TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 00 (2025) 1–27

ISSN: 1756-8765 online
DOI: 10.1111/tops.70034

This article is part of the topic "Marks and Meanings: New Perspectives on the Evolution of Human Visual Culture," Izzy Wisher and Michelle C. Langley (Topic Editors).

Cross-Contextual Variability in Children's Early Understanding of Visual Media

Rebecca Zhu,[a] Tabitha Nduku Kilonzo,[b] Lily Zihui Zhu,[c] Judith E. Fan,[a] Michael C. Frank[a]

[a]*Department of Psychology, Stanford University*
[b]*International Education, RTI International Nairobi*
[c]*Department of Psychology, Harvard University*

Received 13 February 2025; received in revised form 4 November 2025; accepted 5 November 2025

Abstract

When and how do children come to understand various kinds of visual media (e.g., pictures, videos, scale models), and how does early experience contribute to variation in the development of visual media comprehension across global contexts? In this selective review, we show that while researchers have investigated how children from Western convenience samples understand visual media, less is known about how this comprehension varies across children in global contexts. Indeed, prior work investigating picture comprehension suggests that children in different contexts may understand pictures at different developmental time points, potentially due to variation in their early picture experiences. These findings demonstrate the need for more research investigating children's comprehension of additional kinds of visual media across contexts. The experience-dependence of visual media comprehension could provide important insight into these abilities' origins, as well as the appropriateness of cross-cultural use of visual media in early childhood measurement.

Keywords: Cognitive development; Picture comprehension; Visual media; Cross-cultural research; Measurement; Education

Correspondence should be sent to Rebecca Zhu, Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, USA. E-mail: ryz@stanford.edu

> "Never before has there been an age like ours when the visual image was so cheap in every sense of the word. We are surrounded and assailed by posters and advertisements, by comics and magazine illustrations. We see aspects of reality represented on the television screen and in the cinema, on postage stamps and on food packages. Painting is taught at school and practised at home as therapy and as a pastime, and many a modest amateur has mastered tricks that would have looked like sheer magic to Giotto."—Ernst Gombrich (1969)

1. Introduction

Visual media—such as pictures, videos, and scale models—are ubiquitous in everyday life, at least in high-income, industrialized societies. In these contexts, humans regularly take and view photographs and videos on their smartphones (Johnson, Flores, & Troseth, 2021), watch television shows and movies, and navigate through their surroundings using maps. Moreover, in these societies, visual media is prevalent not only in everyday adult life, but also in childhood experiences. From a developmental perspective, the most widely used learning materials (e.g., books, posters) and assessments of early cognitive and linguistic abilities (e.g., Bayley Scales of Infant and Toddler Development, Peabody Picture Vocabulary Test, Wechsler Preschool and Primary Scale of Intelligence) rely on visual media, particularly picture stimuli (Fernald, Prado, Kariger, & Raikes, 2017). But how and when do children come to understand visual media?

Moreover, as the Gombrich (1969) quotation above notes, the fact that humans living in contemporary high-income Western contexts are inundated with visual media may be a historical and evolutionary anomaly. In many historical societies, as well as many contemporary global societies, children do not have access to visual media through picture books, televisions, and smartphones. What kinds of experiences do children need to learn about visual media, and how might these experiences, and consequently children's comprehension of visual media, vary across contexts?

Many initial discussions in the cognitive science literature focused purely on the question of whether various kinds of visual media, such as black-and-white line drawings or color photographs, were universally understood early in development (Barrera & Maurer, 1981; DeLoache, Pierroutsakos, & Uttal, 2003). These discussions have given way to a more nuanced understanding of children's developing expertise in visual media comprehension, for example, by asking what kind of experience, if any, facilitates this development (Callaghan, 2020). Here, we review the state of the field, specifically focusing on previous research investigating infants', toddlers', and preschoolers' comprehension of visual media across different kinds of media types and different learning environments (see Table 1).

We first review the existing literature—which is primarily based on empirical research with Western convenience samples—on variation in the development of visual media comprehension across different kinds of visual media. Our goal was to assemble a representative, but not exhaustive, set of empirical studies on the development of visual media comprehension. We began our literature search with a seed set of notable studies on infants' and toddlers'

Table 1
Summary of evidence, primarily drawn from Western convenience samples

	Infants (0-18 Months)	Toddlers (18-36 Months)	Preschoolers (36-60 Months)
RECOGNITION			
Photographs	**Strong evidence of success from early infancy** (Barrera & Maurer, 1981; Dirks & Gibson, 1977; DeLoache et al., 1979; DeLoache et al., 1998; Ganea et al., 2008; 2009; Hochberg & Brooks, 1962; Pierroutsakos & DeLoache, 2003; Simcock & DeLoache, 2006; 2008; Ziemer et al., 2012)	**Evidence of success** (Ganea et al., 2008; 2009; Simcock & DeLoache, 2006; 2008; Tare et al., 2010)	**Evidence of success** (Wimmer et al., 2014; Tare et al., 2010)
Cartoons	**Evidence of success** (Frank et al., 2009; Ganea et al., 2008; Pierroutsakos & DeLoache, 2003)	**Evidence of success** (Ganea et al., 2008; Tare et al., 2010)	**Evidence of success** (Ganea et al., 2011; Tare et al., 2010; Wimmer et al., 2014)
Black-and-White Line Drawings	**Evidence of success** (DeLoache et al., 1979)	**Mixed evidence of both success** (Gelman & Ebeling, 1998; Hochberg & Brooks, 1962; Preissler & Bloom, 2007; 2008; Salsa & Vivaldi, 2017; Simcock & DeLoache, 2006) **and failure** (Callaghan, 1999; 2000; Simcock & DeLoache, 2006)	**Strong evidence of success** (Armitage & Allen, 2015; Bloom & Markson, 1998; Callaghan, 1999; 2000; Hartley & Allen, 2014; Long et al., 2024; Salsa & Vivaldi, 2017; Wimmer et al., 2014), **though anecdotal evidence of failure in picture-sparse context** (Saitoti, 1986)
Scale Models	**Evidence of success** (Rosengren et al. 2009)	**Strong evidence of success** (Casler et al., 2011; DeLoache et al., 2004; DeLoache et al., 2013; Grzyb et al., 2017; Hunley & Hahn, 2016; Hagihara et al., 2022; Ishibashi et al., 2021; Ishibashi & Moriguchi, 2017; Rosengren et al., 2009; 2010; Ware et al., 2010)	**Evidence of success** (Casler et al., 2011; Ware et al., 2010)
Videos	**Evidence of success** (Diener et al., 2008; Pierroutsakos & Troseth, 2003; Rosengren et al., 2021)	**Evidence of success** (Revencu & Csibra, 2021;Rosengren et al., 2021)	**Evidence of success** (Flavell et al., 1990)

(*Continued*)

Table 1
(Continued)

	Infants (0-18 Months)	Toddlers (18-36 Months)	Preschoolers (36-60 Months)
REFERENCE			
Photographs	**Evidence of success at 15 months** (Ganea et al., 2008; 2009; Tare et al., 2010) **but not before** (DeLoache et al., 1998; Pierroutsakos & DeLoache, 2003; Ziemer et al., 2012); **No evidence of success in picture-sparse context** (Walker et al., 2013)	**Evidence of success** (DeLoache and Burns, 1994; DeLoache et al., 1998; Ganea et al., 2008; 2009; Tare et al., 2010; Walker et al., 2013)	**Evidence of success** (Ganea et al., 2011; Hartley & Allen, 2015; Uttal et al., 2008; Walker et al., 2013; Wimmer et al. 2014; Zaitchik, 1990)
Cartoons	**No evidence of success** (Ganea et al., 2008; Pierroutsakos & DeLoache, 2003; Tare et al., 2010)	**Evidence of success** (Tare et al., 2010) **though not in picture-sparse context** (Zhu et al., 2025)	**Evidence of success** (Wimmer et al., 2014)
Black-and-White Line Drawings	**No evidence of success** (Pierroutsakos & DeLoache, 2003)	**Evidence of success** (Preissler & Carey, 2004)	**Evidence of success** (Apperly et al., 2004; Hartley & Allen, 2015; Wimmer et al., 2014)
Scale Models	**No evidence of success** (Rosengren et al., 2009; Ware et al., 2010)	**No evidence of success** (Casler et al., 2011; DeLoache et al., 2004; DeLoache et al., 2013; Grzyb et al., 2017; Hunley & Hahn, 2016; Hagihara et al., 2022; Ishibashi et al., 2021; Ishibashi & Moriguchi, 2017; Rosengren et al., 2009; 2010; Ware et al., 2010)	**Success from 4 years on** (Grzyb et al., 2017) **but not earlier** (Casler et al., 2011; Ware et al., 2010)
Videos	**No evidence of success** (Pierroutsakos & Troseth, 2003; Rosengren et al., 2021)	**No evidence of success** (Rosengren et al., 2021; Troseth & DeLoache, 1998) **though some success with training** (Troseth, 2003)	**No evidence of success** (Flavell et al., 1990)

recognition and referential understanding of different visual media (e.g., DeLoache & Burns, 1994; DeLoache, Pierroutsakos, Uttal, Rosengren, & Gottlieb, 1998, DeLoache, Uttal, & Rosengren 2004; Ganea, Pickard, & DeLoache, 2008; Pierroutsakos & DeLoache, 2003; Pierroutsakos & Troseth, 2003; Preissler & Carey, 2004). From this base, we conducted backward searches (i.e., reference lists of the seed papers) and forward searches (i.e., records

citing the seed set) in Google Scholar. We then filtered within these citation lists using search terms that describe visual media type (i.e., photo*, cartoon, line drawing, scale model, video, screen), comprehension (i.e., recogn*; refer*), and age group (i.e., infants, toddlers, preschoolers). As a soft constraint, we sought to include approximately 10−15 of the most cited papers for each combination of visual media type and comprehension, while covering different developmental stages. This goal was not always achievable (e.g., relatively few studies addressed cartoons). Most articles identified through this process were empirical studies with Western convenience samples. To identify additional articles for cross-contextual comparisons, we returned to the seed citation lists and applied additional search terms such as "cross-cultural," "cross-contextual," and "SES." Unless otherwise specified, all children were tested in high-income, convenience sample, Western contexts. To date, much of the published research conducted in these Western convenience sample contexts does not provide, or provides extremely sparse, descriptions of the research contexts (Draper et al., 2023). Then, we ask what particular kinds of experiences may account for variation in the development of visual media comprehension across early environments. Finally, we discuss future directions for empirical research to help better articulate the mechanisms underlying children's emerging understanding of various kinds of visual media, as well as further implications of these findings to inform both basic and applied developmental science research.

2. Distinct types of media comprehension: Recognition versus reference

Drawing inspiration from Deloache and colleagues' (2003) definition of "pictorial competence," we define visual media comprehension as the set of abilities and processes involved in perceiving, interpreting, understanding, and using various kinds of visual media. A full understanding of visual media involves, though is certainly not limited to, at least two cognitive capacities, namely, the capacity to *recognize* the content depicted in visual media, as well as the capacity to understand the *referential* nature of visual media. We focus on two critical aspects of visual media comprehension, recognition and reference, because these two capacities emerge early in development, have been studied to some degree across cultures and contexts, and are relevant for a wide array of early childhood learning materials and assessments.

Recognition is the capacity to recognize the content of a visual representation (e.g., that a photograph depicts a *truck*; or that a Picasso painting depicts a *woman*—albeit a very abstract woman). In infant studies, evidence of recognition capacities come from behavioral experiments in which infants successfully track entities (e.g., cartoon faces; Frank, Vul, & Johnson, 2009), or differentiate between certain kinds of entities (e.g., photographs of smiling and frowning faces; Barrera & Maurer, 1981), presented using visual media. Later in development, children provide further evidence of their recognition capacities by accurately naming depicted objects and matching them to corresponding real-world referents (Wimmer, Robinson, Koenig, & Corder, 2014). Experimental tasks that require children to learn new information from visual media also provide evidence of recognition capacities (Ganea, Pickard, & DeLoache, 2008; 2011; Preissler & Carey, 2004; Simcock & DeLoache, 2006;

2008; Tare, Chiong, Ganea, & DeLoache, 2010), as children must recognize the content depicted in the visual media in order to learn from it.[1]

Reference is the capacity to understand that visual media refers to, rather than instantiates, an actual entity (e.g., a *photograph* of a truck is a depiction of a truck, and not a kind of very flat, small truck; a Picasso *painting* is a depiction of a woman, and there is not an actual woman stuck inside the frame). Often, existing research simultaneously highlights children's capacities to recognize visual media, and their failures to understand the referential nature of the same visual media. For example, a predominantly White, middle-class, midwestern U.S. sample of 9-month-old infants grasp at depictions of toys, reaching out and curling their fingers around the depiction as if trying to pick it up off the page (DeLoache et al., 1998). Similarly, toddlers frequently commit scale errors, interacting with replicas as if they were actual objects, for example, by attempting to sit in a miniature chair (DeLoache et a l., 2004). These behaviors demonstrate that children recognize the content of visual media, but are unable to understand that the visual media refers to, rather than instantiates, actual objects. However, another possibility is that infants and toddlers possess some referential understanding, but have poor executive function (i.e., inhibitory control) capacities (Jiang & Rosengren, 2019). Specifically, infants and toddlers may understand that various kinds of visual media are representations, rather than instantiations of actual objects, but the "pull of the real" may be too strong, and they may attempt to interact with visual media as if with actual objects, despite their referential understanding. Indeed, DeLoache and colleagues (1998) note that infants grasp at visual media less than at actual objects, supporting the view that infants' grasping behaviors are not driven by a strong belief that visual media depictions are actual objects. Overall, these behaviors (i.e., grasping, scale errors) at least indicate an absence of positive evidence for referential understanding.

Additional measures of referential understanding include picture-to-object extension tasks (Ganea, Preissler, Butler, Carey, & DeLoache, 2009; Preissler & Carey, 2004) and retrieval tasks (DeLoache, 1987; DeLoache & Burns, 1994; Troseth, 2003; Troseth & DeLoache, 1998). In picture-to-object extension tasks, infants and toddlers are taught a novel word through a visual media depiction of a novel object. On the test trial, participants are prompted with the novel word and then shown both the actual object and the depiction of the object. Researchers argue that if the participants select the actual object, or both the visual media depiction and the actual object, they demonstrate some understanding that the visual media refers to actual objects (Ganea et al., 2009; Preissler & Carey, 2004; Zhu, Pitchik et al., 2025; though see Sloutsky, 2009 for a learner statistical association account arguing that children select the actual object because they possess stronger word-object than word-picture associations in general). In retrieval tasks, toddlers and preschoolers must use information presented through visual media to search for and retrieve items in the actual world.

Overall, visual media comprehension involves at least two related but distinct capacities, namely, recognition and referential understanding. These two capacities can be measured through various kinds of developmentally appropriate experimental paradigms.

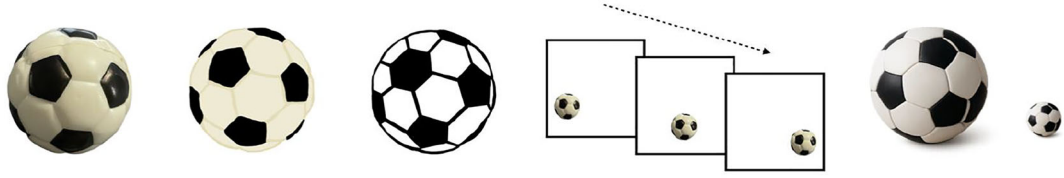

Fig. 1. From left to right, a photograph, cartoon, black-and-white line drawing, video, and scale model of a soccer ball to the right of an actual soccer ball (photograph, cartoon, and black-and-white line drawing are stimuli from Zhu, Kilonzo et al., 2025).

3. Variation in development across kinds of visual media

The capacity to understand different kinds of visual media might have differing developmental trajectories, even when considering only middle-to-high-income Western children (e.g., Pierroutsakos & DeLoache, 2003; Simcock & DeLoache, 2006). Here, we review comprehension of five kinds of visual media—specifically, photographs, cartoons, and black-and-white-line drawings (see Fig. 1), as well as scale models and videos. These visual modalities are important to consider because they have been studied extensively in Western convenience samples and are also frequently used in childhood learning materials and assessments.

Our review of existing evidence suggests that children might readily acquire the ability to *recognize* the contents of highly realistic visual representations, including photographs (Barrera & Maurer, 1981; DeLoache et al., 1998; Dirks & Gibson, 1977; Rose, 1977), scale models (DeLoache et al., 2004), or videos (Pierroutsakos & Troseth, 2003). By contrast, they seem to have more difficulty understanding the *referential* nature of these very same photographs (Pierroutsakos & DeLoache, 2003), scale models (DeLoache, 1987; 1995; 2003), and videos (Pierroutsakos & Troseth, 2003). Meanwhile, prior research suggests that children might have more difficulty recognizing the content of less realistic visual media (e.g., black-and-white line drawings; Callaghan, 2000), but less difficulty understanding these kinds of visual media as referential (Preissler & Carey, 2004).

3.1. Photographs

Photographs are a highly realistic kind of visual media, typically produced using a camera. Infants recognize the content of photographs from very early in life. Three-month-olds recognize and discriminate between photographs of smiling and frowning faces (Barrera & Maurer, 1981). Five-month-olds who were familiarized with an actual doll, and then presented with a photograph of the original doll and a photograph of the new doll, showed a novelty preference for the photograph of the new doll (DeLoache, Strauss, & Maynard, 1979). Two experiments involving White, middle-class, midwestern American samples showed that 9-month-olds grasp at photographs of toys, as if trying to play with the depicted toys (DeLoache et al., 1998; Ziemer, Plumert, & Pick, 2012). Furthermore, 18- and 24-month-olds in Brisbane, Australia, may learn new information more easily from photographs than other, less realistic, kinds of visual media (Simcock & DeLoache, 2006; 2008).

However, while recognizing the content of photographs may be an early emerging capacity, understanding the referential nature of photographs may be a more protracted developmental process. Though infants can differentiate between photographs and physical objects (see DeLoache et al., 1979 for evidence from 5-month-olds; and Slater, Morison, & Rose, 1984 for evidence from Exeter, U.K. newborns), they fail to demonstrate a full-fledged understanding that photographs refer to, rather than instantiate, actual objects. Both middle-class White 9-month-old infants in the United States and low-income 8- to 18-month-old Beng infants in Cote D'Ivoire grasp at color photographs of toys, as if trying to play with the depicted toys (DeLoache et al., 1998). Moreover, 9-month-olds tend to produce grasping behaviors most frequently with color photographs, somewhat less frequently with cartoons, and least frequently with black-and-white line drawings (Pierroutsakos & DeLoache, 2003). At 19 months, infants respond to photographs by pointing rather than grasping, demonstrating a qualitative developmental change in their responses toward photographs (DeLoache et al., 1998).

As discussed above, beyond signaling a failure to understand reference, there are alternative interpretations of infants' grasping behaviors. In addition to the inhibitory control explanation given above, another alternative interpretation is that 9-month-olds do not demonstrate the same kind of pointing behaviors as 19-month-olds, not because of a developmental change in infants' referential understanding, but because of a developmental change in infants' pointing behaviors. Indeed, pointing behaviors typically emerge around 12 months of age (see Liszkowski, Carpenter, & Tomasello, 2007 for evidence from German infants; and Tomasello, Carpenter, & Liszkowski, 2007 for review). Still, 9-month-olds not only fail to engage in an appropriate behavior that demonstrates referential understanding (i.e., pointing), but also engage in an inappropriate behavior that demonstrates a lack of referential understanding (i.e., grasping for the depicted toy). Furthermore, 9-month-olds do not engage in other appropriate responses that may demonstrate referential understanding (e.g., attempting to pick up the entire photograph, rather than the toy depicted in the photograph). At least, there is no positive evidence that 9-month-olds understand the referential nature of photographs.

However, there is some evidence that older infants and toddlers understand the referential nature of color photographs. Specifically, 15-, 18-, and 24-month-olds were taught a new word for a novel object depicted in a color photograph. On the test trial, participants were shown the color photograph of the novel object and the actual object. In all three age groups, approximately 75% of participants selected either the actual object alone, or selected both the actual object and the color photograph. However, 25% of participants selected only the color photograph. This experiment showed some evidence that toddlers, and even older infants, understand that color photographs refer to actual objects (Ganea et al., 2009). Furthermore, children's performance on retrieval tasks lends additional evidence for the developing understanding of the referential nature of photographs. Twenty-seven- and 30-month-olds, but not 24-month-olds, spontaneously succeed at using information from color photographs as retrieval cues in the actual world (DeLoache & Burns, 1994). Overall, it may be relatively easy for children to recognize the content of highly realistic photographs, but more difficult for children to understand the referential nature of highly realistic photographs.

3.2. Cartoons

Cartoons are colorful schematic depictions that typically bear less perceptual resemblance to actual objects than photographs. There has been less work investigating children's understanding of cartoons relative to other kinds of pictures (i.e., photographs, black-and-white line drawings). Children's capacity to understand cartoons may follow a similar developmental trajectory to their capacity to understand photographs (DeLoache, Uttal, & Pierroutsakos, 2000; Ganea et al., 2008; Tare et al., 2010). By at least 9 months, infants look at faces during free viewing of cartoons at levels higher than expected based on salient features, potentially signaling their recognition of these faces as worthy of attention (Frank et al., 2009). Moreover, preschoolers can match cartoons to actual objects (Wimmer et al., 2014). However, while toddlers and preschoolers can learn new information from cartoons, they might learn less effectively from cartoons than from photographs (Ganea et al., 2008; Tare et al., 2010). For example, 18-month-olds more often extend newly learned labels between pictures and objects when the pictures were photographs than cartoons (Ganea et al., 2008). Overall, like the capacity to recognize the content of photographs, the capacity to recognize the content of cartoons might emerge early in development.

Similarly, in terms of referential understanding, infants might struggle with the capacity to understand that both photographs and cartoons refer to, rather than instantiate, actual objects. Nine-month-olds grasp at the depicted objects in cartoons, producing more grasping behavior with cartoons than black-and-white line drawings, but less grasping behavior with cartoons than with color photographs (Pierroutsakos & DeLoache, 2003). Moreover, 2-year-olds living in two communities in urban and rural western Kenya showed no positive evidence of referential understanding of cartoons (Zhu, Pitchik et al., 2025). These participants were taught a new word for a novel object depicted in a cartoon. On the test trial, participants were shown the cartoon of the novel object and the actual object. Thirty percent of toddlers in both the urban and rural contexts selected only the object, or both the object and the cartoon, whereas 70% of toddlers in both contexts selected the cartoon. Thus, the literature provides some evidence that children have difficulty understanding the referential nature of cartoons.

Overall, the development of children's capacities to understand photographs and cartoons may track together closely, though effects may be more pronounced for photographs (Ganea et al., 2008; Pierroutsakos & DeLoache, 2003), which bear more similarity to actual objects than cartoons do. Thus, children may have less difficulty recognizing the content of highly realistic visual media (e.g., photographs, cartoons), but more difficulty understanding these kinds of visual media as referential.

3.3. Black-and-white line drawings

Black-and-white line drawings are schematic depictions that highlight only the most relevant object features, using relatively few strokes or lines. One case study suggests that recognizing the content of black-and-white line drawings is early emerging. In this study, psychologists Hochberg and Brooks (1962) did not expose their own child to visual media until 19 months of age, but their child was still able to recognize and name objects depicted in color

photographs and black-and-white line drawings upon his very first encounters. However, another case study suggests that recognizing the content of black-and-white line drawings may be difficult: in an autobiographical account, a Maasai warrior from rural Tanzania recounts his confusion when first encountering black-and-white drawings as a child attending school, identifying only scratches on a page rather than a drawing of a person (Saitoti, 1986). Thus, anecdotal case studies provide conflicting evidence on children's capacity to recognize the content of black-and-white line drawings and should not be taken as definitive.

Experimental evidence suggests that children may have more difficulty recognizing the content of black-and-white line drawings than recognizing the content of photographs and cartoons. Three-year-olds can match black-and-white line drawings to corresponding real-world objects (Callaghan, 1999; 2000; Wimmer et al., 2014), whereas 2- and 2.5-year-olds cannot (Callaghan, 1999; 2000). These results hold in multiple experimental paradigms, such as relatively contextless matching paradigms (i.e., asking children to find the picture that "matches" the object; Wimmer et al., 2014) and more communicative matching paradigms (i.e., asking children to select a toy for the experimenter, based on the picture that the experimenter shows them; Callaghan, 1999). However, while 3-year-olds can match objects to black-and-white line drawings, they prefer to match objects to photographs and cartoons over black-and-white line drawings (Wimmer et al., 2014).

Furthermore, while children can learn various kinds of new information, such as words, actions, concepts, from photographs (Ganea et al., 2008; 2011; Simcock & DeLoache, 2006; 2008), there are mixed results suggesting that children may have more difficulty learning new information from black-and-white line drawings than color photographs (Simcock & DeLoache, 2006; 2008). Fifteen- and 18-month-olds can learn novel words from color photographs and generalize these novel words to objects (Ganea et al., 2008). Middle-class, predominantly White, 18-, 24-, and 30-month-olds living in Brisbane, Australia, can learn a novel action sequence from color photographs (Simcock & DeLoache, 2006; 2008). However, in this same context, only 30-month-olds succeed at learning the same novel action sequence from black-and-white line drawings (Simcock & DeLoache, 2006). Thus, children may have more difficulty recognizing the content of black-and-white line drawings than the content of photographs or cartoons.

When children do begin to recognize the content of black-and-white line drawings, their interpretations of these depictions are not based solely on visual resemblance. Rather, from 2.5 to 3 years of age onward, American and British children are sensitive to the role of the creator's intentions when interpreting the contents of black-and-white line drawings (Armitage & Allen, 2015; Bloom & Markson, 1998; Gelman & Ebeling, 1998; Hartley & Allen, 2014; Preissler & Bloom, 2008), suggesting that children use at least two different kinds of information (i.e., both perceptual and social cues) to recognize the content of black-and-white line drawings. Overall, recognizing the content of black-and-white line drawings may be a more complex capacity (i.e., involving the integration of multiple kinds of perceptual and social cues) than recognizing the content of more realistic kinds of visual media, and thus may emerge later in development.

In contrast, it is possible that children's capacity to understand that black-and-white line drawings refer to, rather than instantiate, objects in the world may emerge early in develop-

Fig. 2. Example of a black-and-white line drawing of a whisk (generated by ChatGPT), similar to the stimuli used in Preissler and Carey (2004).

ment, relative to the same kind of referential understanding for more realistic photographs and cartoons. As mentioned earlier, 9-month-olds tend to grasp at black-and-white line drawings, but this tendency is least frequent with black-and-white line drawings, somewhat more frequent with cartoons, and most frequent with color photographs (Pierroutsakos & DeLoache, 2003); however, it is also possible that 9-month-olds grasp less at black-and-white line drawings not because they understand these drawings as referential, but because they do not even recognize the content of these drawings. Preissler and Carey (2004) provide more compelling evidence of both recognition and referential understanding through an experiment teaching 18-month-olds in Boston and 24-month-olds in Boston and New York City a new word (i.e., "whisk") through the labeling of a black-and-white line drawing of a whisk (see Fig. 2). On the test trial, participants were shown the black-and-white line drawing of a whisk and an actual whisk, and asked to indicate the whisk. Almost all 18- and 24-month-olds selected either the actual whisk alone, or selected both the actual whisk and the black-and-white line drawing of a whisk (only one child, across four experiments, selected only the black-and-white line drawing of a whisk). This experiment showed that toddlers, and even older infants, understand that black-and-white line drawings refer to actual objects in the world. Thus, experiments demonstrating children's early emerging referential understanding typically use black-and-white line drawing (Preissler & Carey, 2004), whereas referential understanding experiments involving more realistic kinds of pictures, such as color photographs, produce less clear-cut evidence of successful comprehension (Ganea et al., 2009). However, further research—involving targeted experimental comparisons of referential understanding across different kinds of pictures—is necessary in order to conclusively determine whether children's referential understanding of black-and-white line drawings emerges earlier than their referential understanding of more realistic kinds of pictures.

Furthermore, researchers have conducted work on children's production—namely, their capacity to generate their own black-and-white line drawings (Callaghan, 1999; Long, Wang, Christie, Frank, & Fan, 2023; 2024; Wisher, Riede, Matthews, Pagnotta, & Tylén, in press).

While children's comprehension and production of visual media might be separate domains of development, children's drawing production capacities may still provide insight into their drawing comprehension capacities (i.e., just as children's language production capacities can sometimes provide insight into their language comprehension capacities; Clark & Hecht, 1983; Pickering & Garrod, 2013). Indeed, researchers have found that, in San Jose, United States, preschool- and elementary-aged children's capacities to produce and recognize black-and-white line drawings are correlated and improve over development (Long et al., 2024). Moreover, over the preschool and elementary school years, children's production of black-and-white line drawings become more recognizable and category-diagnostic over development, in both San Jose, United States, and Beijing, China (Long et al., 2023; 2024). By 2.5−3 years of age, children are sensitive to social cues such as communicative intent when interpreting line drawings made by other people (Bloom & Markson, 1998; Gelman & Ebeling, 1998; Preissler & Bloom, 2008). Moreover, by 5 years of age, children also tailor their production of line drawings based on pragmatic and communicative needs (Light & McEwen, 1987). Consequently, in both comprehension and production, children demonstrate a developing sensitivity to communicative intent, a critical component for understanding black-and-white line drawings (e.g., given that black-and-white line drawings bear less perceptual resemblance to their referents, viewers must consider artists' intentions when interpreting and recognizing the content of drawings). Overall, children's comprehension and production of black-and-white line drawings both improve over early childhood, and both might rely on similar cognitive underpinnings, such as conceptual knowledge and a sensitivity to communicative goals.

3.4. Scale models

Scale models are miniature replicas that are identical to their larger counterparts, except in size. We briefly summarize a large and robust literature showing that infants and toddlers make scale errors—namely, failure to use visual information about size when interacting these replicas, for example, by attempting to sit in a dollhouse chair or to get into a small toy car (Casler et al., 2011; DeLoache et al., 2004, 2013; Grzyb, Cagelosi, Cattani, & Floccia, 2017; Hagihara, Ishibashi, Moriguchi, & Shinya, 2022; Hunley & Hahn, 2016; Ishibashi & Moriguchi, 2017; Ishibashi, Twomey, Westermann, & Uehara, 2021; Rosengren, Gutiérrez, Anderson, & Schein, 2009; Rosengren, Schein, & Gutiérrez 2010; Ware et al., 2010). Children commit these scale errors across a range of cultural and economic contexts in midwestern America (Rosengren et al., 2010), and across disparate global contexts such as urban southeastern America (Hunley & Hahn, 2016), the U.K. (Grzyb et al., 2017; Ishibashi et al., 2021), and Japan (Hagihara et al., 2022; Ishibashi & Moriguchi, 2017; Ishibashi et al., 2021). Like infants' grasping behavior with pictures, toddlers' scale errors with scale models may demonstrate that they can recognize the content of the visual media, but are unable to understand that the visual media is referential. Another nonmutually exclusive possibility is that toddlers' poor inhibitory control does not prevent them from acting upon the replica as if it was an actual object, regardless of their actual knowledge about replicas and their small size (DeLoache et al., 2004). Regardless, it is well-documented that children have

difficulty responding appropriately to scale models in the first few years of life, specifically by recognizing the content of the scale model, but treating the scale model as an actual object and not a kind of visual media.

3.5. Videos

Videos are a kind of visual media featuring moving images. Typically, these moving images are captured with a camera or animated by an illustrator, such that stills of these images would be categorized as either photographs or cartoons, and thus bear high degrees of perceptual resemblance to actual objects. Infants demonstrate a distinction between live events over events depicted on a video screen (see Diener, Pierroutsakos, Troseth, & Roberts, 2008, for evidence from 9- to 10-month-olds, and Revencu & Csibra, 2021, for evidence from 19-month-olds). Thus, infants show an early emerging capacity to recognize the content of videos.

Still, infants seem to have difficulty understanding the referential nature of videos. Nine-month-olds grasp at objects depicted on video screens (Pierroutsakos & Troseth, 2003). Parental report data suggest that children make these kinds of media errors with some regularity in their first 2 years (Rosengren et al., 2021), and even 3-year-olds struggle to understand the property differences between video-depicted objects and their real-world referents (Flavell, Flavell, Green, & Korfmacher, 1990). Furthermore, toddlers have difficulty using information from videos as cues for retrieval in the actual world (Troseth & DeLoache, 1998), though live video experience improved retrieval task performance for White, middle-class toddlers in Champaign and Urbana, Illinois (Troseth, 2003). Moreover, despite the fact that children can recognize the content depicted on video screens, they may learn less effectively from video screens than from actual objects (Strouse & Samson, 2021), though research on this "video deficit effect" is an ongoing subject of debate and out of scope of the current review (see Zack, Barr, Gerhardstein, Dickerson, & Meltzoff, 2009 for evidence from middle-to-high-income American children in D.C. and Binghamton; and Sommer, Slaughter, Wiles, & Nielsen, 2023 for counterevidence from urban Australian children). Overall, children's understanding of video screens follows a similar developmental trajectory as their understanding of other kinds of visual media that are perceptually similar to actual objects in the world—namely, that these kinds of highly realistic visual media are easy to recognize, but difficult to understand as referential.

3.6. Interim summary

Our review of existing evidence suggests that visual media that bear high degrees of perceptual similarity to the physical world (i.e., color photographs, cartoons, scale models, videos) might be easier for children to recognize, but more difficult to understand as referential. In contrast, other kinds of visual media that bear less perceptual similarity to the physical world (i.e., black-and-white line drawings) may be more difficult for children to recognize, but easier to understand as referential.

Animal comprehension capacities

Many evolutionarily conserved capacities in humans may also be useful capacities for nonhuman animals. The degree to which various nonhuman animal species might also possess different kinds of visual media comprehension could provide important insight about these abilities' origins. Insofar as some kinds of visual media comprehension are shared widely across nonhuman animal species, the ability to make sense of them is more likely to be the product of evolutionarily conserved mechanisms. By contrast, other kinds of visual media comprehension that are uniquely human or are not found in a few nonhuman animal species might instead recruit perceptual and cognitive mechanisms that are strongly shaped by early environments.

Recognition

There is some evidence that the capacity to recognize the content of photographs and videos exists across species, specifically among nonhuman primates (Bovet & Vauclair, 2000), pigeons (Friedman, Spetch, & Ferrey, 2005), lizards (Jenssen, 1970), and fish (Kodric-Brown & Nicoletto, 1997), though the overall literature is mixed. Photograph and video recognition experiments involving nonhuman animals frequently use a restricted set of ecologically relevant stimuli, such as food and conspecifics (e.g., Kodric-Brown & Nicoletto, 1997; Savage-Rumbaugh, Rumbaugh, Smith, & Lawson, 1980). Moreover, contrary to experiments demonstrating successful photograph and video content recognition by nonhuman primates, other experiments show that marmoset monkeys fail to recognize photographs of food (Oh, Šlipogor, & Fitch, 2019) and capuchin monkeys fail to recognize photographs of conspecifics (Morton et al., 2016). While evidence for the recognition of content within highly realistic kinds of visual media in nonhuman animals is somewhat mixed, there is still more evidence for this kind of visual media comprehension capacity in nonhuman animals than for any other kinds of visual media comprehension capacities.

References

Nonhuman primates succeed on retrieval tasks involving photos (Hermann, Melis, & Tomasello, 2006; Kuhlmeier, Boysen, & Mukobi, 1999), videos (Poss & Rochat, 2003), and scale models (Hermann et al., 2006; Kuhlmeier & Boysen, 2002; Kuhlmeier et al., 1999), demonstrating some evidence for referential understanding of visual media. However, there is less evidence that other nonhuman animal species also succeed on retrieval tasks.

In terms of recognition, children might leverage their existing perceptual capacities for recognizing objects in the real world (Reynolds, 2015) to help recognize the content of visual media. Consequently, kinds of visual media that bear high degrees of perceptual similarity to the physical world may be the easiest for children to recognize. This claim contradicts previous ideas that abstract drawings may be even better than highly realistic ones for recognition

(Benson & Perrett, 1991; Rhodes, Brennan, & Carey, 1987) because black-and-white line drawings highlight the most relevant features for recognition (Fan, Yamins, & Turk-Browne 2018; Hertzmann, 2020; Sayim & Cavanagh, 2011), such as edges (Biederman & Ju, 1988).

In terms of reference, children must accurately identify something as visual media to understand its referential nature. Kinds of visual media that bear high degrees of perceptual similarity to the physical world may be more easily mistaken for actual objects, rather than accurately identified as visual media, and thus may be more difficult for children to understand as referential. However, these mechanistic claims are presently speculative: future research directly contrasting when and how children's recognition and referential understanding capacities develop across photographs, cartoons, black-and-white drawings, and other kinds of visual media is necessary in order to build more confidence in these theoretical claims.

4. Variation in development across learning environments

Some kinds of visual media comprehension capacities are not only early emerging in high-income Western infants, but also exhibit lower degrees of cross-contextual variability. This suggests that these capacities require less experience. In contrast, other kinds of visual media comprehension capacities require substantial experience and show higher levels of cross-cultural variability, and might instead recruit perceptual and cognitive mechanisms that are strongly shaped by environments. We tentatively suggest that the capacity to recognize the content of highly realistic kinds of visual media (e.g., photographs) may develop relatively consistently across environments, whereas the capacities to recognize the content of less realistic kinds of visual media (e.g., black-and-white line drawings) and understand the referential nature of visual media may be shaped by children's environments, specifically their experiences with visual media. Thus, in this section, we review child development research conducted across cultures and contexts, which provides initial evidence for consistency or variability of the development of visual media comprehension across early environments.

The capacity to recognize the content of highly realistic visual media may leverage broadly shared capacities for perceiving and acting upon the actual world. As we have discussed several times above, both middle-class White infants living in the United States and low-income Beng infants living in Cote D'Ivoire grasp at color photographs of objects (DeLoache et al., 1998). Though this experiment only involved eight Beng infants, this result initially suggested that infants' capacity to recognize the content of photographs is consistent across environments.

In contrast to this single finding on cross-cultural consistency in photograph comprehension, the vast majority of cross-cultural literature suggests that the capacity to understand many other kinds of visual media requires substantial experience and shows higher levels of cross-contextual variability. Indeed, evidence from adults' production of black-and-white line drawings suggests that even adults living in different cultures may possess somewhat different visual representations of basic-level categories (Lewis, Balamurugan, Zheng, & Lupyan, 2021). Relatedly, cross-cultural research suggests that the capacity to recognize the content of black-and-white line drawings, which often bear little perceptual similarity

to actual objects, may be a relatively late-emerging capacity. Specifically, middle-income Argentinian children succeeded at referent-picture matching tasks with black-and-white line drawings at 30 months, approximately 1 year earlier than low-income Argentinian children, who succeeded at the same task at 42 months (Salsa & Vivaldi, 2017). Though the precise mechanism driving the differing performance between the two groups of children is unclear, these results suggest that some kind of environmental factors may impact children's developing ability to recognize the content of black-and-white line drawings. Moreover, 2- and 3-year-olds from a Western convenience sample improved their performance on a black-and-white line drawing recognition task when given longitudinal training (Callaghan & Rankin, 2002), thus providing further support for the important role of experience in the capacity to recognize the content of less realistic visual media.

Critically, variation in visual media experience may relate not only to variation in children's visual media comprehension, but also their capacity to learn from visual media and to perform accurately on experimental tasks involving visual media. For example, 15- to 29-month-old infants and toddlers in Kwala, a rural village in the Kibaha-Pwani District of Tanzania's Coast Region—an environment with very few pictures—were able to learn a new word for an actual object, but not for a color photograph. By 31−38 months, Kwala toddlers and preschoolers were able to learn a new word from both an actual object and a color photograph (Walker, Walker, & Ganea, 2013). Similarly, research conducted with toddlers living in urban Kisumu city, in Western Kenya, showed that toddlers' experience with pictures (i.e., the frequency of looking at picture books with their caregivers in the past week) related to their success on a cartoon-based word learning task (Zhu, Pitchik et al., 2025).

Moreover, children with relatively little picture experience may also perform less accurately on experiments involving picture stimuli than on experiments involving object stimuli. For example, while preschoolers living in rural environments in India (i.e., a village 70 kilometers from Vijayawada, Andhra Pradesh) and Peru (i.e., a village in the rural Montaro Valley area of the Central Highlands) performed similarly to their rural middle-class Canadian counterparts on a false belief task involving objects, but performed worse than their rural Canadian counterparts on a false belief task involving black-and-white line drawings (Callaghan et al., 2012). Similarly, while middle-to-high-income toddlers in Berkeley, United States performed equally accurately on a vocabulary assessment involving objects and a vocabulary assessment involving cartoons, low-to-middle-income toddlers and preschoolers in Mombasa County, Kenya performed more accurately on a vocabulary assessment involving objects than a vocabulary assessment involving cartoons (Zhu, Kilonzo, Engelmann, & Gopnik, in press). These experiments provide preliminary evidence of environmental variability not only in children's basic visual media comprehension skills, but also in consequent learning and assessment tasks. Thus, the capacities not only to understand, but to use, some kinds of less realistic visual media may rely on children's learning environment and experiences.

Moreover, the capacity to understand visual media as referential may rely to an especially large degree on children's early experiences. Indeed, if it is the case that realistic visual media are easy for children to recognize because these kinds of visual media leverage humans' existing visual capacity to perceive the actual world, it is sensible that children would

then mistake realistic visual media for objects in the actual world. Consequently, children may require a substantial amount of visual media experience in order to understand that visual media *refers to*, rather than *instantiates*, actual objects. This claim is consistent with previously mentioned evidence on difficulties in symbol use (DeLoache, 1987; DeLoache et al., 1998), and with previous research showing that prior experience with the referential understanding of one kind of visual media (i.e., a picture search task) facilitates toddlers' successful referential understanding of another kind of visual media (i.e., a scale model search task; DeLoache, 1991). Moreover, research with British and Japanese 18- to 24-month-olds shows some evidence of cross-cultural consistency in children's likelihood of committing scale errors, suggesting that infants and toddlers across disparate cultures struggle to understand the referential nature of visual media (Ishibashi et al., 2021). Thus, there is evidence for a more protracted, and possibly highly experience-dependent, developmental trajectory for children's referential understanding of various kinds of visual media, relative to their developmental trajectory of the recognition of highly realistic visual media.

Finally, the capacity to produce visual media also varies across early environments. For example, middle-class children, ages 2.5−5 years, in Canada succeeded at picture drawing tasks approximately 1 year earlier than their counterparts in India and Peru (Callaghan et al., 2011). Children in Beijing, China, ages 4−9 years, also produce more recognizable drawings than their American peers in the San Francisco Bay Area (Long et al., 2023). Thus, there is evidence of variability in children's capacity to produce visual media across cultures and contexts. Overall, while some kinds of visual media comprehension might be evolutionarily conserved, initial empirical work suggests that there is a large degree of cross-contextual variability in many other kinds of visual media comprehension. Thus, more work is needed to delineate exactly what experiences might drive this variability.

5. Implications for educational practice

Given the evidence for variability in visual media comprehension, researchers should further investigate cross-contextual variation in children's visual media comprehension. This research is not only necessary for the articulation of better theories of visual media development, but also to improve applied education research and policy. Specifically, popular learning materials (e.g., books, posters) and assessments of early cognitive and linguistic abilities (e.g., Bayley Scales of Infant and Toddler Development, Wechsler Preschool and Primary Scale of Intelligence, Denver Developmental Screening Test) frequently use stimuli involving visual media (Fernald et al., 2017). As these learning materials and assessments spread globally, educators and researchers make an implicit assumption: that children across cultures and contexts understand these kinds of visual media in the same way, at the same developmental time point. What if this assumption does not hold? Differences in early visual media comprehension can drastically change the efficacy of learning materials and the validity of assessments (Cromwell et al., 2014; Jukes et al., 2024; Liddell, 1996; 1997; Ranjitkar et al., 2018; Shin, Han, Oh, Oh, & Ha, 2002). Some children may mainly encounter pictures in school materials, and consequently may not learn as effectively from existing school materials or perform

well on standard assessments. Acquiring knowledge about the cross-contextual development of picture comprehension will help teachers and policy-makers tailor childhood education programs (e.g., formal school programs, home visiting programs, parenting interventions) to children from various backgrounds, and appropriately assess these children's abilities. Thus, in order to appropriately translate learning materials and assessments globally (Draper et al., 2023), it is important to investigate when and how children understand pictures across cultures and contexts.

Indeed, initial studies suggest that children living in environments without many pictures succeed on cognitive tasks involving objects, but underperform on cognitive tasks involving pictures (Callaghan et al., 2012; Walker et al., 2013; Zhu et al., in press; Zuilkowski et al., 2016). Indeed, there is some evidence of differences between object- and picture-based task performance even in children and adults in Western convenience samples (Beaucage et al., 2020; Gomez, Skiba, & Snow, 2018; Sensoy et al., 2021; Snow & Culham, 2021), and these differences might be exacerbated in picture-sparse environments. Since children in picture-sparse environments often still succeed, to some degree, on picture-based tasks (e.g., Zhu et al., in press; Zhu, Kilonzo et al., 2025), researchers may have assumed that picture-based tasks are valid in these environments. However, picture-based tasks may actually underestimate children's performance, relative to object-based tasks.

A better understanding of the experiences facilitating visual media comprehension will also lead to more effective childhood learning programs. For example, researchers have shown the benefits of parent−child picture book-sharing programs on children's language acquisition in global contexts (Knauer, Jakiela, Ozier, Aboud, & Fernald, 2020), but it is unclear whether the benefits arise from an increase in linguistic comprehension, visual media comprehension, or both. Researchers should also closely collaborate with local community members and stakeholders who hold greater cultural and linguistic competency (Eirdosh et al., 2025): for example, local teachers may already be attuned to the characteristics of their classrooms, and may have already adapted curricula accordingly. With a better understanding of the mechanisms facilitating visual media comprehension, and a better understanding of the mechanisms driving stimuli-based task performance differences, researchers can work with parents, teachers, policymakers, and other local and global stakeholders in order to create interventions or policy changes to benefit children in global contexts.

6. Future directions

The development of picture comprehension across cultures and contexts is still a relatively new and emerging area of research investigation. In this section, we highlight future directions for empirical research on variability in the development of picture comprehension, as well as its relation to other early childhood cognitive capacities.

First, while previous research has investigated differences in performance between object- and visual media-based learning and assessment tasks, more research is necessary to investigate the generalizability of this effect across contexts, age ranges, and kinds of visual media. For example, much of this work demonstrating differences in object- and picture-based

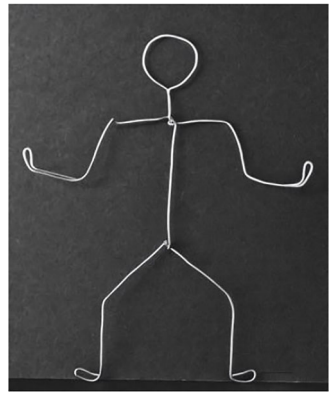

Fig. 3. Example of a wire model of a person (generated by ChatGPT), similar to the stimuli used in Serpell (1979).

learning and assessment has been conducted only with children prior to, or just beginning, formal schooling (Callaghan et al., 2012; Walker et al., 2013; Zhu et al., in press) and these differences may not hold for older children with more formal schooling experience (Zhu et al., 2025). By investigating the generalizability of this effect, researchers will also be able to generate mechanistic explanations—that is, articulating exactly what quality and quantity of experiences that lead to visual media comprehension—with the same precision as in other domains of development (e.g., language learning; Kachergis, Marchman, & Frank, 2022). For example, acquiring an understanding of visual media may require direct interactions with adults and explicit conversations about visual media as referential (Callaghan, 2020). However, it is also possible that child-led experiences—for example, passive experience in environments with frequent visual media (e.g., seeing posters on the wall or watching television) or active experience creating visual media (e.g., drawing, painting)—might facilitate children's visual media comprehension. Thus, another promising line of future research involves conducting training studies investigating whether providing young children with various kinds of visual media experience (i.e., passive viewing, active drawing, social scaffolding) might improve their performance on tasks involving visual media.

Moreover, variables other than picture experience (e.g., age) may contribute to the development of visual media comprehension. Differences in performance between object- and visual media-based learning and assessment tasks may be driven by differences in familiarity, rather than differences in comprehension. For example, urban Zambian children perform more accurately on a pattern reproduction task with wire models than with drawings, whereas urban British children perform more accurately on the same task with drawings rather than wire models (see Fig. 3; Serpell, 1979). Indeed, this variation may not be limited to under-represented global contexts: future research might also investigate the possibility of unexplored variation within Western convenience sample contexts (e.g., between children growing up in middle- and high-income contexts, or urban, suburban, and rural contexts). Thus, researchers might also consider conducting longitudinal studies tracking the developmental trajectory of children's visual media comprehension and its relation to environmental

factors (e.g., visual media experience, socioeconomic status, neighborhood demographics, and formal schooling experience).

Advances in psychological methods, such as the development of head-mounted cameras, also allow researchers to collect richer data relating to young children's everyday experiences, and thus more accurately characterize their early visual, social, and linguistic environments (Long et al., 2024). Moreover, gathering rich datasets characterizing the variability in children's environments and the development of visual media comprehension across contexts may be useful not only for researchers in developmental psychology and education, but also for computer scientists working to efficiently train artificial intelligence systems (Long et al., 2024) and neuroscientists seeking to better understand developing brain networks (O'Doherty et al., 2025).

In addition to a need for empirical research on variability in visual media comprehension across contexts, there is also a need for empirical research on variability in visual media comprehension across different kinds of visual media. For example, in the present article, we speculate that children sometimes respond differently to different kinds of visual media (Pierroutsakos & DeLoache, 2003) because some kinds of visual media bear more or less perceptual resemblance to actual objects in the world. However, more research is required to test these claims. For instance, future research might directly manipulate the degree of perceptual resemblance between visual media and actual objects, across and within kinds of visual media, to demonstrate that the degree of perceptual resemblance is actually the variable driving some of children's behavioral differences across kinds of visual media.

Moreover, while the present review outlines five different kinds of visual media (i.e., photographs, cartoons, black-and-white line drawings, videos, and scale models), more work on how children interact with other kinds of visual media (e.g., smartphones, tablets, computers, and conversational artificial intelligence) is also necessary. Various kinds of interactive visual media are becoming an increasingly prevalent aspect of young children's daily lives (Coyne et al., 2023), as well as experimental paradigms in developmental psychology (e.g., Frank et al., in press); consequently, researchers should systematically investigate how and when children understand and use not only traditional kinds of visual media, but also newer kinds of interactive visual media technology (Dietz, Outa, Lowe, Landay, & Gweon, 2023). Variation in children's experience with different kinds of visual media across contexts may also lead to variations in the degree of difference in visual media comprehension trajectories across contexts. For example, children in most global contexts have some access to picture books, but some children have notably less access to novel kinds of interactive artificial intelligence. Thus, contextual variation in the development of visual media comprehension may be larger for some kinds of visual media than others.

Moreover, future research should investigate the relation between children's developing visual media comprehension and other cognitive capacities. For example, U.S. preschoolers and U.S. adults use both visual and social information to determine the meaning of black-and-white line drawings (Bloom & Markson, 1998; Gelman & Ebeling, 1998; Hawkins, Sano, Goodman, & Fan, 2023), suggesting that, at least in some instances, visual media comprehension relies on social cognition. Children's sensitivity to social information may be particularly important in instances where the visual media bears less perceptual resemblance

to actual objects (e.g., black-and-white line drawings), and less important in instances where the visual media bears high perceptual resemblance to actual objects (e.g., photographs, scale models). Since a sensitivity to communicative intent develops early in infancy (Liszkowski et al., 2007), it is possible that toddlers' and preschooler's capacity to understand some kinds of relatively abstract visual media builds on these existing, earlier emerging kinds of social, communicative capacities. However, more research is required to articulate whether social cognitive capacities are necessary for visual media comprehension, and also whether social cognitive capacities might be more important for understanding some kinds of visual media than others.

Similarly, future research might investigate the relation between children's developing visual media comprehension and their executive function skills. While children's developing executive function capacities (e.g., working memory, inhibitory control, cognitive flexibility) are related to many kinds of linguistic and cognitive capacities (Carriedo et al., 2016; Emslander & Scherer, 2022; Setoh, Scott, & Baillargeon, 2016), some components of executive function, such as inhibitory control, might be particularly important for visual media comprehension (Kuhlmeier, 2005). Specifically, in experimental tasks testing referential understanding of highly realistic kinds of visual media such as scale models and photographs, children must suppress a prepotent response that would lead to a scale error or grasping error (i.e., treating a visual representation as an instantiation of a kind) in order to perform the correct response (i.e., treating a visual representation as a representation of a kind). In other domains, such as theory of mind, researchers have found success at earlier developmental time points using experimental paradigms with reduced processing demands (Setoh et al., 2016). Thus, it is possible that in the domain of visual media comprehension, generating new experimental paradigms with reduced processing demands may also show evidence of referential understanding at earlier developmental time points. Overall, more research is needed to articulate when and how other cognitive capacities (i.e., social cognition, executive function) might influence children's visual media comprehension.

7. Conclusion

This paper reviews evidence of children's developing capacities to understand five kinds of visual media, specifically photographs, cartoons, black-and-white line drawings, videos, and scale models. We suggest that visual media comprehension involves two capacities: the capacity to recognize the content of visual media and the capacity to understand visual media as referential. Previous research suggests that realistic kinds of visual media (e.g., photographs, cartoons) may be easy to recognize but more difficult to understand as referential, whereas less realistic kinds of visual media (e.g., black-and-white line drawings) may be more difficult to recognize but easier to understand as referential, though more empirical work is necessary to build full confidence in these claims. Moreover, we review some initial evidence of variation in children's developing visual media comprehension across cultures and contexts, potentially due to variation in their experiences with visual media. Future research should investigate how and when children's visual media compre-

hension varies across contexts, in order to create more contextually appropriate learning materials and assessment tools, and thus improve early childhood research and education worldwide.

Acknowledgments

This work was supported by an SSHRC Postdoctoral Fellowship to RZ.

Note

1 Instances where children learn from visual media and then apply the information to real-world objects may provide evidence of both recognition and referential understanding capacities. However, instances where children learn from visual media and then apply the information to visual media may provide evidence of only recognition capacities.

References

Apperly, I. A., Williams, E., & Williams, J. (2004). Three- to four-year-olds' recognition that symbols have a stable meaning: Pictures are understood before written words. *Child Development*, *75*, 1510–1522.

Armitage, E., & Allen, M. L. (2015). Children's picture interpretation: Appearance or intention? *Developmental Psychology*, *51*, 1201–1215.

Barrera, M. E., & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development*, *52*, 203–206.

Beaucage, N., Skolney, J., Hewes, J., & Vongpaisal, T. (2020). Multisensory stimuli enhance 3-year-old children's executive function: A three-dimensional object version of the standard dimensional change card sort. *Journal of Experimental Child Psychology*, *189*, 104694.

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, *3*, 105–135.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38–64.

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, *9*, 200–204.

Bovet, D., & Vauclair, J. (2000). Picture recognition in animals and humans. *Behavioral Brain Research*, *109*, 143–165.

Callaghan, T. (1999). Early understanding and production of graphic symbols. *Child Development*, *70*, 1314–1324.

Callaghan, T. (2000). Factors affecting children's graphic symbol use in the third year: Language, similarity, and iconicity. *Cognitive Development*, *15*, 185–214.

Callaghan, T. (2020). The origins and development of a symbolic mind: The case of pictorial symbols. *Interchange*, *51*, 53–64.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., & Liszkowski, U. (2011). Early social cognition in three cultural contexts. *Monographs of the Society for Research in Child Development*, *76*, 1–128.

Callaghan, T., & Rankin, M. P. (2002). Emergence of graphic symbol functioning and the question of domain specificity: A longitudinal training study. *Child Development*, *73*, 359–376.

Callaghan, T., Rochat, P., & Corbit, J. (2012). Young children's knowledge of the representational function of pictorial symbols: Development across the preschool years in three cultures. *Journal of Cognition and Development*, *13*, 320–353.

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS ONE*, *11*, e0150289.

Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental Psychology*, *47*, 857–866.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology*, *35*, 325–349.

Coyne, S., Gale, M., Ashby, S., Memmott-Elison, M. K., Holmgren, H. G., Barr, R., Christensen-Duerden, C., & Brown, S. (2023). Media in the moment: An observational assessment of the digital media context in early childhood. *Translational Issues in Psychological Science*, *9*, 186–198.

Cromwell, E. A., Dube, Q., Cole, S. R., Chirambo, C., Dow, A. E., Heyderman, R., & Van Rie, A. (2014). Validity of U.S. norms for the Bayley Scales of Infant Development-III in Malawian children. *European Journal of Pediatric Neurology*, *18*, 223–230.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*, 1556–1557.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, *62*, 736–752.

DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Psychological Science*, *4*, 109–113.

DeLoache, J. S. (2003). Dual representation and young children's use of scale models. *Child Development*, *71*, 329–338.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition*, *52*, 83–110.

DeLoache, J. S., LoBue, V., Vanderborght, M., & Chiong, C. (2013). On the validity and robustness of the scale error phenomenon in early childhood. *Infant Behavior and Development*, *36*, 63–70.

DeLoache, J. S., Pierroutsakos, S. L., & Uttal, D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science*, *12*, 114–118.

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science*, *9*, 205–210.

DeLoache, J. S., Strauss, M. S., & Maynard, J. (1979). Picture perception in infancy. *Infant Behavior & Development*, 2, 77–89.

DeLoache, J. S., Uttal, D. H., & Pierroutsakos, S. L. (2000). What's up? The development of an orientation preference for picture books. *Journal of Cognition and Development*, *1*, 81–95.

DeLoache, J. S., Uttal, D. H., & Rosengren, K. S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science*, *304*, 1027–1029.

Diener, M. L., Pierroutsakos, S. L., Troseth, G. L., & Roberts, A. (2008). Video versus reality: Infants' attention and affective responses to video and live presentations. *Media Psychology*, *11*, 418–441.

Dietz, G., Outa, J., Lowe, L., Landay, J. A., & Gweon, H. (2023). Theory of AI Mind: How adults and children reason about the "mental states" of conversational AI. In *Proceedings of the 45th Annual Conference of the Cognitive Science Society* (pp. 3367–3374). Cognitive Science Society.

Dirks, J., & Gibson, E. (1977). Infants' perception of similarity between live people and their photographs. *Child Development*, *48*, 124–130.

Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, G., Obradović, J., Scerif, G., Valentini, N. C., Venetsanou, F., & Yousafzai, A. K. (2023). Publishing child development research from around the world: An unfair playing field resulting in most of the world's child population under-represented in research. *Infant and Child Development*, *32*, e2375.

Eirdosh, D., Prasetijo, A., Aprilia, C., Greenfield, P. M., Lavi, N., Muchukunnu, A., Guerrero-Meyer, B., Hanisch, S., Kakkoth, S., Chellan, V., Simatende, B., Trumbull, E., & Rothstein-Fisch, C. (2025). *Guide to cross-cultural research on childhood learning: Theoretical, methodological, practical, and ethical considerations for an interdisciplinary field*. Cambridge: Open Book Publishers.

Emslander, V., & Scherer, R. (2022). The relation between executive functions and math intelligence in preschool children: A systematic review and meta-analysis. *Psychological Bulletin*, *148*, 337–369.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*, 2670–2698.

Fernald, L. C. H., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development*. Washington, DC: World Bank.

Flavell, J. H., Flavell, E. R., Green, F. L., & Korfmacher, J. E. (1990). Do young children think of television images as pictures or real objects? *Journal of Broadcasting & Electronic Media*, *34*, 399–419.

Frank, M. C., Baumgartner, H., Braginsky, M., Kachergis, G., Lightbody, A., Sparks, R. Z., Zhu, R., Carlson, S. M., Graham, S., Lipina, S. J., Newcombe, N. S., Odgers, C. L., Pianta, R. C., Siegler, R. S., Snowling, M., Yoshikawa, C., A., & Dodge, K. A. (in press). Learning Variability Network Exchange (LEVANTE): A global framework for measuring children's learning variability through collaborative data sharing. *Child Development*, *96*, 1867–1884. https://doi.org/10.1111/cdev.70011

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.

Friedman, A., Spetch, M. L., & Ferrey, A. (2005). Recognition by humans and pigeons of novel views of 3-D objects and their photographs. *Journal of Experimental Psychology: General*, *134*, 149–162.

Ganea, P. A., Ma, L., & DeLoache, J. S. (2011). Young children's learning and transfer of biological information from picture books to real animals. *Child Development*, *82*, 1421–1433.

Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, *9*, 46–66.

Ganea, P. A., Preissler, M. A., Butler, L., Carey, S., & DeLoache, J. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology*, *104*, 283–295.

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, *66*, 35–47.

Gombrich, E. H. (1969). *Art and illusion: A study in the psychology of pictorial representation*. Princeton, NJ: Princeton University Press.

Gomez, M. A., Skiba, R. M., & Snow, J. C. (2018). Graspable objects grab attention more than images do. *Psychological Science*, *29*, 206–218.

Grzyb, B. J., Cagelosi, A., Cattani, A., & Floccia, A. (2017). Decreased attention to object size information in scale errors performers. *Infant Behavior and Development*, *47*, 72–82.

Hagihara, H., Ishibashi, M., Moriguchi, Y., & Shinya, Y. (2022). Object labeling activates young children's scale errors at an early stage of verb vocabulary growth. *Journal of Experimental Child Psychology*, *222*, 105471.

Hartley, C., & Allen, M. A. (2014). Intentions vs. resemblance: Understanding pictures in typical development and autism. *Cognition*, *131*, 44–59.

Hartley, C., & Allen, M. A. (2015). Iconicity influences how effectively minimally verbal children with autism and ability-matched typically developing children use pictures as symbols in a search task. *Autism*, *19*, 570–579.

Hawkins, R. D., Sano, M., Goodman, N. D., & Fan, J. E. (2023). Visual resemblance and interaction history jointly constrain pictorial meaning. *Nature Communications*, *14*, 2199.

Hermann, E., Melis, A. P., & Tomasello, M. (2006). Apes' use of iconic cues in the object-choice task. *Animal Cognition*, *9*, 118–130.

Hertzmann, A. (2020). Why do line drawings work? A realism hypothesis. *Perception*, *49*, 439–451.

Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *American Journal of Psychology*, *75*, 624–628.

Hunley, S. B., & Hahn, E. R. (2016). Labels affect preschoolers' tool-based scale errors. *Journal of Experimental Child Psychology*, *151*, 40–50.

Ishibashi, M., & Moriguchi, Y. (2017). Understanding why children commit scale errors: Scale error and its relation to action planning and inhibitory control, and the concept of size. *Frontiers in Psychology*, *8*, 826.

Ishibashi, M., Twomey, K. E., Westermann, G., & Uehara, I. (2021). Children's scale errors and object processing: Early evidence for cross-cultural differences. *Infant Behavior and Development*, *65*, 101631.

Jenssen, T. A. (1970). Female response to filmed displays of *Anolis nebulosus* (Sauria, Iguanidae). *Animal Behavior*, *18*, 640–647.

Jiang, M. J., & Rosengren, K. S. (2019). Action errors: A window into the early development of perception-action system. *Advances in Child Development and Behavior*, *55*, 145–171.

Johnson, C. R., Flores, I., & Troseth, G. L. (2021). Do young children of the "selfie generation" understand digital photos as representations? *Human Behavior and Emerging Technologies*, *3*, 512–524.

Jukes, M. C. H., Ahmed, I., Baker, S., Draper, C. E., Howard, S. J., McCoy, D. C., Obradović, J., & Wolf, S. (2024). Principles for adapting assessments of executive function across cultural contexts. *Brain Sciences*, *14*, 318.

Kachergis, G., Marchman, V. A., & Frank, M. C. (2022). Toward a "standard model" of early language learning. *Current Directions in Psychological Science*, *31*, 20–27.

Knauer, H. A., Jakiela, P., Ozier, O., Aboud, F. E., & Fernald, L. C. H. (2020). Enhancing young children's language acquisition through parent-child book-sharing: A randomized trial in rural Kenya. *Early Childhood Research Quarterly*, *50*, 179–190.

Kodric-Brown, A., & Nicoletto, P. F. (1997). Repeatability of female choice in the guppy: Response to live and videotaped males. *Animal Behavior*, *54*, 369–376.

Kuhlmeier, V. A. (2005). Symbolic insight and inhibitory control: Two problems facing young children on symbolic retrieval tasks. *Journal of Cognition and Development*, *6*, 365–380.

Kuhlmeier, V. A., & Boysen, S. T. (2002). Chimpanzees *(Pan troglodytes)* recognize spatial and object correspondences between a scale model and its referent. *Psychological Science*, *13*, 60–63.

Kuhlmeier, V. A., Boysen, S. T., & Mukobi, K. L. (1999). Scale-model comprehension by chimpanzees *(Pan troglodytes)*. *Journal of Comparative Psychology*, *113*, 396–402.

Lewis, M., Balamurugan, A., Zheng, B., & Lupyan, G. (2021). Characterizing variability in shared meaning through millions of sketches. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (pp. 1970–1976). Cognitive Science Society.

Liddell, C. (1996). Every picture tells a story: South African and British children interpreting pictures. *British Journal of Developmental Psychology*, *14*, 355–363.

Liddell, C. (1997). Every picture tells a story – Or does it? Young South African children interpreting pictures. *Journal of Cross-Cultural Psychology*, *28*, 266–283.

Light, P., & McEwen, F. (1987). Drawings as messages: The effect of a communication game upon production of view-specific drawings. *British Journal of Developmental Psychology*, *5*, 53–59.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language*, *34*, 1–20.

Long, B., Fan, J. E., Huey, H., Chai, Z., & Frank, M. C. (2024). Parallel developmental changes in children's production and recognition of line drawings of visual concepts. *Nature Communications*, *15*, 1191.

Long, B., Goodin, S., Kachergis, G., Marchman, V., Radwan, S., Sparks, R., Xiang, V., Zhuang, C., Hsu, O., Newman, B., Yamins, D. L. K., & Frank, M. C. (2024). The BabyView Camera: Designing a new head-mounted camera to capture children's early social and visual environment. *Behavior Research Methods*, *56*, 3523–3534.

Long, B., Wang, Y., Christie, S., Frank, M. C., & Fan, J. E. (2023). Developmental changes in drawing production under different memory demands in a U.S. and Chinese sample. *Developmental Psychology*, *59*, 1784–1793.

Long, B., Xiang, V., Stojanov, S., Sparks, R. Z., Yin, Z., Keene, G. E., Tan, A. W. M., Feng, S. Y. Zhuang, C., Marchman, V. A., Yamins, D. L. K., & Frank, M. C. (2024). The BabyView dataset: High resolution egocentric videos of infants' and young children's everyday experiences. *ArXiv*.

Morton, F. B., Brosnan, S. F., Pretot, L., Buchanan-Smith, H. M., O'Sullivan, E., Stocker, M., D'Mello, D., & Wilson, V. A. D. (2016). Using photographs to study animal social cognition and behaviour: Do capuchins' responses to photos reflect reality? *Behavioural Processes*, *124*, 38–46.

O'Doherty, C., Dineen, A. T., Truzzi, A., King, G., D'Arcy, E., Caldinelli, C., Holloway, T., Molloy, E., & Cusack, R. (2025). Models trained on infant views are more predictive of infant visual cortex. In *Proceedings of the Computational Cognitive Neuroscience Conference*.

Oh, J., Šlipogor, V., & Fitch, W. T. (2019). Artificial visual stimuli for animal experiments: An experimental evaluation in a prey capture context with common marmosets (*Callithrix jacchus*). *Journal of Comparative Psychology*, *133*, 72–80.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Brain and Behavioral Sciences*, *36*, 329–347.

Pierroutsakos, S. L., & DeLoache, J. S. (2003). Infants' manual exploration of pictorial objects varying in realism. *Infancy*, *4*, 141–156.

Pierroutsakos, S. L., & Troseth, G. L. (2003). Video verité: Infants' manual investigation of objects on video. *Infant Behavior and Development*, *26*, 183–199.

Poss, S. R., & Rochat, P. (2003). Referential understanding of videos in chimpanzees (*Pan troglodytes*), orangutans (*Pongo pygmaeus*), and children (*Homo sapiens*). *Journal of Comparative Psychology*, *117*, 420–428.

Preissler, M. A., & Bloom, P. (2007). Two-year-olds appreciate the dual nature of pictures. *Psychological Science*, *18*, 1–2.

Preissler, M. A., & Bloom, P. (2008). Two-year-olds use artist intention to understand drawings. *Cognition*, *106*, 512–518.

Preissler, M. A., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development*, *5*, 185–212.

Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. M., Shrestha, L., & Hysing, M. (2018). Acceptability and reliability of the Bayley Scales of Infant and Toddler Development-III among children in Bhaktapur, Nepal. *Frontiers in Psychology*, *9*, 1265.

Revencu, B., & Csibra, G. (2021). For 19-month-olds, what happens on-screen stays on-screen. *Open Mind*, *5*, 71–90.

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, *285*, 34–43.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, *19*, 473–497.

Rose, S. A. (1977). Infants' transfer of response between two-dimensional and three-dimensional stimuli. *Child Development*, *48*, 1086–1091.

Rosengren, K. S., Gutiérrez, I. T., Anderson, K. N., & Schein, S. S. (2009). Parental reports of children's scale errors in everyday life. *Child Development*, *80*, 1586–1591.

Rosengren, K. S., Kirkorian, H., Choi, K., Jiang, M. J., Raimer, C., Tolkin, E., & Sartin-Tarm, A. (2021). Attempting to break the fourth wall: Young children's action errors with screen media. *Human Behavior and Emerging Technologies*, *3*, 468–483.

Rosengren, K. S., Schein, S. S., & Gutiérrez, I. T. (2010). Individual differences in children's production of scale errors. *Infant Behavior and Development*, *33*, 309–313.

Saitoti, T. O. (1986). *The worlds of a Maasai warrior: An autobiography*. Berkeley, CA: University of California, Press.

Salsa, A. M., & Vivaldi, R. A. (2017). Developmental changes in early comprehension and production of drawings: Evidence from two socioeconomic backgrounds. *Journal of Genetic Psychology*, *178*, 217–228.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., Smith, S. T., & Lawson, J. (1980). Reference: The linguistic essential. *Science*, *210*, 920–925.

Sayim, B., & Cavanagh, P. (2011). What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, *5*, 118.

Sensoy, O., Culham, J. C., & Schwarzer, G. (2021). The advantage of real objects over matched pictures in infants' processing of the familiar size of objects. *Infant and Child Development*, *30*, e2234.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology*, *70*, 365–380.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, *113*, 13360–13365.

Shin, H., Han, K., Oh, K., Oh, J., & Ha, M. (2002). Denver II Developmental Screening Test: A cross cultural comparison. *Journal of Korean Academy of Community Health Nursing*, *13*, 89–97.

Simcock, G., & DeLoache, J. S. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology*, *42*, 1352–1357.

Simcock, G., & DeLoache, J. S. (2008). The effect of repetition on infants' imitation from picture books varying in iconicity. *Infancy*, *13*, 687–697.

Slater, A., Morison, V., & Rose, D. (1984). New-born infants' perception of similarities and differences between two- and three-dimensional stimuli. *British Journal of Developmental Psychology*, *2*, 287–294.

Sloutsky, V. (2009). Theories about 'theories': Where is the explanation? Comment on Waxman and Gelman. *Trends in Cognitive Science*, *13*, 331–332.

Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Science*, *25*, 506–519.

Sommer, K., Slaughter, V., Wiles, J., & Nielsen, M. (2023). Revisiting the video deficit in technology-saturated environments: Successful imitation from people, screens, and social robots. *Journal of Experimental Child Psychology*, *232*, 105673.

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development*, *92*, e20–e38.

Tare, M., Chiong, C., Ganea, P., & DeLoache, J. S. (2010). Less is more: How manipulative features affect children's learning from picture books. *Journal of Applied Developmental Psychology*, *31*, 395–400.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*, 705–722.

Troseth, G. L. (2003). TV guide: Two-year-old children learn to use video as a source of information. *Developmental Psychology*, *39*, 140–150.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*, 950–965.

Uttal, D. H., Gentner, D., Liu, L. L., & Lewis, A. R. (2008). Developmental changes in children's understanding of the similarity between photographs and their referents. *Developmental Science*, *11*, 156–170.

Walker, C. M., Walker, L. B., & Ganea, P. A. (2013). The role of symbol-based experience in early learning and transfer from pictures: Evidence from Tanzania. *Developmental Psychology*, *49*, 1315–1324.

Ware, E. A., Uttal, D. H., & DeLoache, J. S. (2010). Everyday scale errors. *Developmental Science*, *13*, 28–36.

Wimmer, M. C., Robinson, E. J., Koenig, L., & Corder, E. (2014). Getting the picture: Iconicity does not affect representation-referent confusion. *PLoS ONE*, *9*, e107910.

Wisher, I., Riede, F., Matthews, J., Pagnotta, M., & Tylén, K. (in press). *Children as playful artists: Integrating developmental psychology to identify children's art in the Upper Palaeolithic*. Hunter Gatherer Research.

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.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, *35*, 41–68.

Zhu, R., Kilonzo, T. N., Arieda, J. O., Verma, A., Fan, J. E., & Frank, M. C. (2025). Investigating children's performance on object- and picture-based vocabulary assessments in global contexts: Evidence from Kisumu, Kenya. In *Proceedings of the 47th Annual Conference of the Cognitive Science Society* (pp. 3126–3132). Cognitive Science Society.

Zhu, R., Kilonzo, T. N., Engelmann, J., & Gopnik, A. (2026). Investigating the validity of picture-based assessments across cultures and contexts: Evidence from young children in Kenya and the U.S. *Developmental Psychology*. https://doi.org/10.1037/dev0002050.

Zhu, R., Pitchik, H. O., Kilonzo, T. N., Engelmann, J., Fernald, L., & Gopnik, A. (2025). The development of picture comprehension across early environments: Evidence from urban and rural toddlers in Western Kenya. *Developmental Science*, *28*, e13579.

Ziemer, C. J., Plumert, J. M., & Pick, A. D. (2012). To grasp or not to grasp: Infants' actions towards objects and pictures. *Infancy*, *17*, 479–497.

Zuilkowski, S. S., McCoy, D. C., Serpell, R., Matafwali, B., & Fink, G. (2016). Dimensionality and the development of cognitive assessments for children in sub-Saharan Africa. *Journal of Cross-Cultural Psychology*, *47*, 341–354.