They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television

CrossMark

Alecia Moser [a], Laura Zimmermann [c], Kelly Dickerson [b], Amanda Grenell [c], Rachel Barr [c], Peter Gerhardstein [a],*

[a] Department of Psychology, Binghamton University, State University of New York (SUNY), Binghamton, NY 13902, USA
[b] Human Research and Engineering Directorate, United States Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
[c] Department of Psychology, Georgetown University, Washington, DC 20057, USA

ARTICLE INFO

ABSTRACT

Despite the ubiquity of touchscreen applications and television programs for young children, developmental research suggests that learning in this context is degraded relative to face-to-face interactions. Most previous research has been limited to transfer of learning from videos, making it difficult to isolate the relative perceptual and social influences for transfer difficulty, and has not examined whether the transfer deficit persists across early childhood when task complexity increases. The current study examined whether the transfer deficit persists in older children using a complex puzzle imitation task constructed to investigate transfer from video demonstrations. The current test adapted this task to permit bidirectional transfer from touchscreens as well. To test for bidirectional transfer deficits, 2.5- and 3-year-olds were shown how to assemble a three-piece puzzle on either a three-dimensional magnetic board or a two-dimensional touch-screen (Experiment 1). Unidirectional transfer from video was also tested (Experiment 2). Results indicate that a bidirectional transfer deficit persists through 3 years, with younger children showing a greater transfer deficit; despite high perceptual similarities and social engagement, children learned less in transfer tasks, supporting the memory flexibility account of the transfer deficit. Implications of these findings for use of screen media (e.g., video, tablets) in early education are discussed.

© 2015 Elsevier Inc. All rights reserved.

 * Corresponding author.
   E-mail address: gerhard@binghamton.edu (P. Gerhardstein).

Introduction

Media sources are advertised as educational tools for young children (Common Sense Media, 2013; Fenstermacher et al., 2010), a perspective widely adopted by parents (Rideout, 2007; Zimmerman, Christakis, & Meltzoff, 2007), where information conveyed via the two-dimensional (2D) source (e.g., television, video, computer, touchscreens) is expected to transfer to contexts beyond the screen. In general, learned skills and knowledge are rarely tested in the same environment and under the same conditions in which learning takes place. More often, access to stored information must be transferred from one context to a novel context; this is termed *transfer learning* (Barnett & Ceci, 2002; Brown, 1990; DeLoache, Simcock, & Marzolf, 2004; Fisch, 2000). Barnett and Ceci's (2002) seminal review of the literature on transfer learning proposes that transfer distance mediates the cognitive resources required to complete the task. Transfer distance is defined as the degree to which context and content of to-be-transferred material match between learning and testing. A *near transfer* task is defined as small differences in context or content, and a more challenging *far transfer* task is defined as a large amount of change in either of these two aspects between learning and test.

Children are particularly sensitive to transfer tasks that involve 2D media because they appear to require more cognitive flexibility (Barr, 2010, 2013; Hayne, 2004) that children lack, making learning from media a far transfer task for young children. The *transfer deficit*, as it pertains to 2D media, is defined as young children consistently learning less from television, touchscreens, and books relative to face-to-face interactions and has been reported across multiple paradigms (for reviews, see Barr, 2010, 2013; Troseth, 2010). Much of the transfer deficit research to date has focused on describing the effect during early development. The transfer deficit emerges during late infancy at around 1 year, peaks at around 1.5 years, and becomes less pronounced in children between 2 and 2.5 years (for reviews, see Barr, 2010; Troseth, 2010). However, some studies have challenged this timeline, demonstrating that these transfer learning deficits may persist until at least 3 years or later for more complex tasks (see Dickerson, Gerhardstein, Zack, & Barr, 2013; Zelazo, Sommerville, & Nichols, 1999, for examples).

Despite being well documented, current explanations for the transfer deficit remain controversial. One explanation is that of *perceptual impoverishment;* degraded perceptual features (e.g., size of objects on screen compared with real-life counterparts, absent depth cues) characteristic of televisions and touchscreens make them a poor source of information (Anderson & Hanson, 2010; Barr & Hayne, 1999; Barr, Muentener, Garcia, Fujimoto, & Chavez, 2007). The perceptual impoverishment explanation is also a component of the poor *memory flexibility* account (Barr, 2013; Hayne, 2004); it is challenging for children to perceptually match features between encoding and retrieval when the features undergo changes in color, brightness, motion, and depth between the demonstration (e.g., 2D) and the test (e.g., three dimensional [3D]). In addition, with televised models there are also changes in social cues (Nielsen, Simcock, & Jenkins, 2008; Troseth, Saylor, & Archer, 2006). Other researchers suggest that the transfer deficit results from children's lack of, or developing understanding of, the dual nature of symbolic objects; two-dimensional images are physical objects that depict other objects, that is, *dual representation* (DeLoache, 2000; DeLoache et al., 2004). By this account, children must be adept at recognizing the relationship between symbolic objects (e.g., images on a tablet) and their real-world referents even though they differ in many attributes. Interestingly, there is mounting evidence of successful dual representation in 2.5- and 3-year-olds in scale model paradigms when perceptual cues between the model and testing room are similar (DeLoache, 2000; DeLoache et al., 2004). However, there is ongoing debate concerning the explanatory value of perceptual impoverishment versus memory flexibility in transfer tasks involving 2D media.

To date, many studies have been limited to the investigation of video demonstrations, which are confounded as to whether the transfer deficit arises from difficulty in mapping perceptual or social cues. There is a small but growing body of studies of learning from now ubiquitous touchscreen devices in which the image itself is the only thing that is 2D. In addition, the presence of a live model ensures the same high level of social engagement for the touchscreen or a live demonstration. Therefore, touchscreen technology functionally isolates perceptually based explanations of memory flexibility and perceptual impoverishment while maintaining the same level of social engagement

across conditions. In a seminal study, Zack, Barr, Gerhardstein, Dickerson, and Meltzoff (2009) tested 15-month-olds using a one-step imitation task in which children performed significantly worse under conditions of far transfer when they observed a demonstration in one dimension (3D button box or on a touchscreen) and were then called on to imitate in a different dimension (2D–3D or 3D–2D transfer) than under conditions of near transfer, that is, when there was a high degree of perceptual similarity between the learning and testing contexts (3D–3D or 2D–2D) (Barnett & Ceci, 2002). Through the implementation of a fully crossed design, the researchers were the first to document a *bidirectional transfer deficit;* transfer of learning across a dimensional change was equally impaired in both directions, suggesting that it was challenging for children to apply what they learned across dimensions even when perceptual differences between 3D and 2D representations were minimal. Zack, Gerhardstein, Meltzoff, and Barr (2013) replicated the bidirectional transfer deficit despite providing more verbal prompts (labeling object and action) during the demonstration. Importantly, if an asynchronous dimensional transfer deficit had emerged, in which 2D–3D had been degraded relative to 3D–2D, this would have supported the perceptual impoverishment account of the transfer deficit. Instead, children in these conditions performed equally poorly and significantly worse than in near transfer conditions (2D–2D or 3D–3D), supporting the transfer deficit as a manifestation of young children's limited memory flexibility.

The current study

One primary goal was to determine whether the bidirectional transfer deficit originally reported by Zack and colleagues (2009, 2013) would persist in 2.5- and 3-year-olds when tested using a more complex puzzle imitation task. We predicted that bidirectional transfer in the current test would mirror the finding of Zack and colleagues and show the same level of deficit regardless of direction. This prediction was motivated by the finding of Dickerson and colleagues (2013), who reported that 2.0- to 3.5-year-olds imitated target actions and performed significantly above baseline but performed significantly worse following a video demonstration of the target actions than following a live demonstration. The puzzle imitation task was also used in the current study but was adapted to be presented as a touchscreen puzzle game in order to enable bidirectional assessment of the transfer deficit. The puzzle imitation task was chosen because it is nonverbal, which enables testing across a wider age range. It is also an ecologically valid imitation task because it involves an everyday activity (puzzle play) and because multiple early educational touchscreen applications use puzzles (Dickerson et al., 2013; Levine, Ratliff, Cannon, & Huttenlocher, 2012; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2014).

A second goal was to test transfer of learning across dimensions from both touchscreen displays and traditional videos. Although generalizing learning from both touchscreens and videos to a 3D test context is defined as far transfer tasks for young children, there are challenges in the video condition that increase the transfer distance beyond that imposed by touchscreens (Barnett & Ceci, 2002). Thus, we predicted that performance would improve over baseline in the video condition and improve still more in the touchscreen–3D manipulation and that children would show the strongest performance in the near transfer manipulation.

An additional objective was to create a more sensitive measure of transfer when mapping information gathered from a 2D media source (e.g., a touchscreen) onto a real-life object and vice versa (Barnett & Ceci, 2002; Subiaul, Anderson, Brandt, & Elkins, 2012). To do this, we relied on a body of literature that has examined the development of social learning strategies (Flynn & Whiten, 2013; Huang & Charman, 2005; Nielsen, 2006; Tennie, Greve, Gretscher, & Call, 2010; Want & Harris, 2002; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009). *Action emulation,* the reproduction of actions in the absence of goal-directed behavior, is characteristic of early development; *goal emulation,* the reproduction of the goals of the task without precise copying of the actions, is more apparent during later development (Tennie, Call, & Tomasello, 2006; Want & Harris, 2002). Goal emulation is thought to be dependent on the ability to form abstract goal representations (Meltzoff, Waismeyer, & Gopnik, 2012). Importantly, social learning theoretically evolved in response to our need to transfer knowledge about tools (Csibra & Gergely, 2006), as a form of gestural language and language itself (Goldstein & Schwade, 2010), which could explain why copying of gestures and copying of goals often occur in parallel (for reviews, see Hopper, 2010; Nielsen et al., 2008; Subiaul et al., 2012).

Definitions of social learning and testing protocols vary widely, making tests of these theoretical assumptions difficult. The current study took a new approach, quantifying the behavioral repertoire of toddlers during an imitation test. In many imitation studies, the task restricts the range of producible behaviors. Even with more flexible imitation tasks, learning is most often assessed as the presence or absence of target behaviors. Rarely are nontarget behaviors that children produce coded. Consequently, these tasks may exaggerate the performance of imitators who produced the target behavior and may underestimate the performance of those who did not. These nontarget behaviors could provide information about the self-generated learning that is taking place within the testing period; that is, they could reveal whether children engage in problem-solving strategies (e.g., trial and error; Sommerville, Hildebrand, & Crane, 2008).

The puzzle imitation task is less constrained compared with other imitation tasks, allowing us to measure age-related changes in strategy use because the puzzle pieces can be acted on in a variety of ways. The current study reports two new measures that determined the proportion of target behaviors relative to all behaviors produced during the puzzle task to assess how faithful children are to the gestures (action fidelity) and how efficiently they complete the puzzle (goal efficiency). We predicted that the imitation performance/social learning strategy would change as a function of age and transfer condition, that is, that older children in the transfer conditions would show greater action fidelity and goal efficiency than younger children, particularly under conditions of near transfer, when demonstration and test present the same stimuli.

Experiment 1: Bidirectional transfer from touchscreens

Method

Participants

This experiment included 172 typically developing children (87 boys) from two metropolitan areas. Independent groups of children were tested at 2.5 years (M_{age} = 30.5 months, SD = 1.33 days) and 3 years (M_{age} = 36.7 months, SD = 0.4 days). Although most participants (96%) were from college-educated families, both ethnicity (Caucasian, 62%) and socioeconomic status (SES) scores (SES range = 34.20–97.16, M = 75.31, SD = 15.00, with 83% of families reporting) (Nakao & Treas, 1994) were diverse. An additional 36 children were excluded from the analysis due to experimenter error (n = 2), technical errors (n = 9), failure to interact with the experimental stimuli for at least 60 s (n = 9), inattention to the demonstration or sliding the stimuli several times during the demonstration (n = 10), or parental interference (n = 6).

Apparatus and stimuli

The apparatus enclosure was made of black plastic (35.5 cm tall, 42 cm wide, 23.3 cm deep). The front of the enclosure contained a removable metal "magnetic" board oriented vertically and painted school bus yellow (Fig. 1a). When the magnetic board was removed, a 17-inch touchscreen display was revealed (Fig. 1a). The apparatus was placed on a child-sized table (13 inches high × 24 inches wide × 18 inches deep) at eye level to the child when seated on a stool.

3D magnetic stimuli. There were two puzzles: a "boat" and a "fish" composed of three small plastic geometric pieces painted different colors. Fig. 1a depicts the stimuli pre- and post-assembly. Pre-assembly, the 0.5-cm puzzle pieces were located in three of four corners. There were two starting positions. A magnetic backing held the pieces on the board but was sufficiently weak that the pieces could be easily slid around on the board.

2D touchscreen stimuli. Mirroring the 3D stimuli, there were "boat" and "fish" puzzles. The 2D stimuli were generated from high-resolution photographs of the magnetic pieces and rendered to match the 3D puzzle pieces in color and size (Fig. 1a depicts the 2D stimuli pre- and post-assembly). Starting positions were the same as those used in the 3D demonstrations.
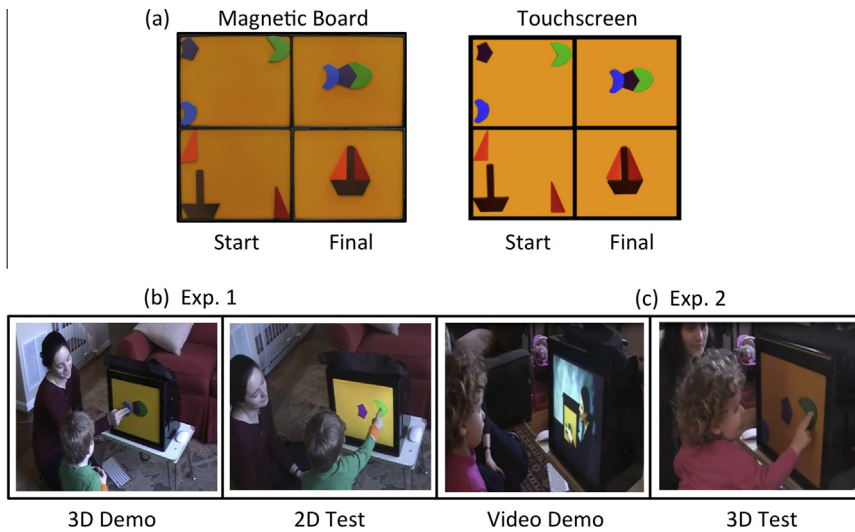
Fig. 1. (a) Stimuli. Left: Magnetic board (3D) puzzle start/end configurations. Right: Touchscreen (2D) start/end configurations. (b) 3D–2D condition. Left: Experimenter demonstrates the target actions on the 3D magnetic board. Right: 36-month-old uses correct gesture on the 2D touchscreen. (c) Video–3D condition. Left: Actions demonstrated on video. Right: 36-month-old uses correct gesture on the 3D magnetic board.

Design and procedure

Children were randomly assigned to one of six conditions (Table 1): two baseline (2D touchscreen or 3D magnetic board), two near transfer (2D–2D or 3D–3D), and two far transfer (2D–3D or 3D–2D). Puzzle (boat or fish) and the starting positions were counterbalanced across all participants. Testing occurred primarily in the home, but all conditions included at least one child tested in the laboratory (n = 37 children were tested in the lab). All conditions, except for baseline conditions, included three consecutive phases: demonstration, test, and manipulation check. Children in the baseline conditions participated only in the test phase, to estimate the rate of spontaneous production of the demonstrated actions during puzzle play, and completed the manipulation check.

Demonstration phase. Participants were seated approximately 50 cm from the apparatus (with the screen covered by a black cloth). The experimenter sat to the side of the apparatus facing the child (Fig. 1b). Once the child was attending to the experimenter and/or apparatus, the experimenter lifted the black curtain to reveal the display (either the magnetic board or the touchscreen). Using the middle and index fingers, the demonstrator slid each puzzle piece into position to make either the fish or the boat in the center of the display, demonstrating a total of three actions (*correct slides*). Each correct slide was accompanied by a nonspecific verbal prompt ("Look at this!," "What was that?," or "Isn't

Table 1
Conditions by phase for Experiments 1 and 2.

	Transfer condition	Condition	Demo	Demonstrator action	Test
Exp. 1	Baseline	3D baseline	None	None	MB
		2D baseline	None	None	TS
	Near transfer	3D–3D	MB	Slide magnet pieces	MB
		2D–2D	TS	Slide virtual pieces	TS
	Far transfer	3D–2D	MB	Slide magnet pieces	TS
		2D–3D	TS	Slide virtual pieces	MB
Exp. 2	Far transfer	Video–3D	V	Slide magnet pieces (on video)	MB

Note. TS, touchscreen; MB, magnetic board; V, video.

that fun?") at the beginning of each gesture. Once the puzzle was complete, the experimenter covered the display with the black curtain to obscure the child's view while the pieces were moved back to their starting positions, either manually on the magnetic board (3D) or by pressing the spacebar for the touchscreen (2D). This sequence was repeated three times (overall demonstration times: touchscreen, $M = 55$ s, $SD = 8$; magnetic board, $M = 52$ s, $SD = 7$).

Test phase.   *Near transfer test.*   A short delay (<30 s) separated the start of the test phase from the end of the demonstration phase. The test phase began when the experimenter lifted the black cloth to reveal either the touchscreen or the magnetic board, with pieces in their starting positions, and said, "Now it's your turn." Each child was given 60 s from his or her first contact with the magnetic board or touchscreen to interact with the apparatus. If the child did not approach the apparatus soon after the test phase began, the experimenter encouraged the child to interact with the puzzle using nonspecific prompts (e.g., "It's your turn to play").

Far transfer test.   A short delay separated the start of the test phase from the end of the demonstration phase, during which time the experimenter covered the display with the black cloth and changed the dimension of the display by either removing the magnetic board from the apparatus to reveal the touchscreen display or vice versa. The experimenter then lifted the cloth to reveal the magnetic board or touchscreen display with the pieces in their starting positions. The test phase was identical to the near transfer test phase.

Baseline (test only).   The test phase began when the experimenter lifted the black cloth and was identical to the near and far transfer test phases.

Manipulation check.   Following the test phase, the experimenter demonstrated the target actions once and gave the child the opportunity to interact with the apparatus. The manipulation check was performed to ensure that the child was physically capable of sliding and connecting the puzzle pieces on the magnetic board and touchscreen. All children passed the manipulation check.

Coding

Each session was video-recorded for later coding in Datavyu (Datavyu Team., 2014), a software program designed for flexible behavioral coding. All behaviors were time-stamped and coded for gestures and goals.

On-task behaviors. Each contact with a puzzle piece (beginning when a piece was touched and ending when the touch ended) was coded along two dimensions: gesture and goal. The combined behavior was considered an on-task behavior. On-task behaviors excluded exploratory play (interactions where the piece was removed from the board for more than 3 s) and micro-gestures (interactions where a piece was "nudged," meaning that it was moved less than one sixth of the board) not resulting in any connection.

Gesture coding. Coded actions included *correct slide, incorrect slide, strategy switch, pick up and move* (3D only), and *swiping* (2D only) (Table 2). All gestures were scored based on the movement of the puzzle pieces and the relation to the point of finger/hand contact. Importantly, children could use any variation of hand shapes when interacting. Despite the differences in the physical properties between 3D and 2D pieces, definitions of nondemonstrated actions were equated across dimension types.

Goal coding. Coded actions that connected puzzle pieces included the following categories of goals: *correct connection, target error connection,* and *connect other* (Table 3).

Reliability. Based on 28% of all test sessions rescored by a second coder, interrater reliability was above the acceptable level of .70 (Landis & Koch, 1977; kappas on the subscales: $\kappa_{gesture} = .741$, $\kappa_{goal} = .847$).

Table 2
Definitions for gesture coding during test on the touchscreen and magnet board.

Gesture code	Definition
Correct slide	Child maintained contact with the face of the puzzle piece using at least one finger for the duration of the slide and an orientation similar to that at the start of the action (a maximum rotation of 45° was allowed)
Incorrect slide	Child made contact with the sides of the piece by any of the fingers or rotated the piece more than 45° (3D pieces were often grasped and slid). The touchscreen equivalent would be dragging or pushing the piece
Pick up and move (3D only)	Child lifted piece off the board for 3 s or less and moved it to a new location more than one sixth of the board away from the onset location
Swipe (2D only)	Child began gesture either before the onset of the piece and swiped across the face of the piece and/or action was evident in the offset by the finger rapidly swinging up and beyond the piece
Strategy switch	Child began correctly sliding the piece and then switched to an incorrect slide or the reverse.

Measure	Calculation
Gesture imitation	Number of correct slides/3 (maximum possible correct slides)
Action fidelity	Number of correct slides/number of on-task behaviors

Table 3
Definitions for coding goals during test session on the touchscreen and magnet board.

Goal code	Definition
Correct connection	Puzzle piece is connected to the central piece in the correct location and orientation within approximately 2 mm of touching
Target error connection	Pieces (a) connected in the correct location but incorrect orientation (e.g., one piece turned sideways), (b) connected not within the 2-mm threshold, or (c) disconnected before the end of the behavior
Connect other	Connections made to nontarget pieces (e.g., connecting the two sails together in the boat puzzle)
No connection	An on-task behavior that did not result in any type of connection

Measure	Calculation
Goal imitation	Number of correct connections/2 (maximum possible correct connections)
Goal efficiency	Number of correct connections/number of on-task behaviors

Results

Coded goals and gestures were used to compute four dependent measures: gesture imitation, action fidelity, goal imitation, and goal efficiency. Analyses of gesture imitation and action fidelity (derived from gesture-coded actions) are presented first, followed by analyses of goal imitation and goal efficiency (derived from goal-directed actions). Action fidelity and goal efficiency are included to more precisely characterize participants' overall behavior during the test phase. See Table 4 for mean proportion scores for each dependent measure for each condition and age group.

Gesture imitation score

Following Dickerson and colleagues (2013), children received credit for each target puzzle piece that they correctly slid, up to a maximum of three, during the 60-s test period. The resulting gesture imitation score was then converted to a proportion to allow for cross-measure comparison. No additional points were given for multiple correct slides with the same puzzle piece.

Action fidelity score

To assess the rate at which correct slides were reproduced relative to other less faithful actions, an action fidelity measure was calculated by taking the sum of all correct slides produced during the testing period (prior to reset following first puzzle completion) and dividing by all on-task behaviors produced (prior to reset following first puzzle completion). Higher proportions indicate more faithful

reproduction of demonstrated actions; lower proportions indicate increasing numbers of nondemonstrated actions produced during the test.

Goal imitation score

Following Dickerson and colleagues (2013), children received one point for each correct connection (max = two correct connections). As with gesture imitation, the goal imitation score was then converted to a proportion (out of two). The goal imitation score is distinct from the gesture imitation and action fidelity scores in that if a child used an incorrect gesture to correctly connect two puzzle pieces, the child still received one point for the goal.

Goal efficiency score

This measure expresses correct connections as a proportion of all on-task behaviors prior to first puzzle completion. This classifies participants on a continuum, with higher proportions indicating highly efficient puzzle reproduction and lower proportions indicating failure to reproduce the puzzle at all. For example, for the boat puzzle, a child might simply move the two sails to most efficiently complete the puzzle, another child might imitate by first moving the mast and then the sails, and still another child might produce 20 on-task behaviors in the course of making the puzzle, a highly inefficient approach.

Preliminary analyses

Preliminary analyses were conducted for each of the four dependent measures (gesture imitation, action fidelity, goal imitation, and goal efficiency) on factors of age, condition, gender, stimulus type (boat or fish), and location (home or lab) and covariates: amount of puzzle play ($M = 27$ min, $SD = 43$), frequency of touchscreen use ($M = 23$ min, $SD = 26$), and delay between end of demonstration and beginning of test (far: $M = 24$ s, $SD = 8$; near: $M = 5$ s, $SD = 2$). No significant ($p < .05$) effects or interactions involving gender, stimulus type, puzzle play, touchscreen use, or delay between demonstration and test emerged in the preliminary analyses; these variables, therefore, were not considered further.

Data analysis plan

For each of the four dependent measures, three separate analyses were conducted according to our objectives. The first analysis determined whether the demonstration or test dimension contributed to performance differences and whether this interacted with age. Importantly, if dimension-specific effects emerged, this would suggest that impoverished perceptual features contributed to the deficit. If no differences emerged, this analysis was omitted for that measure. A second analysis determined whether transfer distance or age significantly affected performance independent of the test or demonstration dimension, which would theoretically represent whether children experienced difficulty in matching demonstration to learning contexts (poor memory flexibility). A third analysis determined whether experimental group performance exceeded baseline because imitation is operationally defined as group performance exceeding baseline. This test reveals whether the manipulation reflected transfer of newly learned information. As expected, baseline scores across all measures were not normally distributed due to children's tendency to perform off-task or incorrect on-task behaviors and did not satisfy the homogeneity of variance assumption of the analysis of variance (ANOVA). Therefore, nonparametric Wilcoxon tests or White's (White, 1980) corrected ANOVAs (for action fidelity and goal efficiency measures in order to use the variance within both measures) were conducted where appropriate. Overall, baseline was stable across age and test dimension; differences in experimental groups cannot be attributed to underlying differences in spontaneous production of the gestures.

Gesture imitation and action fidelity analyses

Gesture imitation. There was no effect of demonstration dimension, so we collapsed across demonstration dimensions and conducted an Age (2.5 or 3 years) × Transfer Condition (near or far) ANOVA on
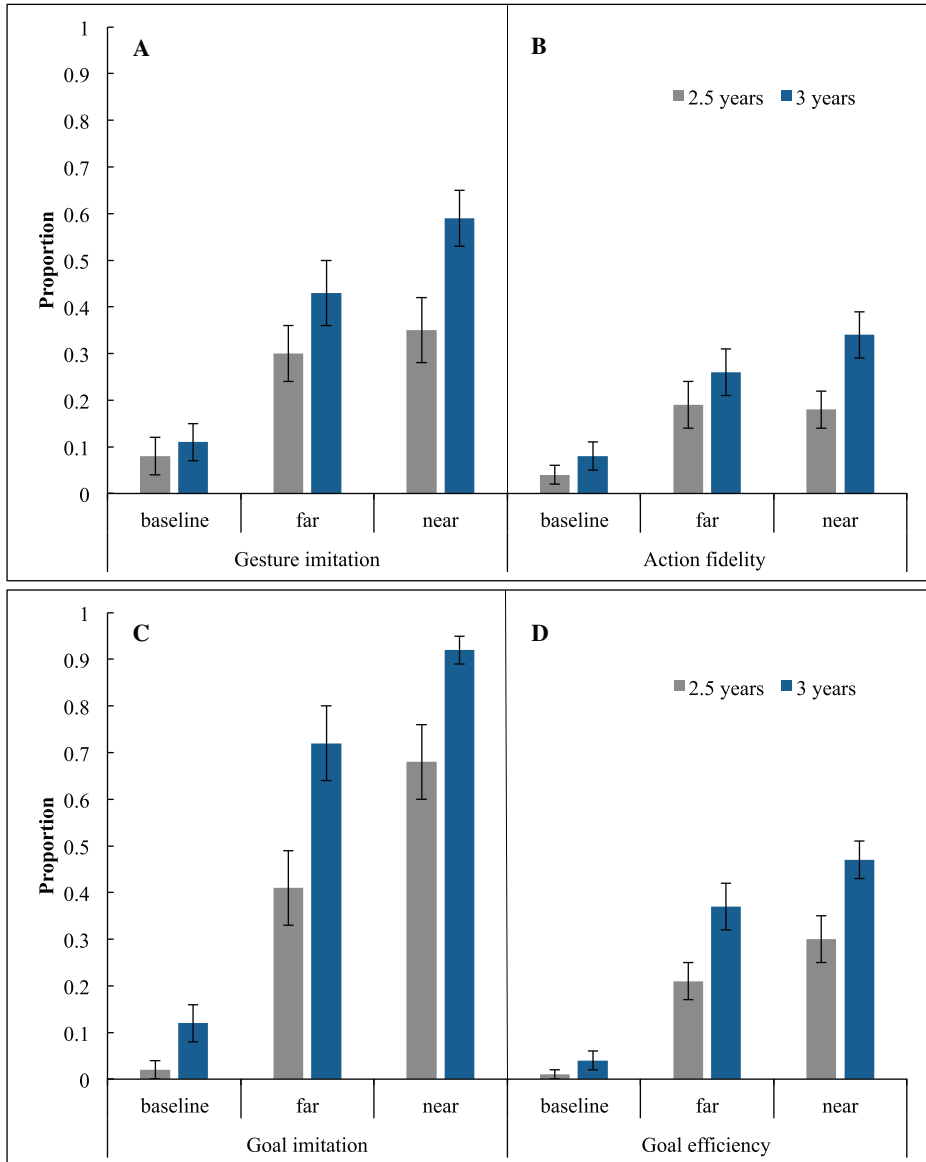
Fig. 2. Top (A,B): Gesture imitation (A) and action fidelity (B) as a function of age and transfer type. Bottom (C,D): Goal imitation (C) and goal efficiency (D) as a function of age and transfer type.

gesture imitation. This analysis revealed a main effect of age, $F(1, 118) = 7.96$, $p < .01$, $\eta_p^2 = .06$ (2.5 years: $M = .32$, $SD = .34$; 3 years: $M = .51$, $SD = .38$), but no effect of transfer condition, $F(1, 118) = 2.62$, $p < .11$, and no interaction ($F < 1$); children's gestures did not differ across transfer conditions (see Fig. 2A and Table 4).

Wilcoxon tests were applied to determine whether experimental groups were significantly above baseline at each age. Children in the near and far transfer conditions performed above baseline (all ps $< .01$) at both ages.

Table 4

Experiments 1 and 2: Mean proportions (and standard deviations) for each dependent measure across age and condition.

Measure									
		Gesture imitation		Action fidelity		Goal imitation		Goal efficiency	
Transfer type	Test	2.5 years	3 years	2.5 years	3 years	2.5 years	3 years	2.5 years	3 years
Baseline	2D	.08 (.06)	.15 (.05)	.04 (.03)	.12 (.04)	.04 (.04)	.15 (.07)	.03 (.03)	.06 (.03)
	3D	.08 (.06)	.06 (.06)	.05 (.03)	.03 (.03)	.00 (.00)	.08 (.06)	.00 (.00)	.03 (.02)
Far transfer	3D (video)	.07 (.05)	.39 (.10)	.02 (.02)	.29 (.10)	.39 (.13)	.67 (.13)	.23 (.09)	.39 (.09)
	2D	.31 (.07)	.44 (.09)	.17 (.06)	.20 (.04)	.46 (.11)	.58 (.14)	.21 (.05)	.20 (.05)
	3D	.29 (.08)	.41 (.11)	.20 (.07)	.29 (.09)	.38 (.10)	.82 (.10)	.22 (.06)	.49 (.07)
Near transfer	2D	.39 (.11)	.67 (.08)	.18 (.07)	.29 (.05)	.54 (.11)	.88 (.06)	.19 (.05)	.36 (.03)
	3D	.31 (.09)	.53 (.09)	.17 (.06)	.37 (.08)	.81 (.09)	.95 (.03)	.40 (.07)	.54 (.06)
Measure totals		.22 (.03)	.39 (.04)	.13 (.02)	.24 (.03)	.38 (.04)	.63 (.04)	.19 (.03)	.33 (.03)

Action fidelity. As above, because there was no effect of demonstration dimension, we collapsed across demonstration dimensions and conducted an analysis of Age (2.5 or 3 years) × Transfer Condition (near or far). This analysis produced a main effect of age, $F(1, 118) = 5.33$, $p < .05$, $\eta_p^2 = .04$, but no other significant effects (all $Fs < 1$). Fidelity increased with age (2.5 years: $M = .19$, $SD = .25$; 3 years: $M = .30$, $SD = .29$) but did not differ as a function of transfer distance (see Fig. 2B and Table 4).

Separate one-way White's corrected ANOVAs across transfer conditions (baseline, near, and far) were conducted at each age. There was a significant main effect of transfer for 2.5-year-olds, $F(2, 82) = 7.14$, $p < .01$, $\eta_p^2 = .09$. Follow-up post hoc Dunnett's t-tests comparing transfer groups with baseline determined that the far transfer group was significantly different from baseline ($p < .05$) but that the near transfer group was only marginally different from baseline ($p = .06$). There was a significant main effect of transfer condition for 3-year-olds, $F(2, 84) = 7.61$, $p < .001$, $\eta_p^2 = .15$; post hoc Dunnett's t-tests determined that both far transfer ($p < .05$) and near transfer ($p < .001$) groups differed from baseline.

Goal imitation score and goal efficiency analyses

Goal imitation. A 2 (Age: 2.5 or 3 years) × 2 (Demonstration Dimension: 2D or 3D) × 2 (Test Dimension: 2D or 3D) ANOVA on goal imitation yielded a main effect of age, $F(1, 114) = 14.61$, $p < .001$, $\eta_p^2 = .11$, no effect of demonstration dimension ($F < 1$), and a nonsignificant effect of test dimension, $F(1, 114) = 3.18$, $p < .10$. The only interaction to reach significance was the Demonstration Dimension × Test Dimension interaction, $F(1, 114) = 11.57$, $p < .001$, $\eta_p^2 = .09$ (see Table 2). Post hoc Tukey HSD tests were conducted to investigate the interaction between demonstration dimension and test dimension. Differences across transfer conditions were evident; 3D–3D performance ($M = .89$, $SD = .24$) was significantly greater ($p < .01$) than the far transfer groups, 2D–3D ($M = .58$, $SD = .49$) and 3D–2D ($M = .52$, $SD = .44$), which did not significantly differ ($p > .05$) (see Fig. 2C). Goal imitation for 2D–2D ($M = 0.72$, $SD = 0.36$) displayed intermediate performance between 3D–3D and the far transfer groups. An Age (2.5 or 3 years) × Transfer Condition (near or far) ANOVA on goal imitation revealed a main effect of age, $F(1, 118) = 16.43$, $p < .001$, $\eta_p^2 = .12$ (2.5 years: $M = .53$, $SD = .44$; 3 years: $M = .83$, $SD = .34$), and confirmed a main effect of transfer, $F(1, 118) = 11.61$, $p < .001$, $\eta_p^2 = .09$ (far: $M = .55$, $SD = .46$; near: $M = .82$, $SD = .31$). There was no Age × Transfer Condition interaction ($F < 1$) (see Fig. 2C and Table 4).

Children of both ages in both near and far transfer conditions produced significantly more correct goal imitation than age- and dimension-matched baseline controls ($ps < .0001$).

Goal efficiency. An Age (2.5 or 3 years) × Demonstration Dimension (2D or 3D) × Test Dimension (2D or 3D) ANOVA yielded significant main effects of age, $F(1, 114) = 10.98$, $p < .01$, $\eta_p^2 = .09$, and test dimension, $F(1, 114) = 15.20$, $p < .001$, $\eta_p^2 = .12$, but no effect of demonstration dimension ($F < 1$). Post hoc Tukey HSD tests showed that goal efficiency proportions were significantly ($p < .001$) higher in the 3D test ($M = .41$, $SD = .30$) than in the 2D test ($M = .24$, $SD = .18$) and that 3-year-olds ($M = .43$, $SD = .27$) were more efficient ($p < .001$) than 2.5-year-olds ($M = .25$, $SD = .25$). A significant interaction

between demonstration dimension and test dimension emerged, $F(1, 114) = 4.29$, $p < .05$, but no other interactions reached significance. A Tukey HSD test ($p < .05$) revealed that children in 3D–3D groups ($M = .49$, $SD = .26$) performed more efficiently than the two conditions with 2D tests, 2D–2D ($M = .28$, $SD = .17$) and 3D–2D ($M = .20$, $SD = .18$), which had equally poor performance. The 2D–3D groups ($M = 0.34$, $SD = 0.33$) showed intermediate performance between 3D–3D groups and the 2D test conditions. The data were then analyzed in an Age (2.5 or 3 years) × Transfer Condition (near or far) ANOVA. There was a main effect of age, $F(1, 118) = 12.80$, $p < .001$, $\eta_p^2 = .10$, and a marginal effect of transfer condition, $F(1, 118) = 3.76$, $p < .06$ (far: $M = .30$, $SD = .30$; near: $M = .40$, $SD = .25$). The interaction was not significant ($F < 1$). These results suggest that both test dimension and transfer distance increase the difficulty of efficiently reaching the goal of the task (see Fig. 2D and Table 4).

A one-way ANOVA on transfer condition (baseline, near, or far) performed at each age (2.5 or 3 years) and test dimension (2D or 3D) using a classical White's correction with post hoc Dunnett's ($p < .05$) *t*-tests revealed that the experimental groups had higher goal efficiency than the baseline group at each age and test dimension.

Discussion

Experiment 1 used a fully crossed design that compared far transfer in both directions (2D–3D and 3D–2D) along with the near transfer conditions (2D–2D and 3D–3D). Analyses across all measures revealed a clear and consistent effect of age, with older children performing better. Given that memory improves with age and experience (Bauer, 2007; Hartshorn et al., 1998; Hayne, 2004; Richmond & Nelson, 2007; Rovee-Collier, 1997), this finding was not surprising. These findings are also consistent with others in the imitation literature (Flynn & Whiten, 2008, 2013; Herbert & Hayne, 2000; Subiaul, Patterson, Schilder, Renner, & Barr, 2015; Subiaul et al., 2012; Williams, Casey, Braadbaart, Culmer, & Mon-Williams, 2014) showing age-related increases in imitation fidelity across the preschool years. Consistent with prior research (Flynn & Whiten, 2008; Nielsen, 2006), children in the current experiment were not faithful to the demonstrated actions but rather produced a wide range of nondemonstrated actions during the test sessions. Transfer and demonstration/test dimension did not, however, alter action fidelity.

Although effects of transfer did not emerge for gesture imitation or action fidelity, the effect of transfer was evident in goal imitation and efficiency. Transfer, however, did not interact with age. Specifically, although 3-year-olds performed better overall than 2.5-year-olds, the transfer deficit was present at both ages tested. This outcome demonstrates that even when the potential disruption to learning imposed via a 2D learning context is reduced by inclusion of a live demonstrator interacting directly with a touchscreen puzzle, the transfer deficit persists. The post hoc tests of goal imitation demonstrated that this effect is bidirectional; although all experimental groups were well above baseline, transfer was adversely affected by a change in dimension both when testing took place on the magnetic board (3D) and when it took place on the touchscreen (2D). Consistent with the findings of Dickerson and colleagues (2013), the transfer deficit is still present in 3-year-olds; furthermore, the transfer deficit is not dependent on the medium.

Goal imitation scores (see Fig. 2C) suggest that 3-year-olds are quite proficient at achieving the goal. The outcome of the goal efficiency analysis, however, demonstrates the utility of the goal efficiency measure; like the 2.5-year-olds, the 3-year-olds are actually relatively inefficient, producing multiple demonstrated and nondemonstrated behaviors before completing the goal (see Fig. 2D). Thus, children at both ages are not "efficient" imitators. This finding is consistent with the idea that young children are at least partially emulating the demonstration. Related to this point, the presence of multiple nondemonstrated gestures suggests that children are learning throughout the 60-s test phase. Some researchers argue that motoric experience with a task actively promotes transfer (Yang, Sidman, & Bushnell, 2010) by providing the means to integrate information from the demonstration and reinforce goal-directed activity (Sommerville et al., 2008). It is likely that this mechanism of integrated learning occurs throughout early childhood (e.g., Barr, Walker, Gross, & Hayne, 2014).

The presence of an interaction between demonstration and test dimension in both goal imitation and goal efficiency shows, however, that the type of test media used (2D touchscreen or 3D magnetic board) affected performance independent of transfer, suggesting that both 2.5- and 3-year-olds are

better imitators with the real puzzle pieces than with the virtual touchscreen puzzle pieces. It is the case, however, that children produced more behaviors on the touchscreen, showing that they were able to use the technology, but they did so less efficiently, producing multiple demonstrated and non-demonstrated behaviors, than when they were using the 3D magnetic pieces. As with the overall age effects, this outcome is not entirely surprising; children in this age range are likely to have less experience with touchscreens than with real objects/puzzles. These difficulties, however, did not interact with the basic task of transferring learning across dimensions. That is, the transfer deficit was not a simple inability to interact with the touchscreen. This issue is considered further in the General Discussion, but first we turn to the question of how increasing transfer distance affects performance by comparing transfer from a video demonstration with transfer from a touchscreen demonstration.

Experiment 2: Comparing video demonstration with touchscreen demonstration

The goal of Experiment 2 was to test a video transfer group (video–3D) and use a cross-experiment analysis to compare performance with the touchscreen (2D–3D) and magnetic board (3D–3D) groups from Experiment 1. Learning from video not only requires the learner to transfer information between dimensions (far transfer: 2D–3D) but also provides diminished perceptual features and lacks socially contingent information known to foster learning (see Flynn & Whiten, 2013; Goldstein, King, & West, 2003; Gros-Louis, West, Goldstein, & King, 2006; Hopper, Flynn, Wood, & Whiten, 2010; Nielsen, 2006; Nielsen et al., 2008; Tennie et al., 2006). Indeed, the demonstrator is 3D and "live" in a touchscreen demonstration but not in a video. Learning from video, therefore, can be characterized as further along the transfer continuum (Barnett & Ceci, 2002) than learning from a touchscreen.

Prior research on age-related changes in transfer of learning indicates that older children have better memory capabilities; they encode more efficiently and are better able to equate and integrate information across different contexts (Hayne, 2004). Therefore, we predicted that older children would be more successful in transfer tasks compared with younger children. Adopting a transfer distance framework, in Experiment 2 we directly tested the effect of transfer distance comparing performance in live, touchscreen, and video demonstration conditions. Due to the greater disparity between the source of information and the testing context, we predicted that children in the video–3D condition would perform worse than those in the touchscreen (2D–3D) condition but that both groups would perform worse than children in the near transfer magnetic board (3D–3D) condition.

Method

Participants

This experiment included 26 typically developing children (16 boys) from two metropolitan areas. Independent groups of children were tested at 2.5 years (M_{age} = 30.6 months, SD = 8 days) and 3 years (M_{age} = 36.7 months, SD = 11). Participants were primarily (92%) Caucasian, all were from college-educated families, and SES ranged from 38.40 to 92.30 (M = 77.90, SD = 14.70, with 69% of families reporting). An additional 3 children were excluded from the analysis for failure to interact with the experimental stimuli for at least 60 s (n = 1), technical error (n = 1), or sibling interference (n = 1). The design for Experiment 2 was a 2 (Age: 2.5 or 3 years) × 4 (Condition: 3D baseline, video–3D, 2D–3D, or 3D–3D) using cross-experimental comparisons (see Table 1).

Apparatus, stimuli, procedure, and coding

The apparatus, test stimuli, procedure, and coding in Experiment 2 were the same as in Experiment 1 except that a video demonstration was played on the touchscreen. In the video demonstration, a female demonstrator (different from the experimenter) sat to the side of the apparatus such that the magnetic board puzzle was centered in the video (see Fig. 1). Using the 3D magnetic stimuli, she demonstrated the sliding actions to make the puzzle three times and used the same verbal cues and actions as in the live and touchscreen demonstration conditions. The video (63 s) began immediately after the cloth was lifted. The experimenter in the room attended to the video along with the child. The child was reoriented to the video with simple verbal cue ("look") and gestural pointing if

attention shifted away during the demonstration. The child was required to look toward the demonstration at least two thirds of the time in order to be included in the final sample. A brief delay separated the start of the test phase from the end of the demonstration phase. The test phase was identical to that of the 3D groups in Experiment 1. As in Experiment 1, children were primarily tested at home ($n = 7$ were tested in the laboratory). When 38% of all test sessions were rescored by a second coder, interrater reliability was very good (kappas on each of the subscales: $\kappa_{gesture} = .759$, $\kappa_{goal} = .827$).

Results

Preliminary analyses

As in Experiment 1, preliminary analyses were carried out on the four dependent measures separately with factors of demonstration condition (video, touchscreen, or magnetic board), age (2.5 or 3 years), gender (male or female), stimulus type (boat or fish), and location (home or lab). Again, daily puzzle play ($M = 32$ min, $SD = 52$), touchscreen frequency ($M = 22$ min, $SD = 25$), and delay between demonstration and test (far: $M = 27$ s, $SD = 6$; near: $M = 6$ s, $SD = 1$) were analyzed as covariates. No main effects or interactions including gender, stimulus type, location, or any of the covariates emerged; thus, these factors were excluded from further analysis.

Gesture imitation and action fidelity analyses

Gesture imitation. A 3 (Demonstration Condition: video, touchscreen, or magnetic board) × 2 (Age: 2.5 or 3 years) ANOVA yielded a main effect of age, $F(1, 91) = 8.96$, $p < .01$, $\eta_p^2 = .09$, with 3-year-olds ($M = .46$, $SD = .39$) performing more gestures than 2.5-year-olds ($M = .23$, $SD = .33$). No other main effects or interactions were observed.

Given the presence of age effects, two age-dependent nonparametric Wilcoxon tests were conducted comparing the video–3D condition with the 3D baseline. The 2.5-year-olds in video–3D did not differ from 3D baseline ($p < .90$), but 3-year-olds in video–3D imitated significantly more gestures than 3D baseline ($p < .01$).

Action fidelity. A 3 (Demonstration Condition: video, touchscreen, or magnetic board) × 2 (Age: 2.5 or 3 years) ANOVA on action fidelity yielded a main effect of age, $F(1, 91) = 8.94$, $p < .01$, $\eta_p^2 = .09$ (2.5 years: $M = .15$, $SD = .24$; 3 years: $M = .34$, $SD = .35$), but there was no significant main effect of demonstration condition, $F(2, 91) = 1.17$, $p > .30$, or interaction ($F < 1$).

To examine performance in experimental conditions against 3D baseline scores, a one-way White's corrected ANOVA with transfer condition (3D baseline, video–3D, 2D–3D, or 3D–3D) was conducted for each age group separately. Results for 2.5-year-olds yielded a significant effect of condition, $F(3, 56) = 4.33$, $p < .01$, but Tukey follow-ups could not be carried out because 2.5-year-olds in the video were at floor. Instead, Welch's two-sample t-tests were conducted and determined that 3D–3D ($M = .18$, $SD = .20$) and 2D–3D ($M = .21$, $SD = .31$) were equivalent and that both were significantly greater than video–3D ($M = .02$, $SD = .06$), which did not differ from baseline ($M = .05$, $SD = .12$). The ANOVA on 3-year-olds' data was significant for condition, $F(3, 57) = 8.52$, $p < .001$. Due to the inclusion of the baseline group, Welch's post-hoc tests were conducted. The results showed that action fidelity in 3D–3D ($M = .37$, $SD = .35$) was equivalent to the two transfer groups, 2D–3D ($M = .29$, $SD = .35$) and video–3D ($M = .29$, $SD = .34$), which did not differ significantly. All experimental groups performed significantly better than 3D baseline ($M = .03$, $SD = .12$) (Table 4 and Fig. 3).

Goal imitation and goal efficiency analyses

Goal imitation. A 3 (Demonstration Condition: video, touchscreen, or magnetic board) × 2 (Age: 2.5 or 3 years) ANOVA on goal imitation yielded a main effect of demonstration condition, $F(2, 91) = 6.72$, $p < .01$, $\eta_p^2 = .13$, and age, $F(1, 91) = 12.44$, $p < .001$, $\eta_p^2 = .12$, but the interaction was not significant, $F(2, 91) = 1.28$, $p = .28$. Follow-up Tukey HSD comparisons ($p < .05$) on condition showed that goal imitation in the near transfer group (3D–3D: $M = .89$, $SD = .24$) was significantly higher than that in the far transfer groups, 2D–3D ($M = .57$, $SD = .49$) and video–3D ($M = .51$, $SD = .48$), which did not significantly differ, demonstrating a transfer deficit.
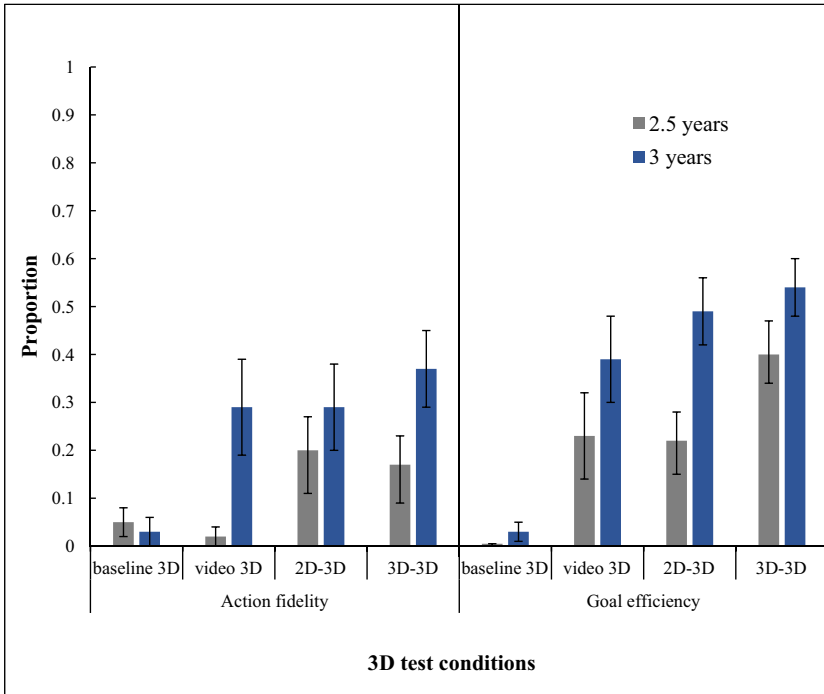
Fig. 3. Left: Action fidelity as a function of age and 3D test condition. Right: Goal efficiency as a function of age and 3D test condition.

Two age-dependent nonparametric Wilcoxon tests were conducted comparing the video–3D condition with the 3D baseline condition. Goal imitation in video–3D was significantly above 3D baseline for both 2.5-year-olds ($p < .05$) and 3-year-olds ($p < .01$).

Goal efficiency. A 3 (Demonstration Condition: video, touchscreen, or magnetic board) × 2 (Age: 2.5 or 3 years) ANOVA on goal efficiency yielded a main effect of age, $F(1, 91) = 11.03$, $p < .01$, $\eta_p^2 = .11$ (2.5 years: $M = .27$, $SD = .30$; 3 years: $M = .50$, $SD = .28$), a very strong trend for a main effect of demonstration condition, $F(2, 91) = 2.20$, $p = .05$, and no significant interaction between age and demonstration condition ($F < 1$).

To further examine the effects of demonstration condition relative to baseline, separate one-way ANOVAs (White's classical correction applied) across transfer conditions (3D baseline, video–3D, 2D–3D, and 3D–3D) were conducted for each age group. Both ANOVAs resulted in significant effects of condition ($p < .001$). Welch's t-tests ($p < .05$) revealed that, although not reaching statistical significance, the 2.5-year-olds in 3D–3D ($M = .40$, $SD = .27$) were higher than those in both transfer conditions, video–3D ($M = .23$, $SD = .33$) and 2D–3D ($M = .22$, $SD = .29$). All experimental groups at this age performed significantly above baseline. A different pattern emerged in 3-year-olds, where all experimental conditions (video–3D, 2D–3D, and 3D–3D) were equally efficient and all were greater than 3D baseline ($p < .01$) (see Table 4 and Fig. 3).

Discussion

This is the first study, to our knowledge, to directly compare learning from touchscreens and video models using the same task parameters. Replicating Dickerson and colleagues (2013), 2.5-year-olds' gesture imitation from video did not exceed baseline, whereas that of 3-year-olds did, suggesting that younger children found it more difficult to learn the gestures from video than from touchscreen. This

age-related change in performance is consistent with Nielsen (2006), who found that younger children were less likely than older children to imitate with high fidelity in a tool-use task in the absence of social cues.

Overall, however, the prediction that transfer distance would result in poorer performance in the video demonstration condition relative to the touchscreen demonstration condition was not supported. Young children imitating from a video model showed similar patterns of action fidelity and goal efficiency when compared with children imitating a live model demonstrating on a touchscreen surface. That is, the transfer and flexible use of perceptual information to real-world situations that have different auditory and haptic cues is challenging, meaning that young children find it surprisingly difficult to transfer learning from a video model and from a touchscreen demonstration (see General Discussion). These findings have important implications for our understanding of the transfer deficit.

General discussion

Our primary objective was to determine whether a transfer deficit from touchscreen demonstrations would persist in 2.5- and 3-year-olds. The results reveal deficits in far transfer conditions (transfer from touchscreens and videos) in both age groups, particularly with respect to our goal-directed measures: goal imitation and goal efficiency. Despite 3-year-olds' better overall performance, goal imitation data in Experiment 1 indicated that near transfer groups performed better than far transfer groups. The same pattern was found in Experiment 2; goal reproduction was significantly higher for near transfer than for far transfer. The transfer deficit was evident despite the fact that a general pattern of overall age-related improvements was seen across measures and that children performed more poorly on touchscreen devices than on the magnetic board. These findings are consistent with Zack and colleagues (2009, 2013) and confirm that the bidirectional nature of the transfer deficit persists across early childhood.

When considering children's goal efficiency in completing the puzzle, the pattern of results across Experiments 1 and 2 reveals that although older children are more efficient, the near transfer advantage is not robust; there are only marginal transfer deficits on this measure. Overall, the use of more traditional measures of imitation performance (gesture and goal imitation), in combination with measures that more effectively characterize the wide repertoire of behaviors that children actually produce, provides strong evidence that (a) the transfer deficit using touchscreen media is reduced, but not ameliorated, in 3-year-olds despite high perceptual similarities and consistent social engagement across conditions, (b) the transfer deficit is present with both video and touchscreen demonstrations, and (c) and the pattern of results across all dependent measures suggests that older toddlers are imitating with increasing fidelity and efficiency. The current study using fine-grain assessments of imitative behavior reveals that imitation might not be as efficient and faithful as previously thought (Csibra & Gergely, 2006; Lyons, Young, & Keil, 2007; Nielsen & Tomaselli, 2010; Whiten et al., 2009). Clearly, young children are proficient and prolific imitators, but future research should more carefully elucidate both goal-related and action-related outcomes as outlined here.

What are the theoretical explanations for the transfer deficit?

In the current set of studies, we conclude that toddlers' poor transfer ability reflects poor memory flexibility. In Experiment 1, the bidirectional transfer deficit was independent of both age and test dimension effects, strongly suggesting that differences in perceptual cues may be particularly difficult for children to equate across dimensions due to memory flexibility limitations and not simply the perceptual impoverishment of the 2D display. These findings parallel those found by Zack and colleagues (2009, 2013), providing converging support for the memory flexibility account and little support for the perceptual impoverishment account of the transfer deficit.

What aspects of the 2D environment are particularly challenging across age?

The most parsimonious explanation would be that the children are simply better at manipulating the 3D magnets compared with the 2D virtual pieces. In actuality, when using the touchscreen, young

children were able to easily move the virtual pieces on the touchscreen but did not necessarily copy the demonstrated gestures and goals. Children in the 2D–baseline conditions produced a similar number of overall on-task behaviors on the screen ($M = 4.0$, $SD = 3.7$) as children in the touchscreen test conditions ($M = 6.8$, $SD = 3.7$). Touchscreens flexibly detect multiple points, require less applied pressure to function properly, and are overall more accurate, so that even children with less experience can move the pieces without difficulty. Therefore, an explanation that children were not able to manipulate the 2D virtual pieces is not sufficient to explain differences based on test dimension. Instead, we propose that other factors may contribute to the differential performance such as lack of object-based feedback in 2D. For instance, whereas the 3D context has haptic and auditory feedback to allow for *online* spatial updating, the 2D environment is void of such feedback.

These salient and multimodal cues may aid in orienting and allocating attention to important areas in space such as interactions between objects; that is, they may help to integrate visuospatial and haptic perception for children and provide a more global frame of reference (Bremner, 1978; Kirkorian & Pempek, 2013; Lockman, 2000). In the current study, we strictly controlled access to the 2D and 3D stimuli and apparatus prior to the test in order to obtain a direct measure of observational learning. Follow-up experiments are in progress, however, to assess whether additional exposure to the touchscreen apparatus prior to the demonstration and test will facilitate learning. Further research examining integration of haptic and visual cues is necessary.

Social learning from screens

The pattern of results across gesture and goal imitation scores supports the growing consensus that young children appear to use social learning strategies in parallel, but these children were disproportionally more goal-directed—obtaining the goal, in this case making the fish or boat, through less precise actions than those demonstrated. In Experiment 1, children reproduced both the demonstrated gestures and goals more than baseline following live demonstrations. Conversely, in Experiment 2, gesture imitation from a video model was lower than that from live demonstrations, showing a greater disturbance in performance than in goal imitation, which exceeded baseline in transfer conditions but was significantly poorer than learning from a live demonstration. Younger children imitate fewer of the gestures from video models than touchscreens, suggesting that social learning strategies are shifting based on the source of information, which is most consistent with the research suggesting that, beyond task difficulty, toddlers' and young children's imitation of actions and goals are mediated by the degree of social engagement (Nielsen, 2006; Nielsen et al., 2008). Csibra and Gergely (2009) argued that ostensive cuing (i.e., demonstrator making eye contact with child) followed by referencing important aspects of the learning environment (i.e., demonstrator shifting gaze to the puzzle pieces and verbal cues during object movement) not only directs attention but also signals that this information is transferrable (see also Wu, Tummeltshammer, Gliga, & Kirkham, 2014).

Fidelity and efficiency of learning

Many tasks measure imitation while focusing on reproduction of the target actions or goals as the sole measure. The action fidelity and goal efficiency measures examined the combination of non-demonstrated behaviors relative to the reproduction of the demonstrated gestures and goals. Action fidelity increased significantly across age. Goal efficiency score increased across age, but the bidirectional transfer deficit was seen with this measure, validating this as a sensitive measure of transfer success. Although our imitation measures show that toddlers learned the demonstrated behaviors, it would be premature to conclude that such behaviors were reproduced accurately and efficiently. In fact, although action fidelity and goal efficiency proportions were above baseline, they were low overall, indicating that children were performing far more nondemonstrated behaviors than demonstrated behaviors during the complex puzzle task.

Note that our definition of "highly efficient" includes the assumption that the most efficient construction of the puzzle involves moving only two pieces to assemble the puzzle around the starting location of the third piece. This strategy was not demonstrated. Even if one assumes, however, that the children perceived the goal as "assemble the puzzle in the middle," necessitating moving all three

pieces, the observed goal efficiency scores relative to this goal were extremely low. One explanation for lower fidelity and efficiency is that applying learned behaviors, especially those requiring motor and spatial skills, comes with some trial and error and is likely to require additional cognitive resources. Documenting performance variability promotes a more accurate description of transfer ability and underscores that learning is continuously updated based on environmental feedback (see also Barr et al., 2014; Sommerville et al., 2008).

Practical implications of these findings for educational settings

Touchscreen technology has a growing presence in early education, with the use of tablets in the classroom and a large number of educational applications being developed and distributed to early educators (Kirkorian & Pempek, 2013). Although television remains the primary form of media exposure for young children (Common Sense Media, 2013), use of touchscreen technology is rapidly increasing (Radesky, Schumacher, & Zuckerman, 2015). Smartphones are the most frequently used touchscreen device (51% have used at least once), although tablets are close behind (44%) (Common Sense Media, 2013).

A recent review of literature on the use of picture books to teach math concepts during early childhood education (Flevares & Schiff, 2014) identified a number of pitfalls and opportunities of using 2D materials to teach math concepts, and the authors strongly argued that direct hands-on experience is necessary for learning. The current findings contribute to this arena by showing a bidirectional deficit in transfer; information acquired via a robust 3D demonstration might not necessarily transfer to the 2D setting as assumed by educators and parents (or the reverse). Learning via an app is presumably supported by children's existing knowledge and representations gained from interacting with the 3D world, but the current study suggests that these representations might not easily transfer from 3D objects to their 2D referents. To provide a concrete example, 3D blocks used to teach math might not be easily translated to 2D block depictions on an app. Together, these findings suggest that to effectively teach content from 2D materials and link it to information that children have already acquired from real-world experiences, teachers and parents should understand that there is a transfer deficit and be made aware of the educational implications. There are potential benefits to including media in the classroom, including enhanced engagement and interactivity, but additional scaffolding (Lauricella, Barr, & Calvert, 2009; Lauricella, Calvert, & Barr, 2014; Levine et al., 2012) by early educators may be necessary to facilitate transfer. Future research is necessary to assess how to facilitate transfer of learning between 2D media and 3D objects in the home and in the classroom.

Acknowledgments

References

Anderson, D. R., & Hanson, K. G. (2010). From blooming, buzzing confusion to media literacy: The early development of television viewing. *Developmental Review, 30*, 239–255.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637.

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*, 128–154.

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives, 7*, 205–210.

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development, 70*, 1067–1081.

Barr, R., Muentener, P., Garcia, A., Fujimoto, M., & Chavez, V. (2007). Age-related changes in deferred imitation from television by 6- to 18-month-olds. *Developmental Science, 10*, 910–921.

Barr, R., Walker, E., Gross, J., & Hayne, H. (2014). Age-related changes in spreading activation during infancy. *Child Development, 85*, 549–563.

Bauer, P. J. (2007). *Remembering the times of our lives: Memory in infancy and beyond*. Mahwah, NJ: Lawrence Erlbaum.

Bremner, J. G. (1978). Egocentric versus allocentric spatial coding in nine-month-old infants: Factors influencing the choice of code. *Developmental Psychology, 14*, 346–355.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science, 14*, 107–133.

Common Sense Media (2013). *Zero to eight: Children's media use in America 2013*. San Francisco: Author.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 249–274). New York: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Datavyu Team. (2014). *Datavyu: A video coding tool*. Databrary Project, New York University. URL: http://datavyu.org.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development, 71*, 329–338.

DeLoache, J. S., Simcock, G., & Marzolf, D. P. (2004). Transfer by very young children in the symbolic retrieval task. *Child Development, 75*, 1708–1718.

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology, 55*, 719–732.

Fenstermacher, S. K., Barr, R., Salerno, K., Garcia, A., Shwery, C. E., Calvert, S. L., et al (2010). Infant-directed media: An analysis of product information and claims. *Infant and Child Development, 19*, 557–576.

Fisch, S. M. (2000). A capacity model of children's comprehension of educational content on television. *Media Psychology, 2*(1), 63–91.

Flevares, L. M., & Schiff, J. R. (2014). Learning mathematics in two dimensions: A review and look ahead at teaching and learning early childhood mathematics with children's literature. *Frontiers in Psychology, 5*. http://dx.doi.org/10.3389/fpsyg.2014.00459.

Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3- and 5-year-olds. *Journal of Experimental Child Psychology, 101*, 228–240.

Flynn, E., & Whiten, A. (2013). Dissecting children's observational learning of complex actions through selective video displays. *Journal of Experimental Child Psychology, 116*, 247–263.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 8030–8035.

Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). New York: Oxford University Press.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*, 509–516.

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Wondoloski, T. L., Klein, P. J., et al (1998). The ontogeny of long-term memory over the first year-and-a-half of life. *Developmental Psychobiology, 32*, 69–89.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review, 24*, 33–73.

Herbert, J., & Hayne, H. (2000). The ontogeny of long-term retention during the second year of life. *Developmental Science, 3*, 50–56.

Hopper, L. M. (2010). "Ghost" experiments and the dissection of social learning in humans and animals. *Biological Reviews, 85*, 685–701.

Hopper, L. M., Flynn, E. G., Wood, L. A., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of Experimental Child Psychology, 106*, 82–97.

Huang, C., & Charman, T. (2005). Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology, 92*, 276–302.

Kirkorian, H. L., & Pempek, T. A. (2013). Toddlers and touch screens: Potential for early learning? *Zero to Three, 33*(4), 32–37.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lauricella, A., Barr, R., & Calvert, S. (2009). Computer skills: Influences of young children's executive functioning abilities and parental scaffolding techniques in the U.S. *Journal of Children and Media, 3*, 217–233.

Lauricella, A., Calvert, S., & Barr, R. (2014). Parent–child interactions during traditional and computer book reading for children's story comprehension. *International Journal of Child-Computer Interaction, 2*(1), 17–25. http://dx.doi.org/10.1016/j.ijcci.2014.07.001.

Levine, S. C., Ratliff, K., Cannon, J., & Huttenlocher, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*, 530–542.

Lockman, J. J. (2000). A perception–action perspective on tool use development. *Child Development, 71*, 137–144.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of over-imitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 19751–19756.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology, 48*, 1215–1228.

Nakao, K., & Treas, J. (1994). Updating occupational prestige and socioeconomic scores: How the new measures measure up. *Sociological Methodology, 24*, 1–72.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology, 42*, 555–565.

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*, 722–731.

Nielsen, M., & Tomaselli, K. (2010). Over-imitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science, 21*, 729–736.

Radesky, J. S., Schumacher, J., & Zuckerman, B. (2015). Mobile and interactive media use by young children: The good, the bad, and the unknown. *Pediatrics, 135*. http://dx.doi.org/10.1542/peds.2014-2251.

Richmond, J., & Nelson, C. A. (2007). Accounting for change in declarative memory: A cognitive neuroscience perspective. *Developmental Review, 27*, 349–373.

Rideout, V. (2007). *Parents, Children, & Media: A Kaiser Family Foundation survey*. Menlo Park, CA: Henry J. Kaiser Family Foundation (http://files.eric.ed.gov/fulltext/ED542901.pdf).

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. *Psychological Review, 104*, 467–498.

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology, 44*, 1249–1256.

Subiaul, F., Anderson, S., Brandt, J., & Elkins, J. (2012). Multiple imitation mechanisms in children. *Developmental Psychology, 48*, 1165–1179.

Subiaul, F., Patterson, E. M., Schilder, B., Renner, E., & Barr, R. (2015). Becoming a high-fidelity – super – imitator: what are the contributions of social and individual learning? *Developmental Science*. http://dx.doi.org/10.1111/desc.12276.

Tennie, C., Call, J., & Tomasello, M. (2006). Push or pull: Imitation vs. emulation in great apes and human children. *Ethology, 112*, 1159–1169.

Tennie, C., Greve, K., Gretscher, H., & Call, J. (2010). Two-year-old children copy more reliably and more often than nonhuman great apes in multiple observational learning tasks. *Primates, 51*, 337–351.

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review, 30*, 155–175.

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development, 77*, 786–799.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2014). Finding the missing piece: Blocks, puzzles, and shapes fuel school readiness. *Trends in Neuroscience and Education, 3*(1), 7–13.

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of imitation in children. *Developmental Science, 5*, 1–13.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817–838.

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation, and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 2417–2428.

Williams, J. H. G., Casey, J. M., Braadbaart, L., Culmer, P. R., & Mon-Williams, M. A. (2014). Kinematic measures of imitation fidelity in primary school children. *Journal of Cognition and Development, 15*, 345–362.

Wu, R., Tummeltshammer, K. S., Gliga, T., & Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology, 5*. http://dx.doi.org/10.3389/fpsyg.2014.00251.

Yang, D., Sidman, J., & Bushnell, E. W. (2010). Beyond the information given: Infants' transfer of actions learned through imitation. *Journal of Experimental Child Psychology, 106*, 62–81.

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch-screen technology. *British Journal of Developmental Psychology, 27*, 13–26.

Zack, E., Gerhardstein, P., Meltzoff, A., & Barr, R. (2013). 15-Month-olds' transfer of learning between touch screen and real-world displays: Language cues and cognitive loads. *Scandinavian Journal of Psychology, 54*, 20–25.

Zelazo, P. D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representations. *Developmental Psychology, 35*, 1059–1071.

Zimmerman, F. J., Christakis, D. A., & Meltzoff, A. N. (2007). Television and DVD/video viewing in children younger than 2 years. *Archives of Pediatrics & Adolescent Medicine, 161*, 473–479.