adult on video held up the other two novel objects. When the on-screen adult had placed all three objects into the boxes out of view of the camera, the co-viewing researcher paused the video and placed the three real objects into the blue storage bin and closed the lid.

In the unsupportive scaffold condition, the same procedure was followed, except the scaffolds were designed to be unsupportive for drawing a symbolic connection between the video and real objects. The initial scaffold was replaced with, "Look, these are some of the toys you played with!" The request for each object was replaced with "Find the one that [description of object that appeared on screen to the *right* of the one labeled]," where the description involved the shape and color of the object (e.g., is long and skinny with black and white stripes; is round and green and has spikes with a handle). When the researcher held up the described object, she said "Look, this is the one that is [description]!" This procedure was designed to match the researcher's actions and the number and timing of scaffolds across conditions, but to avoid highlighting the correspondence between the real and on-screen objects.

In the partial scaffold condition, the researcher followed the supportive scaffold script for the first two object sets and the unsupportive scaffold procedure for the last object set. Condition differences are displayed in Table 1.

2.4.4. Video labeling

Once the researcher had placed the three novel objects into the blue bin, she said "Now let's listen to what she says!" and unpaused the video. All children watched the video labeling phase without any scaffolds or pauses.

2.4.5. Test

The researcher shook the three objects the child had seen in the video inside the blue bin, opened the lid, tilted the bin toward the child, and said, "Now let's find the dax/wug/tebu! Show me the dax/wug/tebu!" Children were encouraged to "show" the researcher by pulling their selection out of the box. If they did not select an object, they were re-prompted. If they did not answer after the re-prompt, the objects were dumped onto the table and the child was given a final prompt to select an object. The researcher thanked the child for their selection.

2.5. Coding

2.5.1. Object selection

Children's object selections during the three tests were recorded by the researcher during the session. Children's object selections were also coded from video by a coder who was blind to the correct answer and the study condition (n = 359 coded; n = 13 were uncodable from video). There were 6 disagreements (kappa = .97), which were resolved by a third coder who reviewed the videos.

2.5.2. Object handling

The amount of time that children spent handling each object during the video familiarization phase was coded from video. With the sound on, a researcher identified the beginning and end timestamps for this phase. A different researcher (master coder) then coded these portions of each video with the sound off, so they could remain blind to condition and to which object was the target. The demonstration video was not visible in the recording of the child, so the coder was also blind to which object was being handled on screen. The master coder marked the timestamps of when the child started and stopped touching each of the objects. The duration of touching was summed for each object.

A reliability coder coded 23 percent of the sample (29 children x 3 videos x 3 objects = 261 total object handling times) using the same procedure. Inter-rater reliability was excellent (Koo & Li, 2016). That is, the intraclass correlation coefficient for a two-way single measures mixed effects model with a consistency definition was ICC = .977, 95 % CI [.971,.982]. Discrepancies were resolved in favor of the master coder.

3. Results

We first present omnibus models including both object selection and object handling as predictors of learning across all trials. We then focus specifically on testing patterns of learning on the third trial, once children in the partial scaffold condition experienced the change in scaffold, to test whether performance on the third trial differed based on prior scaffolding experience.

3.1. Condition differences in learning

To test for differences in learning, we used a binomial GEE with logit link, with test trial as a repeated measure and condition, test trial, condition x test trial, and handling proportion as fixed predictors. Because our research question was based on children's performance on the test of learning after the third video, when the scaffolding in the supportive and unsupportive conditions had not changed, whereas children in the partial scaffold condition had experienced prior support that was replaced with unsupportive statements, we also tested for condition differences on the third trial, regardless of the results of the omnibus model.

There was no significant effect of handling for either age group. For 30-month-olds, there was only a significant condition x test trial interaction, Wald χ^2 (df = 4) = 11.03, p = .026. Bonferroni corrected follow-up tests indicated that there was no significant difference in learning across the three trials for children in the unsupportive or partial scaffold conditions. For the supportive scaffold condition, there was an effect of test trial number, Wald χ^2 (df = 2) = 6.88, p = .032. In this condition only, there was a significant increase in performance from trial 2 to trial 3 (p = .013 after Bonferroni correction). On the third trial, children selected the correct object more often in the supportive (71 %) than the unsupportive (33 %) scaffold condition, p = .009. Performance in the partial scaffold condition

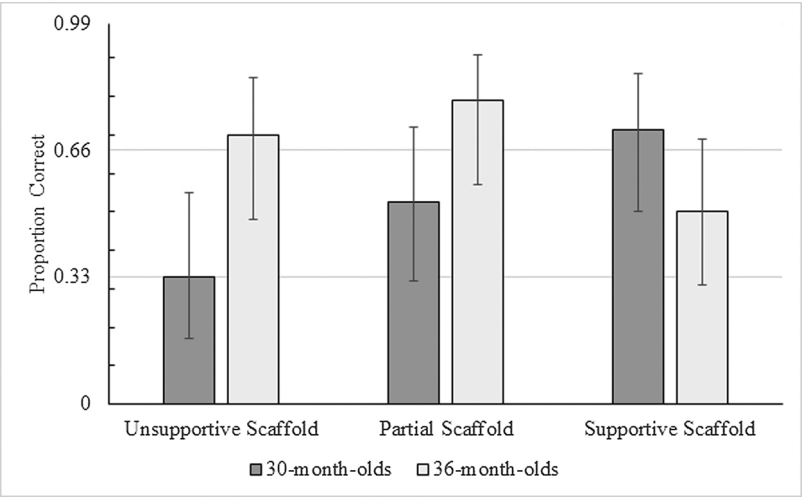

Fig. 2. Proportion of children who chose the correct target on the third learning test with Wilson confidence intervals. Chance performance = 0.33.

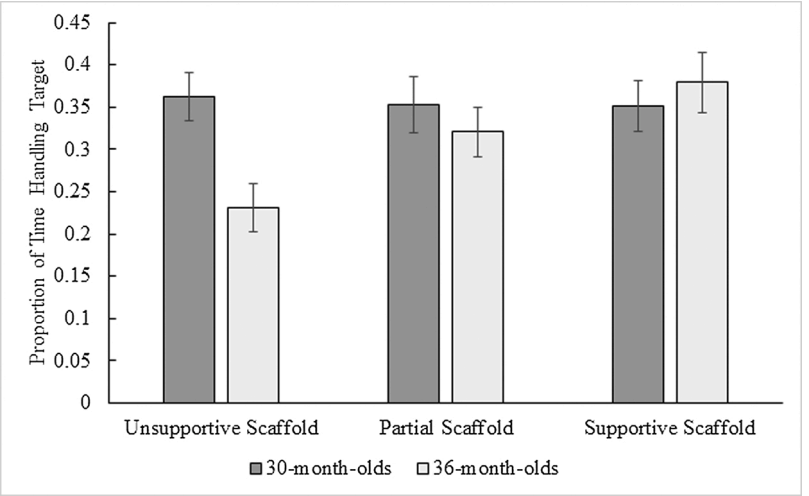

Fig. 3. Proportion of time children spent touching the target object during the video familiarization phase.

(52 %) was not significantly different from the supportive condition, $p = .398$, or the unsupportive condition, $p = .314$.

For 36-month-olds, there were no condition differences or interactions. There was only a significant main effect of trial number, Wald χ^2 (df = 2) = 9.46, $p = .009$. Follow-up comparisons with Bonferroni correction indicated a significant decline (across all conditions combined) from trial 1 to trial 2, $p = .02$ and no other significant differences. On the third trial, performance across conditions did not differ significantly (supportive 50 %, unsupportive 70 %, partially supportive 79 %).

3.2. Tests against chance

We then conducted tests against chance to determine whether children identified the target word at above chance levels after the third video (Fig. 2). We ran binomial tests against a chance value of 0.33 because three objects were available for children to choose between at test. At 30 months, only children in the full scaffold condition chose correctly more often than chance, $p < .001$. At 36 months, children in the unsupportive ($p = .001$) and partially supportive ($p < .001$) scaffold conditions chose above chance. Contrary to our prediction, children in the supportive scaffold condition did not chose significantly differently than chance ($p = .074$).