### Supplementary Information

#### Participants

Participants were recruited from the greater Boston area, and informed consent was obtained from their parents following guidelines approved by the Harvard University Committee on the Use of Human Subjects. Sample sizes were chosen in advance, based on similar paradigms in prior studies (i.e., Kinzler, Dupoux & Spelke, 2007; Hamlin, Wynn & Bloom, 2007). Participants who did not meet the prespecified inclusion criteria (see below) were excluded and replaced. Participant demographics and exclusions are given in Table S1.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Included N</th>
<th>Age Range</th>
<th>Mean Age</th>
<th>Excluded N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>48 (20 F), 24 per condition</td>
<td>4 mos, 1 day – 5 mos, 14 days</td>
<td>4 mos, 20 days</td>
<td>7 (2 F, 5 IAP)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>24 (13 F)</td>
<td>4 mos, 2 days – 5 mos, 14 days</td>
<td>4 mos, 24 days</td>
<td>4 (1 F, 1 IAI, 1 IAP)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>72 (35 F), 24 per condition</td>
<td>4 mos, 0 days – 5 mos, 12 days</td>
<td>4 mos, 19 days</td>
<td>22 (3 F, 1 PI, 4 IAI, 14 IAP)</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>16 (10 F)</td>
<td>11 mos, 19 days – 13 mos, 13 days</td>
<td>12 mos, 14 days</td>
<td>4 (2 F, 2 NR)</td>
</tr>
</tbody>
</table>

**Table S1.** Information on participants in each experiment. The final column tallies the number of excluded participants for each experiment and the reasons for exclusion (F – excessive fussiness that prohibited participation; PI – parental interference; IAI – inattentiveness to the interaction phase [see definition in Exp. 1 supplemental methods]; IAP – inattentiveness to the preferential looking test [see definition in Exp. 1 supplemental methods]; NR – no visually guided reach to a single character [see definition in Exp. 4 supplemental methods].)

Approach event data for the majority of the participants in the target and responder approach conditions of Experiments 1 and 3 overlap with data reported in Experiment 4 of Powell and Spelke (2018), in which we analyzed looking times to the approach events, but not to the interactions or preference tests (i.e. these data come from the same participants, and the same experimental session). A small amount of non-overlap in the samples for
each condition resulted from cases in which either (1) an infant did not meet the attentional criterion for the preferential looking test, and was thus excluded from and replaced in the current experiments but not Powell & Spelke (2018), or (2) an infant did not watch the critical sequences in the approach events, and was thus excluded from and replaced in Powell & Spelke (2018) but not the current experiments. The number of infants who do not overlap between the current sample and the one reported in Powell & Spelke (2018) for one of these two reasons are as follows: Experiment 1, responder approach – 5; Experiment 1, target approach – 3; Experiment 3, responder approach – 3; Experiment 3, target approach – 5. No infants provided data to Experiments 2 or 4 (or the “no approach” condition of Experiment 3) that overlapped with any other experiments.

**Reliability Coding**

For Experiments 1-3, looking times to interaction events, approach events, and preferential looking tests were coded by a condition-blind offline coder, as described below. For each phase, 20-25% of participants across the three experiments were coded by an additional blind coder, and the correlations between the coders’ times were calculated as a measure of reliability. Correlations were 0.94, 0.99, and 0.93 for the interaction, approach, and preferential looking phases, respectively.

For Experiment 4, a display-blind offline coder judged which of the two 3D test objects the infant touched first during the preferential reaching test. The experimenter who presented the figures also made an independent judgment of first touch while still blind to the identity of the imitator in 14 of 16 cases, and agreement between these two sets of judgments was 100%.

**Supplementary Methods and Results**
**Experiment 1.**

**Procedure.** Infants were positioned in a car seat approximately 60 cm from a 40 x 60 cm screen. Parents, seated behind the infants, were instructed not to speak or to look at the screen and were monitored for compliance (see exclusion rate for "PI" in Table S1). The animated events were created and displayed using Keynote and featured three circular characters, each 6.5 cm in diameter, with schematic faces: red and blue characters located on the sides of the screen, near the top two corners, and a purple character centered in the bottom half of the screen (Figure 1, main text). Each infant viewed two rounds of six events: four interaction events, followed by two approach events (see descriptions in main text and Movie S1 for one such round). Each event was preceded by a verbal cue from the experimenter (“Look, [baby's name]!”). Looking times to the interaction events were coded offline by a condition-blind coder. Participants who did not attend to the display screen during some portion of both interaction partners' actions during at least two interaction events of each type (i.e. imitative and non-imitative) across the two rounds were excluded for inattentiveness to the interaction events (see exclusion rate for “IAI” in Table S1).

The approach events were labeled as “imitation-congruent” when they involved the target and imitator and “imitation-incongruent” when they involved the target and non-imitator. Display-blind experimenters initiated coding according to a knocking sound at the beginning of each approach event that was common across all trials, and the looking thresholds for ending trials were set prior to the start of data collection. When analyzing looking times to the approach events, we excluded trials for which infants missed the portion of the display during which one character approached another, as well as those for which infants looked less than 0.5 s following the onset of the “dance” portion of the
display. Because infants spent significantly more time overall looking to the first pair of approach events than the second pair, if looking to one event was excluded on either of these grounds, we also excluded the looking time to the other approach event in that pair (i.e. if the first imitation-congruent event had to be excluded, we also excluded the first imitation-incongruent event, occurring in the same round of interaction and approach events). As a result of these exclusion criteria, two infants in the responder approach condition and one infant in the target approach condition had no valid approach event pairs and were thus excluded from analyses for this phase only. (In Powell & Spelke, 2018, these infants were replaced by additional participants, but here we only replaced participants excluded from the preferential looking analyses reported in the main text.) We log-transformed the raw looking times for valid approach trials to deal with skew common in distributions of infant-controlled looking times (Csibra, Hernik, Mascaro, Tatone & Lengyel, 2016), as in the supplementary analyses of Powell & Spelke (2018).

We then compared mean log-transformed looking times to the imitation-congruent and –incongruent approach events for each condition. In analyses of the largely overlapping data sets in Experiment 4 of Powell & Spelke (2018), we found evidence that infants expected responding individuals to approach targets they imitated over ones they did not, but also that these expectations were graded, and depended upon the complexity of the displays and the age of the infants. The specific configurations of the responder approach conditions presented here (as well as the one reported below, in Experiment 3) occupied a middle ground in terms of the complexity of displays we presented, and did not elicit differential looking to congruent and incongruent approaches. More complex and simpler displays elicited significantly longer looking to congruent and to incongruent
approaches by imitators, respectively. Target approach conditions, in contrast, consistently yielded null results regardless of display complexity (Powell & Spelke, 2018). We thus expected to find equivalent looking times to congruent and incongruent approaches in both the responder and target conditions of the current experiment, when analyzing data sets that largely overlapped with those previously reported in Powell & Spelke (2018).

When administering the preferential looking test, the experimenter ensured that the infant looked at each physical character, centered the figures in the infant’s visual field and then asked “Who do you like?” while moving each figure outward approximately 20 cm in opposite directions, bowing her head, and closing her eyes. The video of each preference test was coded frame-by-frame (at a rate of 30 frames/second). We determined in advance that we would exclude and replace any infant who did not look at the red and blue characters for at least 1 s each during the test, to ensure infants’ continued awareness that both characters were present (see exclusion rates for “IAP” in Table S1). These looking times were not log transformed prior to analyses. Csibra and colleagues’ (2016) analysis demonstrating skew in large samples of infant looking time data dealt specifically with infant-controlled looks to event outcomes, not with preferential looking to two potential targets over a fixed duration. We did not think it likely that the same skew would be found in data from preferential looking tests, both because the fixed duration of such tests truncates the potential for an extended tail of exceptionally long looking times and because exceedingly long looks to one target are necessarily balanced by exceedingly short looks to the other target. We thus saw no reason to expect that sample variance would deviate from
a normal distribution, and tests conducted with the ANOVAs indicated homogeneity of variance.

**Counterbalancing.** Within each approach condition, we counterbalanced which side character imitated the center character (red, right-side character vs. blue, left-side character), which sound was imitated (high vs. low pitched), the order of imitative and non-imitative interactions (imitation first or second), and the order of approach events (incongruent first or second).

**Supplementary results.** In addition to the main effect of longer looking to non-imitative than imitative interactions, there was also a three-way interaction between interaction type, interaction order, and imitated sound ($F(1,32) = 4.27, P < 0.05$). Neither of these effects was found in the subsequent experiments using the same or similar interactions, however.

A repeated-measures ANOVA comparing log-transformed looking times to the two approach trial types (imitation-congruent and -incongruent), with interaction order (imitation first or second), approach order (incongruent first or second), and approach type (responder vs. target approach) as between-subjects factors revealed an interaction between trial type and approach order, $F(1,37) = 4.86, P < 0.05$, reflecting longer looking to the trial type shown first ($M = 1.31$) vs. second ($M = 1.22$). There was no main effect of trial type, $F(1,37) = 1.27, P > 0.2$, nor an interaction between trial type and approach type, $F(1,37) = 0.32, P > 0.5$. Paired sample t-tests confirmed the lack of differentiation between congruent and incongruent approaches in both the responder approach condition (congruent $M = 1.34$; incongruent $M = 1.28$; $t(21) = 0.90, P > 0.3$) and the target approach condition (imitation congruent $M = 1.23$; imitation incongruent $M = 1.20$; $t(22) = 0.56, P > 0.5$).
These statistical conclusions are the same as those reported for the largely overlapping samples presented in Powell & Spelke (2018).

There was no evidence that infants’ looking patterns during the preferential looking test were driven by differential looking to events involving the imitator and non-imitator during the interaction or approach phases, as looking time to the congruent and incongruent approach events did not differ and difference scores from the preferential looking test were uncorrelated with those contrasting looking to imitative vs. non-imitative interactions ($r(46) = -0.02$) and looking to incongruent vs congruent approach events ($r(43) = -0.03$).

**Experiment 2.**

**Procedure.** Infants viewed eight contiguous interaction events, alternating between imitative and non-imitative responses, followed by the same preference test conducted in Experiment 1. In order to maintain the presence of infant-controlled pauses and processing time prior to the preference test, we inserted infant-controlled pauses following each pair of interaction events. As in Experiment 1, these pauses lasted until online coding indicated that the participant had looked for 60 s cumulatively or looked away for 2 s consecutively. As in Experiment 1, the identity of the imitating character (red, right hand character vs. blue, left hand character), the sound that was imitated (high vs. low pitched), and the order in which the imitative and non-imitative interactions occurred during familiarization (imitation first or second) were counterbalanced across participants.

**Supplementary Results.** On the final preference test, there were longer combined looking times when the red, right-side character was the imitator ($M = 12.27$ s) than when the blue, left-side character was the imitator ($M = 9.93$ s, $F(1,16) = 5.23, P < 0.05$).
Individual differences in looking to the imitator vs. the non-imitator were not correlated with differential looking to the imitative and non-imitative interactions, $r(22) = 0.17$.

By independent samples t-tests comparing preferential looking difference scores, the preference for the imitator in Experiment 2 was as strong as in the responder approach condition of Experiment 1 ($t(46) = 0.65, P > 0.5$), and stronger than in the target approach condition of Experiment 1 ($t(46) = 2.18, P < 0.05$). A repeated-measures ANOVA with approach condition (responder, target, or no approach), interaction order (imitation first or second), imitator identity (red, right-side vs. blue, left-side character), and imitated sound (high or low pitched) as between-subjects factors, revealed a significant main effect of looking to the imitator vs. non-imitator, $F(1,48) = 10.97, P < 0.005$, an interaction of that effect with approach condition ($F(1,48) = 3.92, P < 0.05$, and the same between-subjects effect found in Experiment 2 (longer overall looking to both characters during the preference test when the imitator was the right-side character [$M = 12.56$ s] vs. the left-side character [$M = 10.90$ s; $F(1,48) = 5.91, P < 0.05$].

**Experiment 3.**

**Procedure.** All conditions began with familiarization events similar to those featured in Experiments 1 and 2, except that the order of actions within each interaction was reversed (see Movie S2).

The identity of the imitating character (red, right-side vs. blue, left-side character), the sound that was imitated (high vs. low), and the order in which the imitative and non-imitative interactions occurred during familiarization (imitation first or second) were counterbalanced across all conditions. In the two conditions including an approach phase, we also counterbalanced the order of the approach events (incongruent first vs. second).
Supplementary Results. For looking times to approach events, the procedure for excluding event pairs and participants from analyses was the same as in Experiment 1 and resulted in the exclusion of two participants in the target approach condition. For remaining participants in the target and responder approach conditions, we compared mean log-transformed looking times to the two trial types (imitation-congruent and – incongruent approach events) using a repeated-measures ANOVA, with interaction order (imitation first or second), approach order (incongruent first or second) and approach type (responder or target) as between-subjects factors. As in Experiment 1, there was a significant trial type x approach order interaction, $F(1,38) = 12.73, P < 0.005$, reflecting longer looking to the trial type that came first ($M = 1.33$) versus second ($M = 1.21$). There was no main effect of trial type, $F(1,38) = 0.41, P > 0.5$, and no interaction between trial type and approach type, $F(1,38) = 2.73, P > 0.1$. Paired sample t tests confirmed that infants did not differentiate the event types in either condition (responder approach – imitation-congruent $M = 1.23$; imitation-incongruent $M = 1.30$; $t(23) = 1.53, P = 0.14$; target approach – imitation-congruent $M = 1.29$; imitation-incongruent $M = 1.25$; $t(21) = 0.77, P > 0.4$).

These statistical conclusions are again the same as those reported for the largely overlapping samples and reported in Experiment 4 of Powell and Spelke (2018), though in the corresponding responder approach condition of the prior report, infants’ tendency to look longer at approaches by the responder toward the non-target (i.e. incongruent approaches) trended more strongly toward statistical significance. A follow-up condition employed the same displays used in this condition (i.e. one responder who alternately approached the target it imitated and the one it did not), but consolidated the interaction
and approach events, rather than alternating the two phases (i.e. it presented 8 interactions followed by 4 approach events, instead of the 4-2-4-2 design used here), in order to improve infants’ encoding of the interaction events prior to viewing approach events. This change resulted in significantly longer looking times to the imitation-incongruent approaches (Powell & Spelke, 2018, Experiment 5).

The key finding from Experiment 4 of Powell & Spelke (2018) was that the three-character displays elicited less relative preference for congruent responder approach events than more complex five-character displays. We checked if the substitution of a small number of participants to meet inclusion criteria for the preference tests affected this finding by repeating the ANOVA comparing looking times from three-character responder approach conditions (the current data sets for responder approach conditions in Experiments 1 and 3) to looking times from five-character responder approach conditions (Powell & Spelke, 2018, Experiments 1a and 2b). We kept the same parameters for the analysis as the previous experiment, including total number of characters (three vs. five), number of approaching characters (one vs. two), order of interaction events (imitation first or second), and order of approach events (incongruent first of second) as between subjects factors. As in the previous report, we observed a significant interaction between congruent/incongruent looking and total number of characters, $F(1,62) = 6.29, P < 0.05$, showing that reducing the complexity of the displays reduced infants’ preferential attention to the congruent events. In contrast, an ANOVA comparing the current target approach conditions from both Experiments 1 and 3 to the five-character target approach conditions (Powell & Spelke, 2018, Experiments 2a and 2b) found no interaction with character number, $F(1,61) = 0.66, P > 0.4$, consistent with the analyses of the overlapping
sample reported in the previous paper and with the conclusion that infants do not
differentiate target approaches regardless of complexity.

In the analysis of looking patterns on the final preference test, there was no
interaction between preferential looking and approach condition, and paired samples t-
tests within each condition also failed to find evidence of a preference in any one
(responder approach – imitated target: 6.17 s; non-imitated target: 5.82 s; $t(23) = 0.40, P >$
0.6; initiator approach – imitated target: 6.69 s; non-imitated target: 6.71 s; $t(23) = 0.01, P >$
0.9; no approach – imitated target: 6.46 s; non-imitated target: 5.55 s; $t(23) = 1.12, P > 0.2$;
Figure 2b).

The repeated-measures ANOVA comparing preferential looking to the characters
involved in imitative vs. non-imitative interactions across the conditions eliciting a
preference for imitators in Experiments 1 and 2 (imitator approach and no approach),
relative to the analogous conditions of Experiment 3 (target approach and no approach),
revealed an interaction between that variable and the position and/or features of the side
character involved in the imitative interaction, $F(1,88) = 8.41, P < 0.01$: infants looked
longer to the imitator or imitated target when that role was assumed by the right-side
character rather than the left-side character. This side bias did not interact with condition
and did not account for infants’ imitator preference in Experiments 1 and 2. (Infants in
those experiments looked longer to the imitator regardless of its identity [red imitator vs.
blue non-imitator: 8.19 s vs. 4.29 s; blue imitator vs. red non-imitator: 6.25 s vs. 4.85 s],
while infants in Experiment 3 looked longer to the right side, regardless of that character’s
role in the interactions [red target vs. blue non-target: 6.73 s vs. 5.21 s; blue target vs. red
non-target: 6.42 s vs. 7.04 s].) Elevated attention to the right side character is consistent
with previous reports of a right visual field bias in young infants in a supine position (Turkewitz, Gordon & Birch, 1965; Michel, 1981); our use of an inclined car seat to position infants in front of the screen may thus have encouraged the observation of this bias.

**Experiment 4.**

**Procedure.** The procedure was based closely on past studies of infants’ preferences for helpful social partners (Hamlin et al., 2007), with the notable exception that the familiarization events were presented via animated displays rather than a live action play (see Movie S3). All three characters made the same sound when initiating movements up or down the hills. The identity of the imitator and non-imitator (the blue or yellow characters) and the order of the familiarization events (imitation or non-imitation first) were counterbalanced across participants.

For the critical test of infants’ preference for an imitator vs. a non-imitator, we presented infants with the same type of physical copies of the relevant characters used in the first three experiments and observed which they chose to reach for first. Prior to moving the characters close enough for infants to reach them, a display-blind experimenter verified that participants had looked at each one and then attracted their attention back to the center, making eye contact. After moving the characters within the infants’ reach, the experimenter tracked their gaze and motion, to determine when the infant made a visually guided reach to one of the characters. If an infant attempted to reach for both characters simultaneously or reached toward one character without also looking at it, the experimenter pulled the characters back and repeated the presentation, moving the characters farther apart to prevent infants from reaching for both characters at once. If the infant was hesitant and did not reach at all after 20 s, the experimenter repeated the
presentation, moving the characters closer to the infant. If the infant repeatedly failed to make a visually guided reach to a single character across four consecutive presentations his or her data were excluded from the final sample and replaced with another participant (see “NR” in Table S1).

References


