Supplementary Materials 1- Pre-test details

The goal of the pre-test was to select stimuli, for the main experiment, which children can map gestures too most easily.

The pre-test used 28 videos, ranging 6-11 seconds, and depicting 14 different manners. These manners were organised into seven video groups (1-7). For each group, each manner resulted in two different end states. There were also, therefore, 14 different end states (2 per group). Thus, each group consisted of two video pairs: Pair 1, manner A resulting in end state A, manner B resulting in end state B; pair 2, manner A resulting in end state B and manner B resulting in end state A. These pairs of videos were shown together in a two-way forced choice task.

Sixteen 3-year-olds (8 females, M= 40.94months, SD=3.28) took part in the pre-test. Two children were removed from the analysis: one for attentional problems and one for a side bias (only selecting videos on one side of the screen). Participants were shown two videos playing simultaneously, side by side. Videos playing together were matched for length, within 1second. These videos were pairs taken from a video group. Counterbalancing ensured that all videos were tested equally, in both manner gesture and end state gesture trials.

Participants were then shown a gesture (either manner or end state) and asked '*which one is like this? (plus iconic gesture)*'. Participants' task was to point to the video they felt best matched the gesture.

The performances for end state and manner gestures were compared using a *t*-test. The results showed that there was no effect of gesture type (*p*>.05). Next the

overall proportion of correct responses was compared to chance (0.5) using a *t*-test. The results revealed that children performed better than chance, $M = .64$, $SD = .12$, t (13) = 4.40, $p=.001$. The results were then split by gesture type and the comparison to chance was repeated. The results showed that children could match end state gestures significantly better than chance, $M = .69$, $SD = .22$, t (13) =3.25, $p=.006$), and match the manner gestures better than chance descriptively but not significantly ($M = .59$, $SD = .19$, t (13) =1.67, $p=.119$).

The groups that would be used in the main study were chosen based on the descriptive results obtained, Performance for manner gestures for two video groups were descriptively poorer than other gestures so these were removed. The analyses were then rerun without these groups in. First, the performances for end state gestures ($M = .69$, $SD = .24$) and manner gestures ($M = .64$, $SD = .26$) were compared using a *t*-test. The results showed that there was no effect of the type of gesture used (t (13) =.416, $p>.05$). Finally, the proportion of correct responses was compared to chance (0.5). Results revealed that children could match the end state gestures significantly better than chance, t (13) = 2.842, $p=.014$ and manner gestures marginally better than chance, t (13) = 2.002, $p=.067$. Therefore, the main experiment used five of the original seven video groups.

Supplementary Materials 2- Results across stimuli

We investigated how consistent the results were across stimulus items. First we considered consistency across the ten manners and then across the ten end states. For each comparison, we looked at the performance of the three gesture groups when each of the ten manners/ end states were used as the training video.

Consistency across manners

If we look at the pattern for all the gesture groups for the ten manners shown in the video stimuli, we see that the manner gesture group selected the same manner video at test more often than either of the other gesture groups for 8/10 manners. The end state gesture group only once selected more same manner videos that the manner gesture group. For seven of the manners, the no gesture group selected more same manner videos than the end state gesture group. See Figure 1 for the descriptive results.
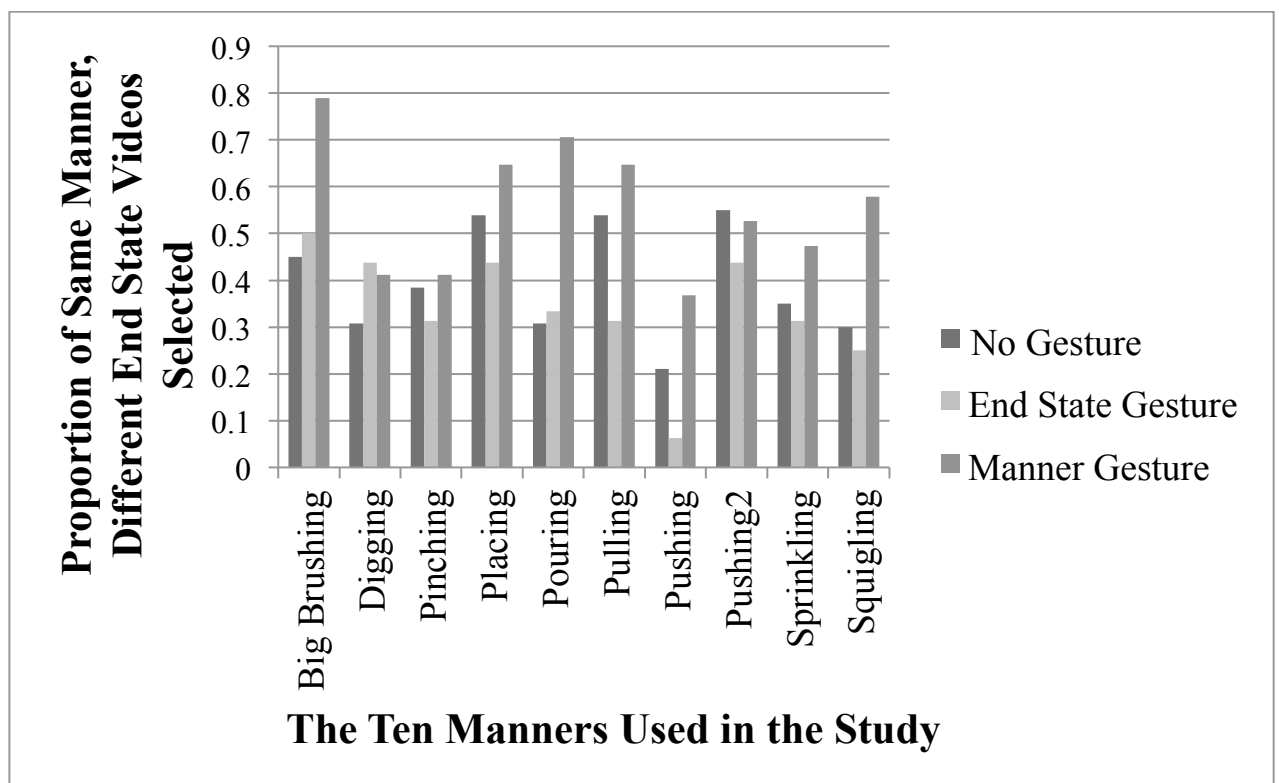

Figure 1. The proportion of trials in which the same manner videos were selected (i.e., generalization based on manner) by the three gesture groups for the ten manners used in the study.

This pattern shows that across the ten manners, the results mimic those found in the main results. Specifically, for 80% of the manners, children who saw manner

gestures were generally more likely than the other two gesture groups to select the same manner video at test.

Consistency across end states

If we look at the pattern for all the gesture groups across all ten end states, we see that the manner gesture group selected the same-manner video more often than the other two gesture groups for 8/10 of the manners, and the end state gesture group selected the same-manner video less often than the manner gesture group for 9/10 of the end states. For 5/10 end states, the end state gesture group selected the same manner video less than either of the other two gesture groups. See Figure 2 for the descriptive results.
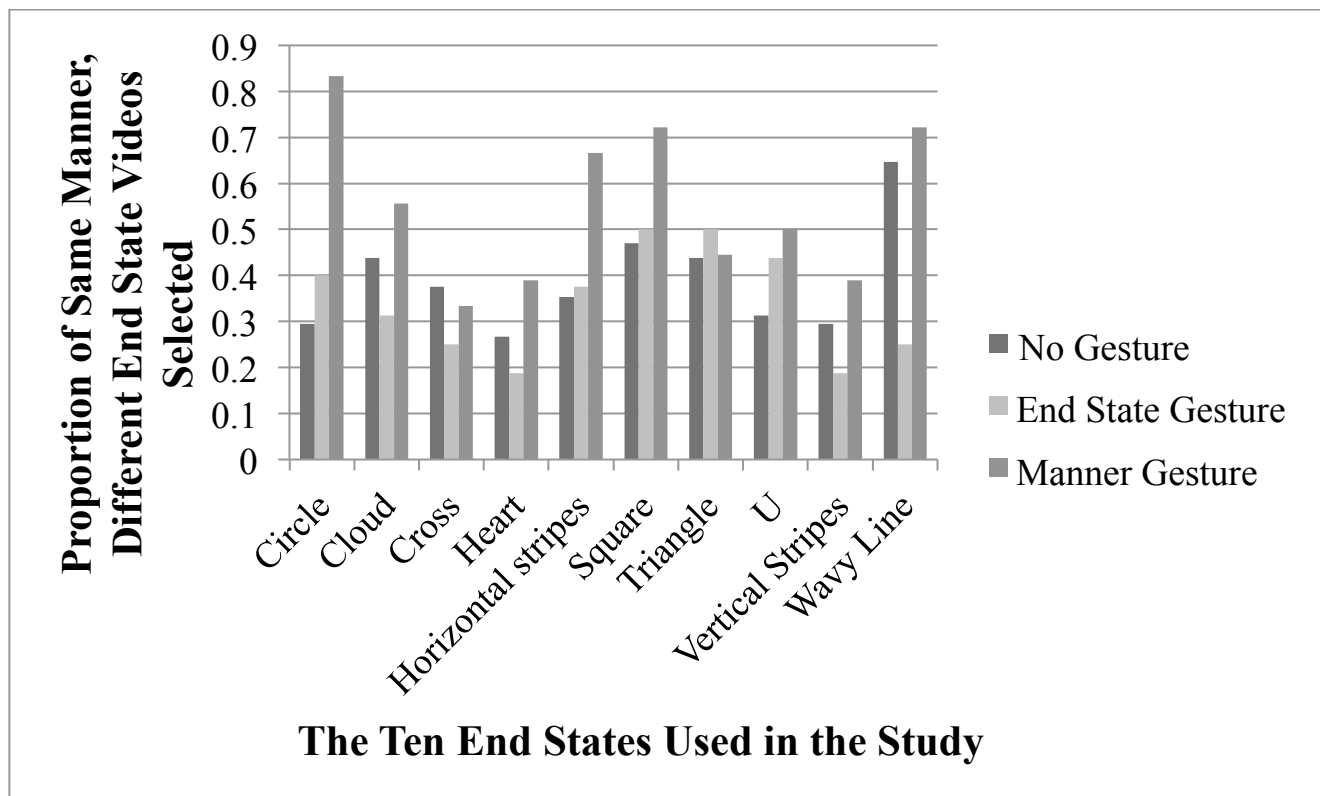

Figure 2. The proportion of trials in which the same manner, different end state video (i.e., generalisation based on manner) was selected for the three gesture groups across the ten end states.

This pattern shows again mimics the main results, such that for 90% of the end states, children who saw end state gestures were less likely than children who saw manner gestures to select the same manner video at test.

The results also reflect the lack of difference in performance between the end state gesture and the no gesture groups, as they each selected a smaller number of same manner videos at test than each other for half of the end states.

Conclusion

Overall, the descriptive patterns obtained are similar to the ones found in the main analysis, such that there is little difference between the end state gesture and no gesture group (particularly when looking across end states), but there is a consistent difference between the manner gesture group and the other two groups, such that children who saw manner gestures selected the same manner video more frequently. This pattern is fairly consistent across stimuli, and it was not the case that a small number of items drove the effect.

Supplementary Materials 3- Additional analysis with age as a covariate

As there was a weak but significant negative correlation between age (in months) and manner bias, additional analysis was conducted to address this. The correlation was such that older children showed a stronger change-of-state bias (Pearson's $R = -.242$, $N = 101$, $p = .015$). Therefore, age was entered into the analysis as a covariate. This should further reduce the error variance (note that the three gesture groups did not significantly differ in age, as reported in the Method section). The proportion of trials with the same-manner video choice were entered into a one-way ANCOVA with gesture type as the between subjects variable (3 levels: no gesture, manner gesture and end state gesture) (see Figure 2) and age as the covariate.

The results revealed again that there is a significant main effect of gesture type (F (2, 97) = 7.453, p=.001).

Supplementary Materials 4- Additional language analysis

If the six children who were acquiring additional languages were excluded, the results did not change. Children who saw manner gestures did not differ from chance (0.5), whereas the other two groups did (end state gesture group: t (29) = -3.581, p=.001; no gesture group: t (29) = -2.223, p=.034). When the data was entered into an ANOVA, there was a main effect of gesture group (F (2, 92) = 7.004, p=.001). LSD post hoc analysis showed that there was a significant difference between the manner gesture group and the no gesture group (p=.009) and the end state gesture group (p=.001).

Supplementary Materials 5- Video Index (see a separate folder for the video files)

For each video group there are four video which were used as the actual stimuli in the experiment. For groups 1 and 3 there are also four videos (per group) demonstrating the associated gestures (two manners and two end states, these were performed live during the training state of the experiment).

Group 1

- Stimuli- Slow Brushing-Square
- Stimuli- Slow Brushing-U

- Stimuli- Digging-Square

- Stimuli- Digging-U

- Manner Gesture- Slow Brushing

- Manner Gesture- Digging

- End State Gesture- Square

- End State Gesture- U

Group 2

- Stimuli- Pinching- Heart

- Stimuli- Pinching- Vertical Stripes

- Stimuli- Pushing- Heart

- Stimuli- Pushing- Vertical Stripes

Group 3

- Stimuli- Pouring- Circle

- Stimuli- Pouring- Cross

- Stimuli- Sprinkling- Circle

- Stimuli- Sprinkling- Cross

- Manner Gesture- Pouring

- Manner Gesture- Sprinkling

- End State Gesture- Circle

- End State Gesture- Cross

Group 4

- Stimuli- Pulling-Triangle

- Stimuli- Pulling- Wavy Line

- Stimuli- Flicking- Triangle

- Stimuli- Flicking- Wavy Line

Group 5

- Stimuli- Placing- Cloud

- Stimuli- Placing- Horizontal Stripes

- Stimuli- ZigZagging- Cloud

- Stimuli- ZigZagging- Horizontal Stripes