VER + E touch; (iv) Play dough + VER + E touch
Alaw: T5 (21)	(i) Stickers + VER + PT; (ii) Stickers + VER + E touch; (iii) Stickers + PT
Eleri: T2 (24)	(i) Stickers + VER; (ii) Placing of stickers; (iii) Placing of stickers + VER; (iv) Anticipation of stickers (v) Spontaneous
Eleri: T3 (35)	(i) Stickers + VER; (ii) Placing of stickers + VER
Eleri: T4 (38)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers
Eleri: T6 (26)	(i) Stickers + VER; (ii) Stickers + VER + E touch; (iii) Stickers + VER + PT
Ceri: T1 (33)	(i) Stickers + VER; (ii) Stickers
Ceri: T2 (29)	(i) Stickers + VER; (ii) Placing of stickers + VER; (iii) Anticipation of stickers + VER; (iv) Placing of stickers; (v) Stickers
Ceri: T5 (34)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers; (iv) Stickers + VER + PT + E touch
Ceri: T7 (22)	(i) Stickers + VER; (ii) Stickers
Rhun: T3 (32)	(i) Stickers + VER; (ii) Stickers; (iii) VER; (iv) Anticipation of stickers; (v) Stickers + VER + PT
Rhun: T5 (34)	(i) Stickers + VER; (ii) Stickers+ VER + PT; (iv) Stickers
Rhun: T6 (32)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + E touch
Rhun: T7 (40)	(i) Stickers + VER; (ii) Stickers; (iii) Placing of stickers

APPENDIX 1

(Continued)

Child: Target (Freq. correct)	Antecedent events that successfully evoked each named target behavior in motor skills training
Caid: T1 (34)	(i) Stickers + VER; (ii) Stickers; (iii) Placing of stickers + VER; (iv) Placing of stickers (v) Anticipation of stickers
Caid: T2 (38)	(i) Stickers + VER; (ii) Stickers; (iii) Stickers + VER + PT; (iv) Placing of stickers
Caid: T5 (30)	(i) Stickers + VER; (ii) Stickers + VER + PT; (iii) Stickers; (iv) Stickers + VER + E touch
Caid: T7 (32)	(i) Stickers; (ii) Stickers + VER; (iii) Stickers + VER + E touch; (iv) Stickers + VER + PT; (v) Placing of stickers