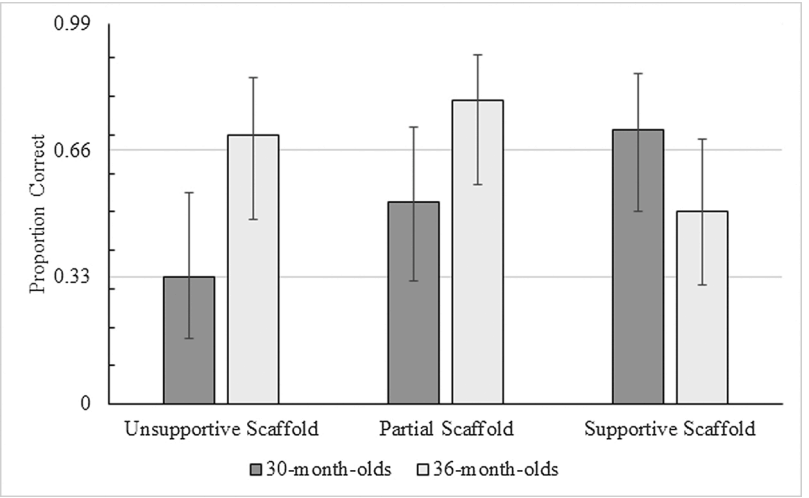

Fig. 2. Proportion of children who chose the correct target on the third learning test with Wilson confidence intervals. Chance performance = 0.33.

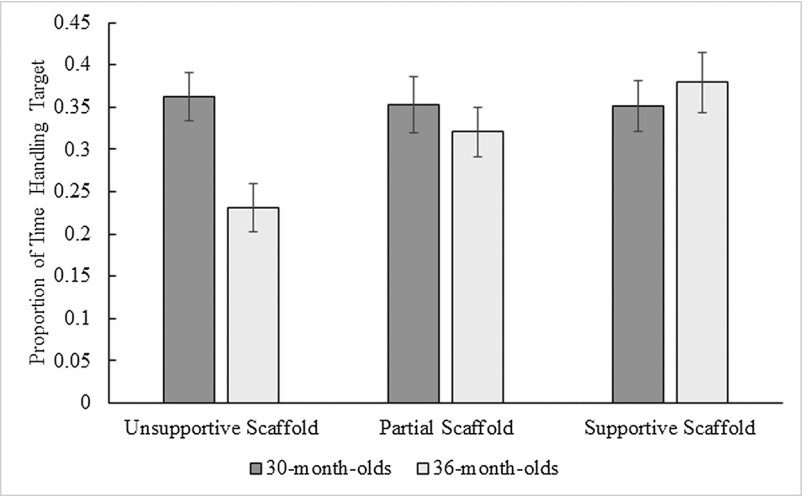

Fig. 3. Proportion of time children spent touching the target object during the video familiarization phase.

(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 ($p = .074$).

3.3. Object handling

To explore whether the supportive scaffolds we used may have simply drawn children's attention toward the target objects, rather than supporting a richer conceptual understanding, we also tested for condition differences in handling of the target objects.[1] We first ran a linear mixed model with test as a repeated measure; age group, condition, and test trial and their interactions as fixed predictors, and the proportion of the familiarization phase spent handling the target object as the dependent variable. We found only a significant age group by condition interaction, $F(2, 345.06) = 3.27$, $p = .039$. We followed-up by conducting separate mixed models for the two age groups. There were no condition differences in handling of the target objects for 30-month-olds, $F(2,165.28) = .03$, $p = .969$ (unsupportive: $M = .36$, $SD = .23$, partial: $M = .35$, $SD = .25$, supportive: $M = .35$, $SD = .24$). However, there was a significant condition difference for 36-month-olds, $F(2,165.69) = 5.53$, $p = .005$. Thirty-six-month-olds in the supportive scaffold condition handled the target object more ($M = .38$, $SD = .28$) than children in the unsupportive scaffold condition ($M = .23$, $SD = .22$), $p = .003$ after Bonferroni correction. Thirty-six-month-olds in the partial scaffold condition fell in between and did not significantly differ from the other conditions (Fig. 3).

4. Discussion

In this study, we compared 30- and 36-month-old children's performance on a label learning task when supportive scaffolds were provided across multiple trials, when supportive scaffolds had previously been present but were replaced by unsupportive scaffolds, and when unsupportive scaffolds were provided across trials. Thirty-month-old children performed above chance on the third trial only when supportive scaffolds were consistently offered across all trials. These children benefited from hearing a scaffold that drew their attention to the correspondence between the video objects and their real-world referents, like 24-month-olds in Strouse and Troseth's (2014) study. When children heard unsupportive scaffolds across all three trials, they did not choose the labeled real object at above chance rates, in line with the performance of 30-month-olds on a similar task in O'Doherty et al.'s (2011) study. Children in the partial scaffold condition, who heard the supportive scaffold for the first two trials and then the unsupportive scaffold on the third trial, also did not identify the labeled object at above chance rates on the third trial. At 30 months, two trials of scaffolding were apparently not enough to support children in learning to think symbolically about video such that they would later apply a new video label on their own. Rather, it is possible that 30-month-olds continued to depend on the presence of the supportive scaffold because it only supported them in associating the specific objects referenced in the scaffold. In fact, it appears that it took many of the 30-month-olds until the third trial to begin making the association to transfer the specific label, as the omnibus models showed no condition or trial differences except for an increase from the second to third trial for the scaffold group. In addition, 30-month-olds handled the target object for about the same amount of time in all three conditions, and their handling was not associated with learning. This suggests that at 30 months, children in the supportive scaffold condition were not simply drawn by the scaffold to pay more attention to the labeled object but were likely supported in conceptually drawing the association between the labeled video object and its specific real-world referent.

At 36 months, there were no condition differences in children's label learning on the third trial, nor across the set of three trials. At this age, many, but far from all, of the children were able to apply the video label to the real-world objects regardless of whether an adult drew their attention to the specific correspondence between the video images and their real-world referents. Although children at this age have benefited from adult scaffolding to learn language in prior studies (Roseberry et al., 2009; Strouse et al., 2013), the label-learning task we used in this study may have been more straightforward for children to solve on their own because it involved learning nouns (Childers & Tomasello, 2006) and the speaker appeared on screen and provided several cues to reference (gaze, movement; see Strouse, 2019).

Although we did not see condition differences at 36 months, only the performance of the unsupportive and partial scaffold groups was significantly above chance. However, with no condition differences on the third trial nor in the omnibus model across trials, further research is needed to determine if the lack of difference from chance in the supportive scaffold group is reliable. It is possibly a type II error. Alternatively, it may reflect a general trend that many 3-year-olds did not maintain engagement with the task. The only pattern we observed in the omnibus models was an overall drop in performance from the first to the second trial, across all groups.

However, it is also possible that performance in this group represents a type of Goldilocks effect, a pattern in which infants attend most to stimuli that are not too complex nor too simple (Kidd et al., 2012). Nussenbaum and Amso (2016) reported this type of pattern when 3- and 4-year-olds learned fewer words from video in a condition in which an on-screen actress provided the most scaffolding (high interactive condition) than in a condition where a more moderate amount of scaffolding was provided (medium interactive condition). Through eye-tracking, they observed that children in the high interactive condition maintained more attention to the labeler's face, whereas children in the medium interactive condition followed the speaker's social cues toward their referents. They argued that an optimal level of social interactivity is needed to engage children without distracting them from the target information. In our study, the scaffolds were provided by a person in the room prior to the labeling, rather than by the on-screen actress during labeling. Eye tracking may help future researchers determine how in-person scaffolding impacts children's visual attention to the on-screen speaker, and whether a moderate level of co-viewer scaffolding best supports children's attention.

[1] We also planned to code children's visual attention during the video familiarization and labeling phases, but we were unable to do so because children's eyes moved out of the camera frame too often.