Routledge
Taylor & Francis Group

Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: https://www.tandfonline.com/loi/hlld20

How Children Identify Events from Visual Experience

Ercenur Ünal & Anna Papafragou

LANGUAGE LEARNING AND DEVELOPMENT
2019, VOL. 15, NO. 2, 138–156
https://doi.org/10.1080/15475441.2018.1544075

Routledge
Taylor & Francis Group

How Children Identify Events from Visual Experience

Ercenur Ünal * and Anna Papafragou

Department of Psychological and Brain Sciences, University of Delaware, Newark, Delaware, USA

ABSTRACT
Three experiments explored how well children recognize events from different types of visual experience: either by directly seeing an event or by indirectly experiencing it from post-event visual evidence. In Experiment 1, 4- and 5- to 6-year-old Turkish-speaking children (*n* = 32) successfully recognized events through either direct or indirect visual access. In Experiment 2, a new group of 4- and 5- to 6-year-olds (*n* = 37) reliably attributed event recognition to others who had direct or indirect visual access to events (even though performance was lower than Experiment 1). In both experiments, although children's accuracy improved with age, there was no difference between the two types of access. Experiment 3 replicated the findings from the youngest participants of Experiments 1 and 2 with a matched sample of English-speaking 4-year-olds (*n* = 37). Thus children can use different kinds of visual experience to support event representations in themselves and others.

Introduction

A central aspect of our day-to-day experiences involves gaining various types of knowledge, such as knowledge about people, objects, and situations. One of the most important pieces of knowledge that allows us to successfully interact with the world around us is knowledge about events. Knowledge about events allows us to predict other people's actions, remember our experiences, and learn new information about the world (Kurby & Zacks, 2008; Papafragou, 2015; Radvansky & Zacks, 2014; Richmond & Zacks, 2017; Shipley & Zacks, 2008; Zacks & Tversky, 2001).

Identifying events in the real world often depends on the perspective of the observer (Gleitman, 1990): a single action can be perceived as chasing or fleeing depending on one's biases or point of view. Furthermore, events—unlike objects—are transient and one's experience of events may be partial and fragmentary: for instance, an event is sometimes visually perceived in its entirety (e.g., Mary sees Susan bite a cookie) but at other times the occurrence of an event can only be identified probabilistically from partial visual evidence from the event aftermath (e.g., Mary sees a cookie with a piece missing on a plate and realizes that Susan took a bite). Yet at other times the occurrence of an event can be identified on the basis of auditory, verbal, or other cues. Thus multiple sources of information can lead to event identification.

Recent evidence shows that adults rapidly compute event representations based on both complete and incomplete visual stimuli. In one study, adults were able to identify a two-participant action such as hitting even after having seen a single frame of the unfolding event for 37 ms (Hafri, Papafragou, & Trueswell, 2013; cf. also Griffin & Bock, 2000; Dobel, Gumnior, Bölte, & Zwitserlood, 2007). In another study, adults were presented with videos of launching and catching events that were missing the moment of contact (Strickland & Keil, 2011). Some participants saw a scene that

CONTACT Ercenur Ünal ercenurunal@gmail.com Department of Psychological and Brain Sciences, University of Delaware, 105 The Green, Room 108, Newark, DE 19716
*Present address: Department of Psychology, Özyeğin University, Nişantepe Mahallesi Orman Sokak Çekmeköy, Istanbul 34794, Turkey.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlld.

strongly implied the moment of contact, whereas other participants saw an irrelevant scene from the same video. In a later memory test, the participants in the first group reported seeing the moment of contact more often than the participants in the second group, suggesting that conceptually coherent post-event information filled in temporally occluded parts of the event and participants perceived the event without the interruption. Other evidence from adults suggests that the co-ordination of different types of input to visual processes in event perception is rapid (cf. Chen & Scholl, 2016; Firestone & Scholl, 2016; Strickland & Scholl, 2015). In one study, participants saw a square that was intruded upon by another shape, such that the contours of the intrusion gave the impression that a piece of the square was missing. Even though this intrusion was sudden, participants reported seeing motion, a fact indicating that they perceived the change as gradual. However, when the contours of the missing piece looked like another shape was superimposed on the square and thus the shape of the second object was not causally related to a change to the first object, participants did not perceive the change as gradual. This suggests that, in some cases at least, a "smart" perceptual system computes backward causal inferences that reconstruct changes of state in an object from the object's shape (Chen & Scholl, 2016; cf. also Landau & Leyton, 1999).

How children identify events from visual experience is the topic of a small but growing literature on event perception (Baillargeon, Li, Gertner, & Wu, 2011; Baldwin, Baird, Saylor, & Clark, 2001; Bunger, Skordos, Trueswell, & Papafragou, 2016; Göksun, Hirsh-Pasek, & Golinkoff, 2010; Kominsky et al., 2017; Radvansky & Zacks, 2014; Spelke, Phillips, & Woodward, 1995; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014; Tatone, Geraci, & Csibra, 2015; Ünal, Trueswell, & Papafragou, 2017) and a larger but separate literature on the acquisition of verbs labeling events of different types (Bowerman & Choi, 2001; Gleitman, 1990; Pinker, 1989; Fisher, Gleitman, & Gleitman, 1991; Tomasello & Merriman, 2014; among many others). However, in this work events are typically visually available in their entirety. As a result, we currently know little about how children use different types of evidence (including indirect or incomplete visual evidence) to recognize events or attribute event representations to others. Understanding events is very central to the daily lives of children. And, clearly children can derive rich interpretations about the events they experience as they often communicate about events with others (Pinker, 1989). Yet the limits of their early abilities remains unknown. The goal of this article is to fill this gap.

Our empirical investigations of this topic use as a starting point the large literature on how children identify *objects* and their properties and attribute object representations to others. This literature has shown that by age 3 children can identify simple objects from a variety of sources: for instance, they can say what object is in a toy tunnel after looking inside, being verbally informed, feeling inside, or figuring out the contents from a clue (Gopnik & Graf, 1988; cf. O'Neill & Gopnik, 1991; Pillow, 1989; Pillow, Hill, Boyce, & Stein, 2000; Wimmer, Hogrefe, & Perner, 1988; Woolley & Bruell, 1996). Furthermore, in simple tasks, even young children seem to grasp that seeing an object can lead to knowing the identity of that object in other people (Pillow, 1989; see also Melis, Call, & Tomasello, 2010; Moll, Carpenter, & Tomasello, 2014; Povinelli & deBlois, 1992; Pratt & Bryant, 1990; Ruffman & Olson, 1989; Sodian, 1988; Wimmer et al., 1988; Wimmer, Hogrefe, & Sodian, 1988). However, the role of indirect visual evidence (especially for what others' minds represent) seems a later development: for instance, 4- and 5-year-olds, unlike 6-year-olds, claim that an observer shown a small, uninformative region of a drawing would know what the drawing depicts (Taylor, 1988; cf. Sodian, Zaitchik, & Carey, 1991). Relatedly, 4-year-olds, unlike 6-year-olds, fail to realize that others can use inference to identify a property of an object even if the object is not visible (Sodian & Wimmer, 1987; cf. Keenan, Ruffman, & Olson, 1994; Pillow, 1999, 2002; Pillow & Anderson, 2006; Pillow et al., 2000; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009). One limitation of this line of work on object knowledge is that it has often assessed the link between a certain experience and the resulting knowledge by asking children questions with explicit mental state language (e.g., "Do you know X?" or "Does < name of character> know X?"; Pillow, 2002; Pillow & Anderson, 2006; Pillow et al., 2000; Pratt & Bryant, 1990; Sodian & Wimmer, 1987; a.o.). However, asking children to explicitly reflect on what they or others know might underestimate their ability to use evidence to form object representations or track such representations in others (see Saxe, 2013, for limitations of explicit methods).

In the present study, we explore 4- and 5- to 6-year-old children's ability to use direct and indirect visual access to recognize events (Experiment 1) and attribute event recognition to others (Experiment 2). We use a novel task that does not involve explicit mental state language but simply asks children to match verbs to events accessed through either direct or indirect visual evidence by themselves or other characters. Our first goal is to sketch the developmental profile of how children compute and attribute event representations (as measured by verb-to-event matching). Two more specific goals are inspired by possibilities raised in the literature on how objects and their properties are identified (see Hogrefe, Wimmer, & Perner, 1986; O'Neill & Gopnik, 1991; Sodian, 1988; Sodian & Wimmer, 1987; Wimmer et al., 1988; Wimmer et al., 1988). First, we ask whether children would be better at event identification (for both themselves and others) when the available visual evidence is direct as opposed to indirect. Second, we explore whether children would be better at identifying events themselves compared to attributing knowledge of event identity to others with identical access to an event. Our goal is to test whether the patterns that characterize children's identification of objects and their properties also characterize children's event identification.

Experiment 1: Identifying events

The goal of Experiment 1 was to assess children's ability to use direct and indirect visual evidence to recognize events. We presented children with two cards. One card had a photograph that provided direct or indirect visual evidence about an event. The second card was facing down. Children were given a verb and were asked to find its picture. Half of the time, the verb matched the visible photograph and half of the time the verb did not match the visible photograph. If the visual evidence provided by the photograph was sufficient for event identification, then children should pick the photograph and not the facing down card. However, the facing down card could potentially provide better evidence for event identification. Thus, if the evidence provided by the photograph was not sufficient for event recognition, children could pick the facing down alternative (cf. Huang, Spelke, & Snedeker, 2013, for a similar method). Furthermore, in order to make sure that children were evaluating the evidence on the photograph (and did not have a response bias) there were some trials when the photograph was not the correct response (as indicated by a mismatch between the verb and the visual evidence in the photograph) and the facing down card was a correct response. Of interest was whether children would be able to link the evidence provided by the photograph to the event described by the verb, and whether performance would differ depending on the type of visual evidence (direct vs. indirect).

Method

The methods reported in this experiment and all subsequent experiments were approved by the Institutional Review Board of the University of Delaware.

Participants

Data were collected from Turkish children in two age groups: 4-year olds (n = 16, 9 females, mean age 4;6, range 4;0–4;11), and 5- to 6-year-olds (n = 16, 8 females, mean age 6;0, range 5;6–6;10). Children were recruited through preschools in Istanbul, Turkey that served middle- to upper-class families. All children were native speakers of Turkish. None of the children had a history of speech-language or any other developmental disorders. Eleven Turkish adults also participated as controls (9 females, mean age 19.3, range 18–21). All were native speakers of Turkish. Adults were students at Koç University in Istanbul, Turkey and participated in the experiment to satisfy a course requirement. Adults performed at ceiling level for all types of access ($M_{dir\text{-}vis}$ = 1.00, $M_{ind\text{-}vis\text{-}familiar}$ = 0.98, $M_{ind\text{-}vis\text{-}unfamiliar}$ = 0.98) and thus adult results will not be reported.

Materials

Photographs giving different types of access to events were used as stimuli.[1] There were three types of access: Direct Visual Access, Indirect Visual Access-Familiar, and Indirect Visual Access-Unfamiliar. For Direct Visual Access cases, the stimuli consisted of photographs of ongoing events including an animate agent (e.g., a man opening a jar). For Indirect Visual Access cases, we manipulated the familiarity of objects used as visual cues. We reasoned that familiar visual cues (e.g., footsteps on snow) might be more helpful in reconstructing what happened compared to other, perhaps less familiar visual cues (e.g., paint footsteps on asphalt).[2] Sample events for each type of access are presented in Figure 1 and a full list is included in the Appendix.

There were a total of 18 events, which consisted of 6 examples of each of the three types of access. Each Indirect Visual Access event had two versions (Familiar and Unfamiliar; cf. Figure 1). In order to counterbalance the familiarity of the objects for Indirect Visual Access events, two lists of events were created. The two versions of each Indirect Visual Access event were never assigned

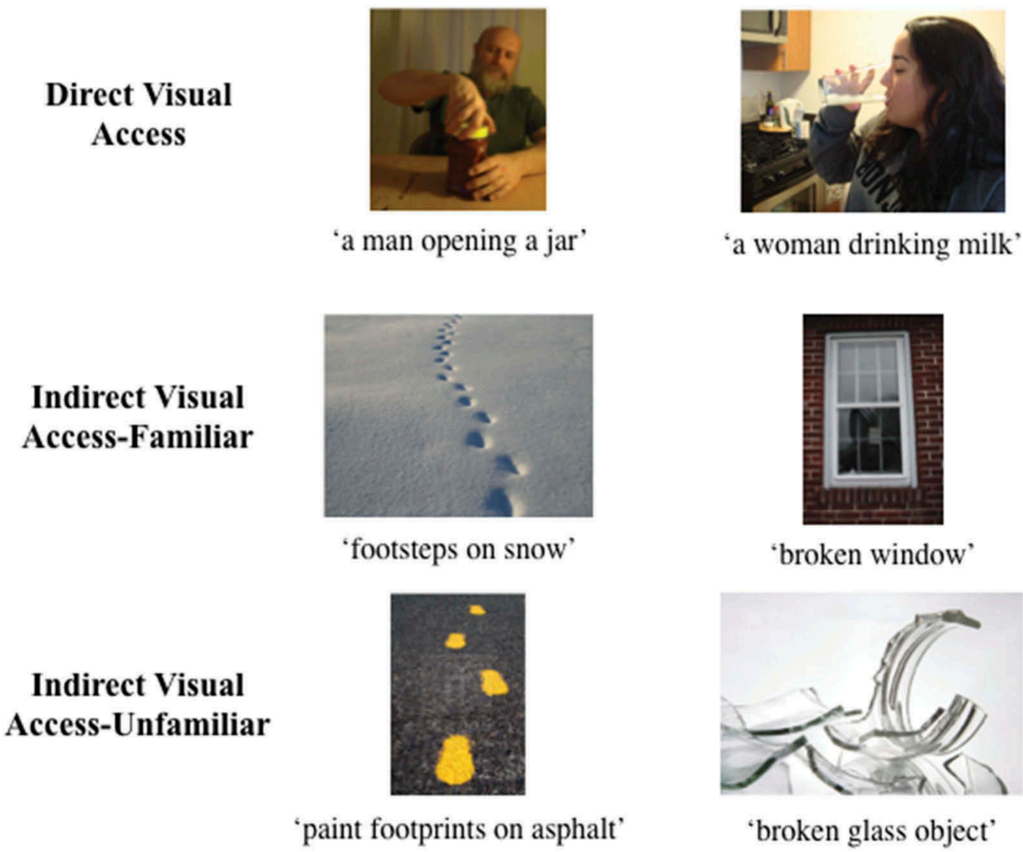

Figure 1. Sample stimuli for the three types of access in Experiment 1.

[1] Our choice for using static pictures as stimuli is motivated by two reasons. First, our aim was to test children's inferences about events on the basis of a single snapshot of the event. Second, prior developmental work has revealed that static depictions of events can elicit rich interpretations about events both in language production tasks and non-linguistic tasks from children of similar ages (Bunger et al., 2016; Göksun et al., 2010; Nappa, Wessell, McEldoon, Gleitman, & Trueswell, 2009; Ünal et al., 2017).

[2] For the selection of stimuli for the indirect visual access events, a list of events that would be depicted by an indirect visual cue was created. The familiar visual access items were constructed by using the most typical object on which this action would be performed (e.g., for "cracking," an "egg" was picked). For unfamiliar visual access items, these objects were replaced by another object that would not be easily identified and/or associated with the action by children (e.g., a walnut, little pieces of wood, etc.)

to a single list. Direct Visual Access events were the same for the two lists. The same random order of events was used for each list.

To select the verbs we would use to describe these events, and to ensure that type of evidence was reasonable, we asked a separate group of Turkish adults (n = 14) to describe each event with one verb and rate how certain they are that this is what has happened in the event on a 5-point scale (1-not certain at all, 5-very certain). Overall, in 80% of the trials adults produced the very same matching verbs that were used in the study ($M_{dir\text{-}vis}$ = 0.94, $M_{ind\text{-}vis\text{-}familiar}$ = 0.79, $M_{ind\text{-}vis\text{-}unfamiliar}$ = 0.66). Failures to use the matching verb to describe the event were mostly due to the tendency to use more general as opposed to more specific verbs. None of the events were described by the mismatching verbs that were assigned to the particular photographs. When adults used the very same matching verbs to identify the events, they were certain that this was indeed what has happened in the event, as indicated by mean certainty ratings ($M_{dir\text{-}vis}$ = 3.92, $M_{ind\text{-}vis\text{-}familiar}$ = 4.04, $M_{ind\text{-}vis\text{-}unfamiliar}$ = 3.97). These means did not differ from each other.

As a final check, we consulted the Turkish Communicative Development Inventory (TIGE, the Turkish adaptation of the MacArthur-Bates Communicative Development Inventory) and found that 89% of the mothers of a nationally representative sample of Turkish 3-year-olds reported that their children understood and produced the verbs used to describe the events in our study (Aksu-Koç et al., 2011).

Procedure

Children were tested individually in a quiet room at their preschool. The experimenter presented children with two facing down cards and told them that there was a picture under each card but they could look at only one of these pictures and the other card had to be upside down. Then the experimenter turned over one of the cards revealing a photograph that gave some type of access to an event (Direct Visual Access, Indirect Visual Access-Familiar, or Indirect Visual Access-Unfamiliar). Then the experimenter said (in Turkish): "Now, I will tell you a word. I want you to find its picture for me. Do not forget that there is a picture under this card (pointing to the facing down card) so, the picture for the word can be under this card as well." Then the experimenter uttered a verb in the infinitive form (to "V") that can be felicitously used on its own. We chose this form because it is an "unmarked" form of the verb that allowed us to refer to the events without encoding tense, aspect, or source of information. Furthermore, this form is already produced by 3-year-old Turkish learners according to 98.7% of the mothers surveyed in TIGE (Aksu-Koç et al., 2011). After the children made a choice, the experimenter went onto the next trial.

In half of the trials, the experimenter uttered a verb that matched the event depicted in the visible photograph (e.g., the photograph showed some footsteps and the experimenter said, "walk," Turkish "yürümek"). In the other half of the trials, the experimenter uttered a verb that did not match the event depicted in the visible photograph (e.g., the photograph showed some footsteps and the experimenter said "stack," Turkish "dizmek"). For each type of evidence, there were three matching verb and three mismatching verb trials. The assignment of verb types to events was counterbalanced across participants. For example, for the photograph showing the footsteps, half of the children heard "walk" and the other half heard "stack."

At the beginning of the experiment there were three practice trials to familiarize the children with the task. For these trials, the visible photographs consisted of photographs of objects. The experimenter followed the same instructions as in the actual experiment. After the children responded in each practice trial, the experimenter gave feedback about their response by turning up the facing down card and showing the photograph under the card. Children were highly accurate (M = 0.96) in these practice trials.

If the children thought that the visible photograph provided sufficient evidence about the event encoded by the verb, then children should pick the visible photograph. Otherwise, they should pick the facing down alternative. Of interest was whether children's likelihood of linking matching verbs (and avoiding linking mismatching verbs) to visible photographs would differ across types of access and age groups. Specifically, we were interested in whether children's success in linking verbs to events would differ across direct and indirect visual access. We were also interested in whether events would be best identified when the indirect visual cues giving access to an event were familiar as opposed to when they were unfamiliar.

Results

Data from all experiments were analyzed using generalized binomial linear mixed effects modeling (*glmer*) with crossed random intercepts for Subjects and Items. All models were fit with the *lme4* package (version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015) in *R* (R Core Team,). Table 1[3] presents the proportion of selecting the visible photograph across Verb Types, split by Type of Access for each age group. The fixed effects that were investigated were Access (Direct Visual Access, Indirect Visual Access-Familiar, Indirect Visual Access-Unfamiliar), Verb Type (Matching, Mismatching) and Age (4-year-olds, 5- to 6-year-olds). The fixed effect of Access was assessed with two planned comparisons using contrast coding (c_1 Direct vs. Indirect Access: 0.66, −0.33, −0.33; c_2 Indirect-Familiar vs. Indirect-Unfamiliar Access: 0, −0.5, 0.5). The fixed effects of Verb Type and Age were analyzed with centered contrasts (−0.5, 0.5). Dependent variable was binary values for picking the visible photograph (1 = photo, 0 = facing down card) at the item level. Fixed factors that varied within subjects were included as random slopes in Subjects and Items when they were not perfectly correlated with the Intercept and when they improved model fit as indicated by a significantly lower value for the Aikake Information Criterion (AIC) based on a chi-square test of the change in −2 restricted log likelihood between the two models. For completeness, we report the estimates from the most inclusive model.[4]

Table 2 presents the fixed effect estimates from the *glmer* model for picking the visible photograph. The model revealed a main effect of Verb Type. As expected, children picked the visible photograph when presented with a Matching Verb and were significantly less likely to pick the visible photograph and instead picked the facing down card when presented with a Mismatching Verb. That is, averaged across types of access and age groups, when presented with a Matching Verb, children frequently selected the visible photograph (86% of the time). However, when presented with a Mismatching Verb, children selected the visible photograph only 14% of the time and instead they frequently selected the facing down card (86% the time).

The model also revealed a significant interaction between Verb Type and Age, indicating that the difference in the likelihood of picking the visible photograph across Matching and Mismatching Verbs was larger for 5- to 6-year-olds (0.91 vs. 0.11) than it was for 4-year-olds (0.81 vs. 0.17). This shows that the older group of children was simply more successful in linking verbs to events

Table 1. Proportion of picking the visible photograph (Experiment 1).

	4-year-olds		5- to 6-year-olds	
	Matching Verb	Mismatching Verb	Matching Verb	Mismatching Verb
Direct	0.79	0.06	0.90	0.06
Indirect-Familiar	0.83	0.25	0.92	0.15
Indirect-Unfamiliar	0.81	0.21	0.92	0.13

[3]Due to the binary nature of the data, Tables 1 and 3 only report the proportion of selecting one of the options, the visible photograph, split across Verb Types, Types of Access, and Age groups. The proportion of selecting the other option, the facing down card, can be deduced by subtracting these proportions from 1. Since the data that contributes to Matching and Mismatching Verb conditions come from different items, the proportions of selecting the visible photographs for Matching and Mismatching Verb trials do not add up to 1.

[4]A different analytical strategy in which we started with the most inclusive model and gradually removed the non-significant effects and interactions to reach to the most parsimonious model also revealed the same findings in Experiment 1 and all subsequent experiments.

Table 2. Fixed effect estimates from the *glmer* model for picking the visible photograph in Experiment 1.

Fixed Effect	*β*	*SE*	*z*	*p-value*
(Intercept)	0.114	0.284	0.402	0.687
Access (Direct vs. Indirect)	0.568	0.502	1.131	0.258
Access (Indirect Familiar vs. Unfamiliar)	−0.201	0.327	−0.615	0.539
Age (4 vs. 5–6)	0.284	0.423	0.673	0.501
Verb Type (Match vs. Mismatch)	−4.918	0.553	−8.893	0.000***
Access (Direct vs. Indirect): Age (4 vs. 5/6)	−0.484	0.660	−0.733	0.464
Access (Indirect Familiar vs. Unfamiliar): Age (4 vs. 5/6)	0.147	0.652	0.226	0.821
Access (Direct vs. Indirect): Verb (Match vs. Mismatch.)	1.458	1.057	1.379	0.168
Access (Indirect Familiar vs. Unfamiliar): Verb (Match vs. Mismatch.)	−0.163	0.652	−0.251	0.802
Age (4 vs. 5/6):Verb (Match vs. Mismatch.)	−1.576	0.608	−2.593	0.010**
Evidence (Direct vs. Indirect):Age (4 vs. 5/6):Verb (Match vs. Mismatch.)	−0.590	1.318	−0.448	0.655
Evidence (Indirect Familiar vs. Unfamiliar):Age (4 vs. 5/6):Verb (Match vs. Mismatch.)	−0.036	1.301	−0.028	0.978

Significance codes: *** 0.001, ** 0.01, * 0.05
Formula in *R*: Response ~ Access * Age * Verb + (1 | Subject) + (1 + Verb | Event).

compared to the younger group of children. No other effects or interactions were significant, indicating that children's success in event identification did not vary across different types of visual access.

Discussion

In this experiment we tested whether 4- and 5–6-year-old children can identify events based on different types of evidence. We used a novel task without any explicit mental state language (e.g., "Do you know…?"): in this task, children had to link a verb to either a face-up picture (providing direct or indirect visual access to an event) or an upside-down picture. Both age groups were highly successful in linking verbs to corresponding events on the basis of the evidence available from the accessible picture, even though 5- to 6-year-olds were better than 4-year-olds. There was no difference in event identification between direct and indirect access. Similarly, within the indirect class, there was no difference in event identification between familiar and unfamiliar cues.

This is some of the first evidence that children as young as 4 years can use not only direct but also indirect perceptual evidence to correctly identify events. Even though prior work has shown that young children can use direct perception and simple indirect perceptual cues to identify objects (Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; Pillow, 1989; Pillow et al., 2000), there is independent evidence on object knowledge showing that 4-year-olds sometimes have difficulty understanding the subtle distinctions between different types of indirect cues (e.g., Sodian et al., 1991; Taylor, 1988; de Villiers et al., 2009). Given that event knowledge is more complex than object knowledge, the fact that the 4-year-olds in our study can use indirect visual cues for event identification is especially novel.

Experiment 2: Event identification in others

The goal of Experiment 2 was to examine whether children understand that others can identify events under conditions in which they have direct or indirect visual access to the events. We used the paradigm of Experiment 1 but replaced the child's access to an event with someone else's (a puppet's) access. Furthermore, since there was no difference between the two types of Indirect Visual Access in Experiment 1, we only used the stimuli for Direct and Indirect-Familiar events in Experiment 2.

Method

Participants

Data were collected from a new group of Turkish children in two age groups: 4-year olds (n = 21, 9 females, mean age 4;5, range 4;0–4;10) and 5- to 6-year-olds (n = 16, 8 females, mean age 6;0, range 5;6–6;10). Children were recruited through preschools in Istanbul, Turkey that serve middle- to upper-class families. All children were native speakers of Turkish. None of the children had a parent reported history of speech-language or other developmental disorders. Seven Turkish adults also participated as controls (5 females, mean age 26, range 25–27). All of them were native speakers of Turkish. All adult participants were college graduates and volunteered to participate in the study. As in Experiment 1, adult performance was at ceiling ($M_{dir\text{-}vis}$ = 0.98, $M_{ind\text{-}vis}$ = 0.94) and will not be reported.

Materials

The stimuli consisted of a subset of the stimuli for Experiment 1. Since there was no effect of familiarity of objects as visual cues for the inference events, we excluded the unfamiliar events. Thus, the visual stimuli for Experiment 2 consisted of all of the Direct Visual Access and Indirect Visual Access-Familiar (henceforth referred to as Indirect Visual Access) events of Experiment 1. There were 6 events per type of evidence, resulting in a total of 12 events.

Procedure

Children were tested individually in a quiet room at their preschool. There were two experimenters: the first experimenter (E1) always interacted with the children and asked the questions, and the second experimenter (E2) only acted out the puppets. E1 presented the children with two cards that were facing down and said (in Turkish): “Look, I have two cards here. There is a picture under each card, but we can look at only one of them. The other has to be upside down.” When E1 turned one of the cards up, E2 placed each of the puppets next to each card. E1 went on to say, “But I have a puppet and it can look at the picture under this card.” Then one of the puppets looked at the visible photograph (the child could also see this card), and the other puppet looked under the facing down card (the child had no access to what was under the facing down card). The second puppet only looked under the facing down card and not at the visible photograph. After that, each puppet put the card they looked at into two different boxes. The two boxes were next to the cards. E1 asked the children: “If you want to find out more about < verb>, which puppet should you ask?” As in Experiment 1, the verb was presented in the infinitive form (“to V”). After the children made a choice the experimenter went onto the next trial.

As in Experiment 1, in half of the trials (three per type of evidence, six in total), E1 asked about an event (i.e., verb) that matched the visible photograph (e.g., the photographs showed some footsteps and E1 asked: “If you want to find out more about ‘walk’, which puppet should you ask?”). In the other half of the trials E1 asked about an event (i.e., verb) that did not match the visible photograph (e.g., the photographs showed some footsteps and E1 asked, “If you want to find out more about ‘stack’, which puppet should you ask?”). (Note that the sentences were both grammatical and natural in Turkish.) The same counterbalancing procedure for verb types as in Experiment 1 was followed.

There were three practice trials at the beginning of the experiment. The visible photographs consisted of the same three photographs of objects as the practice trials of Experiment 1. E1 followed the same instructions as in the main experiment. After the children responded in each practice trial, E1turned up the facing down card and showed the photograph to give them feedback about their response. Children made errors and received feedback in 22% of the practice trials.

Of interest was whether children would pick the puppet looking at the accessible event when they were asked about a verb that matched the accessible event and avoid picking the puppet looking at the accessible event and instead pick the puppet looking at the inaccessible alternative when the verb did not match the accessible event. Furthermore, we were interested in whether children’s performance in attributing event representations to others would differ from their performance in forming event representations for themselves in Experiment 1. Finally, we wanted to test whether, unlike Experiment 1, there might be a source asymmetry, with indirect visual access leading to more errors compared to direct visual access.

Results

Table 3 presents the proportion of picking the puppet that accessed the visible photograph across Matching and Mismatching Verbs split by type of Access for each age group. The same data analysis and model fitting procedures were used as in Experiment 1. The fixed effects that were investigated were Access (Direct Access, Indirect Access), Verb Type (Matching, Mismatching) and Age (4-year-olds, 5- to 6-year-olds). All fixed effects were assessed using centered contrasts (−0.5, 0.5). The dependent variable was binary values for picking the puppet that accessed the visible photograph (1 = photo, 0 = facing down card) at the item level. Subjects and Items were added as random intercepts. Adding random slopes for factors that varied within subjects did not significantly improve model fit, thus they were not included. As in Experiment 1, we report the estimates from the most inclusive model.

Table 4 presents the fixed effect estimates from the *glmer* on picking the puppet that accessed the visible photograph. As expected, the model revealed a main effect of Verb Type: children in both age groups were more likely to pick the visible photograph when presented with a Matching Verb (0.62) compared to when presented with a Mismatching Verb (0.38), suggesting that they were able to link different types of evidence to events. Furthermore, there was a trend indicating that the difference between difference in picking the visible photograph across Matching and Mismatching Verbs (0.68 vs. 0.34) was larger for 5- to 6-year-olds than it was for 4-year-olds (0.57 vs. 0.40). Nevertheless, this trend, as indicated by an Age by Verb Type interaction, did not reach statistical significance (p = .08). No other effects or interactions were significant.

Self vs. others in event representation

We compared the data from the Self Task (Experiment 1) to the data from the Others Task (Experiment 2) with a generalized binomial linear mixed effects modelling (*glmer*) with crossed random intercepts for Subjects and Items. The fixed factors that we assessed were Verb Type (Matching, Mismatching), Age (4-year-olds, 5- to 6-year-olds), and Perspective (Self, Other). (We

Table 3. Proportion of picking the visible photograph (Experiment 2).

	4-year-olds		5- to 6-year-olds	
	Matching Verb	Mismatching Verb	Matching Verb	Mismatching Verb
Direct Visual Access	0.54	0.41	0.69	0.33
Indirect Visual Access	0.60	0.39	0.67	0.35

Table 4. Fixed effect estimates from the *glmer* model on picking the visible photograph in Experiment 2.

Fixed Effect	*β*	*SE*	*z*	*p-value*
(Intercept)	−0.020	0.237	−0.084	0.933
Access	0.024	0.371	0.065	0.948
Verb Type	−1.262	0.227	−5.554	0.000***
Age	0.150	0.368	0.409	0.683
Access: Verb Type	0.055	0.444	0.124	0.902
Access: Age	−0.097	0.437	−0.223	0.824
Verb Type: Age	−0.768	0.441	−1.743	0.081
Access: Verb Type: Age	0.644	0.875	0.735	0.462

Significance codes: *** 0.001, ** 0.01, * 0.05
Formula in *R*: Response ~ Access * Verb * Age + (1 | Subject) + (1 | Event)

collapsed the data from Familiar and Unfamiliar trials to get a single score for Indirect Visual Access trials in Experiment 1.) All fixed effects were assessed with centered contrasts (−0.5, 0.5). Table 5 presents the estimates from the model.

The model revealed a main effect of Verb Type. Across tasks and age groups, children picked the visible photo when presented with a Matching Verb (0.74) and avoided picking the visible photo (0.24) and instead picked the facing down card when presented with a Mismatching Verb. Furthermore, the difference in the likelihood of picking the visible photo across Matching and Mismatching Verb trials was larger in the Self Task (0.86 vs. 0.11) than it was in the Others task (0.62 vs. 0.37), as indicated by a significant interaction between Verb Type and Perspective. This shows that performance was better in the Self task than in the Others task. Finally, there was an interaction between Verb Type and Age: the difference in the likelihood of picking the visible photo across Matching and Mismatching Verbs was larger for 5- to 6-year-olds (0.79 vs. 0.21) than it was 4-year-olds (0.69 vs. 0.28). This indicated that, across tasks older children performed better than younger children. No other effects or interactions were significant. We conclude that thinking about how others identify events is more difficult than identifying events for oneself.

Discussion

Experiment 2 examined children's ability to attribute event representations to characters who had either direct or indirect visual access to an event. A major finding was that both 4- and 5- to 6-year-olds recognized whether others represented an event. Prior literature had shown that 3-year-olds understand simple links between seeing an object and identifying an object: for instance, they know that someone who looked inside a box knows what is in the box unlike someone who has simply tapped the box (e.g., Pillow, 1989; Pratt & Bryant, 1990; Woolley & Bruell, 1996). These studies required considerable training and linguistic support (e.g., Pillow, 1989, children were told that the person "looked and therefore he knows"). In this article, others' access to an event was not highlighted or commented upon in any special way; yet even 4-year-olds could reliably link someone's direct or indirect visual access to an event with the ability to identify the event (and thus become a good candidate for supplying further information about the event when asked).

Two further aspects of our findings are noteworthy. First, as in Experiment 1, there was no difference in children's use of direct or indirect visual access: children were equally likely to identify someone who saw the point at which an event took place and someone who filled in incomplete visual information as knowledgeable about the event. Second, despite their good performance in the present task, children were better at recognizing events from different types of visual access for themselves (Experiment 1) compared to attributing event recognition to others with identical information sources (Experiment 2). We expand on the significance of these aspects of our findings in the General Discussion.

Table 5. Fixed effect estimates from *glmer* model on picking the visible photograph across Experiments 1 and 2.

Fixed Effect	*β*	*SE*	*z*	*p-value*
(Intercept)	−0.076	0.182	−0.420	0.674
Verb Type	−2.911	0.236	−12.345	0.000***
Perspective	0.108	0.298	0.362	0.717
Age	0.074	0.298	0.248	0.804
Verb Type: Perspective	3.368	0.446	7.552	0.000***
Verb Type: Age	−1.293	0.429	−3.011	0.003**
Perspective: Age	0.142	0.595	0.238	0.812
Verb Type: Perspective: Age	1.089	0.858	1.269	0.204

Significance codes: *** 0.001, ** 0.01, * 0.05
Formula in *R*: Response ~ Verb * Perspective * Age + (1 | Subject) + (1 | Event)

Experiment 3: A cross-linguistic replication

Our results leave open the possibility that children's success with identifying events in Experiment 1 and attributing event representations to others in Experiment 2, as well as the relative difficulty of Experiment 2 compared to Experiment 1, might be linked to the fact that the language of the experiments was Turkish. Two features of Turkish are especially relevant.

A first feature has to do with the Turkish form of the question used to probe others' event representations. Because Turkish has a flexible word order, the order in which information becomes available in an utterance may vary (Erguvanlı, 1984). To maintain naturalness, in the Self task, the target verb in the infinitive form appeared at the end of the test question ("Hangisinde < verb>?", literally "Which one has < verb>?") but in the Others task the target verb appeared at the beginning of the question ("< Verb> hakkında daha çok şey öğrenmek için kime sormalıyız?", literally "< Verb> about more find out who do we ask?"). In order to ensure that the self-other asymmetry was not driven by such order differences, in Experiment 3 we replicated both experiments with a sample of English-speaking 4-year-olds (since no such verb order differences exist in English).

Second, Turkish is a language with grammaticalized evidentiality and obligatorily encodes the distinction between direct and indirect sources of information through verb morphology. One might hypothesize that Turkish-speaking children in our sample might have special sensitivity to the distinction between direct perception and indirect visual cues as a consequence of acquiring Turkish. Prior cross-linguistic comparisons provide preliminary evidence against such advantages for both child and adult speakers of languages with grammaticalized evidentiality (Papafragou, Li, Choi, & Han, 2007; Ünal, Pinto, Bunger, & Papafragou, 2016). Nevertheless, in Experiment 3 we tested this assumption more directly by testing English preschoolers with the current paradigm and stimuli.

Method

Participants

Thirty-seven English-speaking 4-year-olds were recruited from daycares in Newark (Delaware) in the US. Each child was age-matched with one of the original Turkish participants from Experiments 1 and 2, such that 16 English speakers (M_{age} = 4;6, range: 4;0–5;0) were tested on the Self task and 21 (M_{age} = 4;5, range: 4;1–4;11) on the Others task. English speaking children were recruited from preschools that served middle- to upper-class families. Fourteen English-speaking adults also participated as controls (7 females, mean age 18.2, range 18–22). All of them were native speakers of English. All adult participants were students at University of Delaware and received course credit for their participation in the study. As in Experiments 1 and 2, adult performance was at ceiling (Self task: $M_{dir\text{-}vis} = 1.00$, $M_{ind\text{-}vis\text{-}familiar} = 0.92$, $M_{ind\text{-}vis\text{-}unfamiliar} = 0.94$; Others task: $M_{dir\text{-}vis} = 1.00$, $M_{ind\text{-}vis} = 0.96$) and will not be reported.

Materials and procedure

Stimuli and procedure was exactly the same as in Experiments 1 and 2 with the following exceptions. First, the verbs used in the Self and Others tasks were translated to English and the infinitival form of the verb was replaced by the *–ing* form. Second, the target verb appeared at the end of the test question in both the Self task ("Which one has < verb>?") and the Others task ("Who do we ask to find out more about < verb>?").

Results

Table 6 presents the proportion of picking the visible photograph across Verb Types split by type of Access in the Self and Others tasks for English speakers. Data were analyzed using the same data analysis, model fitting, and effect coding procedures as in Experiments 1 and 2.

Data from the Self task were analyzed using a *glmer* model with crossed random intercepts for Subjects and Items that tested the effects of Access (Direct Visual Access, Indirect Visual Access-Familiar, Indirect Visual Access-Unfamiliar), and Verb Type (Matching, Mismatching). The dependent variable was binary

Table 6. Proportion of picking the visible photograph across self and others tasks (Experiment 3).

	Self Task			Others Task	
	Direct	Indirect-Familiar	Indirect-Unfamiliar	Direct	Indirect
Matching Verb	0.80	0.70	0.73	0.66	0.66
Mismatching Verb	0.29	0.33	0.36	0.37	0.42

values picking the visible photograph (1 = photo, 0 = facing down card) at the item level. The model only revealed a significant effect of Verb Type ($\beta = -4.643$, $SE = 0.612$, $z = -7.589$, $p < .001$): children picked the visible photograph when presented with a Matching Verb (0.75) and avoided picking the visible photograph (0.33) and instead picked the facing down card when asked about a Mismatching Verb.

Data from the Others task were analyzed using a *glmer* model on binary values picking the puppet that had access to the visible photograph (1 = photo, 0 = facing down card) at the item level with crossed random intercepts for Subject and Items. The fixed effects of Access (Direct Visual Access, Indirect Visual Access) and Verb Type (Matching, Mismatching) were tested. The model reveled only a main effect of Verb Type ($\beta = -1.322$, $SE = 0.283$, $z = -4.666$, $p < .001$): children picked the puppet that had access to the visible photograph when asked about a Matching Verb (0.66) and avoided picking the puppet that had access to the visible photograph when asked about a Mismatching Verb (0.40) and instead picked the puppet that accessed the facing down card.

As in Experiments 1 and 2, English-speaking children's performance in the Self and Others tasks was compared using a *glmer* model with crossed random intercepts for Subjects and Items. The fixed effects of Verb Type (Matching, Mismatching) and Perspective (Self, Other) were tested. The dependent variable was binary values picking the visible photograph (1 = photo, 0 = facing down card) at the item level. The model revealed a significant effect of Verb Type ($\beta = -2.819$, $SE = 0.348$, $z = -8.090$, $p < .001$) as well as a significant interaction between Verb Type and Perspective ($\beta = 2.735$, $SE = 0.662$, $z = 4.129$, $p < .001$). This indicated that the difference in the likelihood of picking the visible photograph across the Matching Verb and Mismatching trials was larger in the Self Task (0.85 vs. 0.22) than in the Others task (0.69 vs. 0.40). In other words, as in the Turkish sample, English-speaking 4-year-olds performed better in the Self task compared to the Others task.

As a final analysis, we compared performance of English-speaking 4-year-olds to the performance of age-matched Turkish peers. We used a *glmer* model with crossed random intercepts for Subjects and Items. The fixed factors that we assessed were Access (Direct Visual Access, Indirect Visual Access), Verb Type (Matching, Mismatching), Perspective (Self, Other), and Language (English, Turkish). All fixed effects were assessed with centered contrasts (−0.5, 0.5). Adding random slopes for factors that varied within subjects did not improve model fit, and thus were omitted from the final model. Table 7 presents the fixed effect estimates from the model. The model revealed a main effect of Verb Type and an interaction between Verb Type and Perspective. As expected, children picked the visible photograph when asked about a Matching Verb (0.72) and did not pick the visible photo and instead picked the facing down alternative when asked about a Mismatching Verb (0.31). Furthermore, for both English and Turkish children, the difference in the likelihood of selecting the visible photograph in the Matching vs. Mismatching Verb trials was larger in the Self task (0.83 vs. 0.19) than in the Others task (0.63 vs. 0.40). There were no effects or interactions involving the fixed factor Language, indicating that English and Turkish speaking children performed similarly across tasks.

Discussion

Experiment 3 replicated the findings of Experiments 1 and 2 with an age-matched group of English-speaking preschoolers. As in Experiments 1 and 2, English-speaking 4-year-olds successfully recognized events from both direct and indirect visual cues and attributed event representations to others, although performance was worse in the Others task than in the Self task. In both tasks, English- and Turkish-speaking children performed similarly. We conclude that the observed self-other asymmetry

Table 7. Fixed effect estimates from the *glmer* model on picking the visible photograph across Experiments 1–3.

Fixed Effect	*β*	*SE*	*z*	*p-value*
(Intercept)	0.040	0.165	0.243	0.808
Access	0.245	0.193	1.267	0.205
Verb Type	−2.580	0.226	−11.401	0.000***
Perspective	0.030	0.327	0.092	0.927
Language	−0.426	0.327	−1.303	0.193
Access: Verb Type	0.558	0.387	1.444	0.149
Access: Perspective	−0.366	0.378	−0.969	0.332
Verb Type: Perspective	2.856	0.433	6.599	0.000***
Access: Language	0.643	0.378	1.701	0.089
Verb Type: Language	0.338	0.416	0.813	0.416
Perspective: Language	0.203	0.654	0.310	0.757
Access: Verb Type: Perspective	−1.337	0.774	−1.729	0.084
Access: Verb Type: Language	0.122	0.772	0.158	0.875
Access: Perspective: Language	−1.206	0.756	−1.596	0.111
Verb Type: Perspective: Language	0.471	0.830	0.567	0.570
Access: Verb Type: Perspective: Language	−1.599	1.544	−1.036	0.300

Significance codes: *** 0.001, ** 0.01, * 0.05
Formula in *R*: Response ~ Access * Verb * Perspective * Language + (1 | Subject) + (1 | Event)

is not linked to surface features of the test probes in Turkish. Furthermore, children's ability to both represent events and attribute event representations to others seems to proceed similarly across different language-learning communities.

General discussion

Events in the world are often experienced in a partial, interrupted or disjointed fashion, and event apprehension and recognition involve a complex co-ordination of cognitive processes. In adults, these processes are seamless and rapid (e.g., Strickland & Keil, 2011). In children, however, the mechanisms underlying event identification are only beginning to be understood (see Bunger et al., 2016; Radvansky & Zacks, 2014). The present experiments were designed to contribute to the developmental literature on event cognition by focusing on children's ability to identify events on the basis of different types of visual input.

Specifically, we examined children's use of direct and indirect types of visual experience as sources of event representation. Unlike past work on sources of object representation that has often asked children for explicit reports on their own and others' mental states (Pillow, 2002; Pillow & Anderson, 2006; Pillow et al., 2000; Sodian & Wimmer, 1987), we devised a task that simply asked children to match verbs to events that had been accessed either through direct visual perception or indirect visual cues by the children themselves or by other characters. Our first goal was to sketch the developmental timetable of how children achieve and attribute event recognition (as measured by verb-to-event matching); we hypothesized that our method might reveal early sensitivity to the way visual input is handled, both for one's own and for others' event representations. Two more specific goals were to compare the contributions of direct and indirect visual access to event identification and to assess the relation between performing event identification for oneself vs. attributing event-identifying knowledge to others.

Our findings revealed that 4- and 5- to 6-year-olds were highly successful in forming event representations both from direct perception of the unfolding event itself and from indirect visual cues that pointed to the occurrence of the event (Experiment 1). Children of similar ages could also reliably attribute event identification to others who had the same kind of access to this set of events (Experiment 2), even though performance was lower compared to the first ("self") task. The ability to handle event identification (both in oneself and in others) increased with age but did not differ depending on how event information was acquired (through direct visual perception vs. indirect visual cues).

One possibility was that children's successful event recognition might be explained by the fact that they were learners of Turkish and might have been especially sensitive to the different types of access to events. However, Experiment 3 showed that an age-matched group of English preschoolers performed similarly to the Turkish-speaking children, suggesting that this possibility is unlikely. These findings are also consistent with cross-linguistic comparisons with children (Papafragou et al., 2007) and adults (Ünal et al., 2016) showing that the ability to handle different sources of information for events proceeds similarly across speakers of different languages. Furthermore, prior work has shown that acquisition of evidentiality lags behind reasoning about different types of evidence in several respects both in learners of Turkish (Ozturk & Papafragou, 2016) and in learners of other languages with grammaticalized evidentiality, such as Korean (Papafragou et al., 2007; see Ünal & Papafragou, 2018; for an overview).

A particularly interesting aspect of our findings is that children's use of direct visual access and indirect visual cues to assess event identification in others lagged behind their ability to use the same types of access to identify events themselves (cf. also Hogrefe et al., 1986; Sodian, 1988; Sodian & Wimmer, 1987; Ünal & Papafragou, 2016; Wimmer et al., 1988). This self-other asymmetry emerged even though children were not asked to explicitly report on their own or others' knowledge or access to events. We argue that the asymmetry stems from the development of perspective-taking abilities needed to compute others' informational sources and resulting representations. In support of this idea, the difficulty in attributing event representations to others seems to characterize children from different communities: English-speaking 4-year-olds responded identically to Turkish-speaking 4-year-olds when tested with the current methods. Our data show that there is considerable development in the ability to link access and event representations in others between the ages of 4 and 5–6; not even the older children in our sample are at ceiling in linking access to an event and recognition of the event in others.

One might wonder whether the Others task of Experiment 2 requires children to take the puppet's perspective and whether children might perform well in this task by merely associating a puppet with a card and retaining these associations instead of assessing puppet's knowledge of the events. The fact that even the older group of children did not perform at ceiling level suggests that they needed to do more than simply retaining the associations between the puppets and the cards. Perhaps, the drop in children's performance in the Others task might simply be attributed to the fact that children had to retain more information in memory in this task. Two pieces of evidence provide support for the idea that this alternative explanation is unlikely. First, an earlier study that used a similar paradigm also found a self-other asymmetry (Ünal & Papafragou, 2016). In that study, the Self and Others tasks had the same set up and visual stimuli (thus the information retained in memory was the same) and the only difference between the tasks was whether the test question required children to reason about their own vs. someone else's evidence for information (i.e., the perspective-taking demands). Second, we conducted a pilot study with six 4-year-olds that that used the paradigm of Experiment 2 but replaced the test question with "Which box has < verb>?," so that children would have to identify the events themselves. Four-year-olds were highly accurate in this task and their performance was similar to the performance in the Self task of Experiment 1. This suggests that the drop in children's performance in Experiment 2 is best explained by the perspective-taking demands of attributing event representations to others.

Our findings have implications for theories of children's event representation, since they reveal that children as young as 4 rely on different types of visual access as they form their own and others' representation of events in the world. It remains an open question whether children's early success with event identification from visual input generalizes to other types of events or situations that involve identification of multiple events. Recent work by Zacks (2017) has shown that children between the ages of 5 and 7 can segment continuous stream of activity into discrete event units in ways similar to adults; nevertheless, the same children were less likely to agree on when an event ended and when a new one began. In adults, agreement on event boundaries has often been taken as an indicator of better event unit identification (Zacks & Tversky, 2001) as it has been found to predict memory and action performance (Kurby & Zacks, 2008; Sargent et al., 2013). This suggests that children's event identification shows early competence but has a protracted development.

The current investigation focused on the identification of change-of-state events based on object states used as visual cues. In adults, recent neuroimaging evidence shows that, when processing verbal descriptions of events where an object undergoes a change of state ("He chopped an onion, then he smelled the onion"), adults track the causal history of the object, as evidenced by the fact that the pre- and post-change states of the object (here, the chopped and the intact onion) appear to be in competition (no such competition is experienced when actions are performed on different object tokens, as in "He chopped an onion, then he smelled another onion"; Altmann, 2017; Hindy, Altmann, Kalenik, & Thomspon-Schill, 2012; Hindy, Solomon, Altman, & Thompson-Schill, 2015; Solomon, Hindy, Altmann, & Thompson-Schill, 2015). This suggests that object states include critical information for event identification and the children in our study were able to use this piece of information for event recognition. Nevertheless, event recognition is a multimodal process and is oftentimes based on non-visual input. It remains to be tested whether children can also use additional cues beyond the ones included in the present study for building event representations for themselves and others.

Our findings also have implications for the relation between event cognition and language acquisition. In our experiments, we measured event identification through verb-to-picture matching. Crucially, verbs mapped onto event referents that were themselves experienced in different ways. Studies of verb learning have often focused on how children map verbs onto unfolding, fully observable actions that act as verb referents (e.g., running, breaking or eating; Tomasello & Merriman, 2014); however, children also need to identify verb referents from partial evidence in the world (for instance, they might hear "cracking" when no cracking is visible—but only its aftermath.) For those cases, as in the present study, children need to know what type of event the speaker has accessed and is now talking about. Our research provides some evidence that learners of different languages are alike in the way they extract event information from encounters with different temporal slices of events (cf. the comparison between English- and Turkish-speaking children in Experiment 3). Studies of verb learning therefore need to consider how children map words onto event meanings that might only incompletely be gleaned from visual experience.

Our work opens up several further issues in terms of how children use different types of information sources to gain and attribute knowledge about events. A major issue concerns the very boundary between direct and indirect visual access. In other words, how much visual evidence counts as direct (full) visual access to an event and how much evidence counts as indirect (partial) visual access. In our stimuli, both direct and indirect access appear to have been highly constrained by perception, such that the visual information constrained the kinds of inferences that could be drawn (i.e., when presented with a cracked egg one knows that the event was performed on the egg; however, hearing the sound "cracking" does not constrain the inferences in the same way). Two patterns in our data support this possibility. First, adult controls were equally certain about the kinds of events instantiated in the direct and indirect access trials in the norming studies preceding Experiment 1. Second, in all three experiments, children's performance was similar across the two types of trials. It seems reasonable to assume that other types of indirect visual access might depart from direct perception more clearly. This is consistent with recent claims that, in some cases at least (e.g., a cookie that is missing a piece in the shape of a bite), reconstructing changes of state in an object from the object's shape is a purely perceptual process (Chen & Scholl, 2016; cf. also Landau & Leyton, 1999) whereas in other cases, reconstructing the causal history of an object on the basis of visual input is likely to require higher-order cognitive inference operations (e.g., a wet umbrella might indicate that it's raining outside).

There is substantial variation in the phenomenological degree to which visual perception and inference are perceived to contribute to event identification. In a recent demonstration, adults were presented with photographs that gave either direct perceptual access to an event (e.g., a woman blowing bubbles) or indirect visual cues about the occurrence of the event (e.g., the woman next to bubbles traveling through the air); their task was to say whether they had seen or inferred the event. For directly perceived events, people reported having "seen" them but for indirectly perceived events, responses varied: when the visual cues were not deterministic, people reported having "inferred" the event; by contrast, when the visual cues strongly constrained the inference, people were equally likely to report having "seen" or "inferred" the event (Ünal et al., 2016). An interesting direction for future work would be to pursue such phenomenological intuitions and test whether they capture a distinction between those types of visual inferences that are characterized by speed, automaticity, inscrutability to introspection and other signature properties of the perceptual system (see, e.g., Firestone & Scholl, 2016), and those that are the workings of slower, more reflective judgments that operate over visual premises.

One complication for extending this line of work to children is that their overt phenomenological reports of information access are not always reliable: even when children can use a piece of evidence to identify an object, they have difficulty reporting how they acquired such knowledge (e.g., whether they saw the object, were told about it, or figured out what it was from a clue; Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; Pillow, 1989; Pillow et al., 2000; Wimmer et al., 1988; Woolley & Bruell, 1996). Furthermore, children's ability to report the sources of their knowledge depends on the type of source, with inference being identified later than other sources such as visual perception, sometimes as late as age 6 or beyond (perhaps due to difficulties with the corresponding abstract verbs; Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; Ozturk & Papafragou, 2016). Nevertheless, it remains an interesting question whether children's success with identifying indirectly experienced visual events or attributing such knowledge to others would decrease in cases where children themselves are less likely to say that they have "seen" (as opposed to having "figured out") the event. It is also an intriguing open possibility that these exact cases might require support from higher-order inference to secure event recognition.

Acknowledgments

This research has been supported by NSF grant BCS0749870 to A. P. We thank Yasemin Sandıkçı, Didem Aykut, Taylor Maher, and Sam Katz for their assistance in data collection, the preschools in Istanbul, Turkey and Newark, DE, and all of the children that participated in these studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation [BCS0749870].

ORCID

Ercenur Ünal http://orcid.org/0000-0002-6794-2129

References

Aksu-Koç, A., Küntay, A. C., Acarlar, F., Mavis¸, İ., Sofu, H., Topbaş, S., & Turan, F. (2011). Türkçe'de erken sözcük ve dilbilgisi gelişimini ölçme ve değerlendirme çalışması. Türkçe İletşim Gelişimi Envanterleri. Tige-I and Tige-II *(Measurement and evaluation of early vocabulary and grammar development in early Turkish. Turkish Communicative Development Inventories. Tige-I and Tige-II)* (p. 107K058). Ankara, Turkey: Tubitak report.

Altmann, G. T. M., (2017). Abstraction and generalization in statistical learning: Implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society B*, *372*, 20160060. doi:10.1098/rstb.2016.0060

Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How do infants reason about physical events? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 11–48). Oxford, UK: Blackwell.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*, 708–717. Retrieved from http://www.jstor.org/stable/1132450

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 465–511). Cambridge, UK: Cambridge University Press.

Bunger, A., Skordos, D., Trueswell, J., & Papafragou, A. (2016). How adults and children encode causative events cross-linguistically: Implications for language production and attention. *Language, Cognition and Neuroscience*, *31*, 1015–1037. doi:10.1080/23273798.2016.1175649

Chen, Y.-C., & Scholl, B. J. (2016). The perception of history: Seeing causal history in static shapes induces illusory motion perception. *Psychological Science*, *27*, 923–930. doi:10.1177/0956797616628525

Core Team, R. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundations for Statistical Computing. Available online at: https://www.R-project.org/

de Villiers, J. G., Garfield, J., Gernet-Girard, H., Roeper, T., & Speas, M. (2009). Evidentials in Tibetan: Acquisition, semantics, and cognitive development. *New Directions for Child and Adolescent Development*, *125*, 29–47. doi:10.1002/cd.248

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*, 129–143. doi:10.1016/j.actpsy.2006.07.004

Erguvanlı, E. E. (1984). *The function of word order in Turkish grammar*. Berkeley, CA: University of California Press.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral Brain Sciences*, *39*, 1–77. doi:10.1017/S0140525X15000965

Fisher, C., Gleitman, H., & Gleitman, L. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*, 331–392. doi:10.1016/0010-0285(91)90013-E

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55. doi:10.1207/s15327817la0101_2

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces: Carving up events for learning language. *Perspectives on Psychological Science*, *5*, 33–42. doi:10.1177/1745691609356783

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371. doi:10.2307/1130499

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. doi:10.1111/1467-9280.00255

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology. General*, *142*(3), 880–905. doi:10.1037/a0030045

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thomspon-Schill, S. L. (2012). The effect of object-state changes on event processing: Do objects compete with themselves? *The Journal of Neuroscience*, *32*, 5795–5803. doi:10.1523/jneurosci.6294-11.2012

Hindy, N. C., Solomon, S. H., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). A cortical network for the encoding of object change. *Cerebral Cortex*, *25*, 884–895. doi:10.1093/cercor/bht275

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in the attribution of epistemic states. *Child Development*, *57*, 567–582. doi:10.2307/1130337

Huang, Y., Spelke, E., & Snedeker, J. (2013). What exactly do number words mean? *Language Learning and Development*, *9*, 105–129. doi:10.1080/15475441.2012.658731

Keenan, T., Ruffman, T., & Olson, D. (1994). When do children begin to understand logical inference as a source of knowledge? *Cognitive Development*, *9*, 331–353. doi:10.1016/0885-2014(94)90010-8

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological. Science*, *28*, 1649–1662. doi:10.1177/0956797617719930

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*, 72–79. doi:10.1016/j.tics.2007.11.004

Landau, B., & Leyton, M. (1999). Perception, object kind, and object naming. *Spatial Cognition and Computation*, *1*, 1–29. doi:10.1023/A:1010073227203

Melis, A. P., Call, J., & Tomasello, M. (2010). 36-month-olds conceal visual and auditory information from others. *Developmental Science*, *13*, 479–489. doi:10.1111/j.1467-7687.2009.00892.x

Moll, H., Carpenter, M., & Tomasello, M. (2014). Two- and 3-year-olds know what others have and have not heard. *Journal of Cognition and Development*, *15*, 12–21. doi:10.1080/15248372.2012.710865

Nappa, R., Wessell, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, *5*, 203–234. doi:10.1080/15475440903167528

O'Neill, D., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology*, *27*, 390–397. doi:10.1037/0012-1649.27.3.390

Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development*, *12*, 199–230. doi:10.1080/15475441.2015.1024834

Papafragou, A. (2015). The representation of events in language and cognition. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 327–345). Cambridge, MA: MIT Press.

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, *103*, 253–299. doi:10.1016/j.cognition.2006.04.001

Pillow, B. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*, 116–129. doi:10.1016/0022-0965(89)90066-0

Pillow, B. (1999). Children's understanding of inferential knowledge. *Journal of Genetic Psychology*, *160*, 419–428. doi:10.1080/00221329909595555
Pillow, B. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development*, *73*, 779–792. doi:10.1111/1467-8624.00438
Pillow, B. H., & Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *British Journal of Developmental Psychology*, *24*, 823–849. doi:10.1348/026151005X83360
Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology*, *36*, 169–179. doi:10.1037//0012-1649.36.2.169
Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
Povinelli, D. J., & deBlois, S. (1992). Young children's (*Homo sapiens*) understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology*, *106*, 228–238. doi:10.1037/0735-7036.106.3.228
Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (So long as they are looking into a single barrel). *Child Development*, *61*, 973–982. doi:10.1111/j.1467-8624.1990.tb02835.x
Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. New York, NY: Oxford University Press.
Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences*, *21*, 962–980. doi:10.1016/j.tics.2017.08.005
Ruffman, T. K., & Olson, D. R. (1989). Children's ascriptions of knowledge to others. *Developmental Psychology*, *25*, 601–606. doi:10.1037/0012-1649.25.4.601
Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Head, D., Kurby, C. A., … Beck, T. M. (2013). Event segmentation ability uniquely predicts memory across the lifespan. *Cognition*, *129*, 241–255. doi:10.1016/j.cognition.2013.07.002
Saxe, R. (2013). The new puzzle of theory of mind development. In M. Banaji & S. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 107–112). New York, NY: Oxford University Press.
Shipley, T. F., & Zacks, J. (2008). *Understanding events: From perception to action*. New York, NY: Oxford University Press.
Sodian, B. (1988). Children's attributions of knowledge to the listener in a referential communication task. *Child Development*, *59*, 378–385. doi:10.2307/1130317
Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, *58*, 424–433. doi:10.1111/1467-8624.ep7253729
Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, *62*, 753–766. doi:10.1111/j.1467-8624.1991.tb01567.x
Solomon, S. H, Hindy, N. C, Altmann, Gerry T. M, & Thompson-Schill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal Of Cognitive Neuroscience*, *27*(12), 2324-2338. doi: 10.1162/jocn_a_00866
Spelke, E. S., Phillips, A. T., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). New York, NY: Oxford University Press.
Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*, 1821–1826. doi:10.1111/cdev.12247
Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*, 409–415. doi:10.1016/j.cognition.2011.04.007
Strickland, B., & Scholl, B. J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General*, *144*, 570–580. doi:10.1037/a0037750
Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62. doi:10.1016/j.cognition.2014.12.007
Taylor, M. (1988). The development of children's understanding of the seeing-knowing distinction. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 207–225). Cambridge, UK: Cambridge University Press.
Tomasello, M., & Merriman, W. E. (2014). *Beyond names for things: Young children's acquisition of verbs*. New York, NY: Psychology Press.
Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language*, *87*, 157–176. doi:10.1016/j.jml.2015.10.009
Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179–199. doi:10.1016/j.jml.2015.12.001
Ünal, E., Trueswell, J., & Papafragou, A. (2017). *How children map event participants onto language. In E. Ünal (Chair), Encoding events in language and cognition*. Paper presented at the14th International Congress for the Study of Child Language, Lyon, France.
Ünal, E., & Papafragou, A. (2018). Relations between language and cognition: Evidentiality and the sources of knowledge. *Topics in Cognitive Science*. Special issue on lexical learning. doi:10.1111/tops.12355
Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development*, *59*, 386–396. doi:10.1111/1467-8624.ep8589275

Wimmer, H., Hogrefe, G. J., & Sodian, B. (1988). A second stage in children's conception of mental life: Understanding informational access as origins of knowledge and belief. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 173–192). Cambridge, UK: Cambridge University Press.

Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology*, *32*, 335–346. doi:10.1037/0012-1649.32.2.335

Zacks, J. M. (2017). *Event segmentation and event memory across the lifespan*. Talk delivered at the workshop on Event Representations in Brain, Language, and Development, Nijmegen, The Netherlands.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3–21. doi:10.1037/0033-2909.127.1.3

Appendix

List of stimuli (verbs translated from Turkish)

Access	Photograph	Matching Verb	Mismatching Verb
Direct Visual	a woman pouring coffee into a cup	pour	open
Direct Visual	a woman squeezing a can	squeeze	comb
Direct Visual	a woman falling off a bicycle	fall	open
Direct Visual	a woman blowing candles on a cake	blow	twist
Direct Visual	a man opening a jar	open	comb
Direct Visual	a woman drinking milk	drink	twist
Indirect Visual-Familiar	footsteps on snow	walk	stack
Indirect Visual-Familiar	broken window	break	bite
Indirect Visual-Familiar	shredded carrot next to a shredder	shred	close
Indirect Visual-Familiar	a piece of a play-doh with a hole in it	punch	wipe
Indirect Visual-Familiar	paper torn into four pieces	tear apart	close
Indirect Visual-Familiar	mashed potatoes next to a masher	mash	stack
Indirect Visual-Familiar	a doll taken apart into pieces	take apart	wash
Indirect Visual-Familiar	eggshells	crack	hit
Indirect Visual-Familiar	a bag on its side with books falling out	drop	wipe
Indirect Visual-Familiar	blocks spread around	knock down	wash
Indirect Visual-Familiar	crushed milk carton	crush	bite
Indirect Visual-Familiar	two halves of a lemon next to a knife	cut	hit
Indirect Visual-Unfamiliar	paint footsteps on asphalt	walk	stack
Indirect Visual-Unfamiliar	broken glass object	break	bite
Indirect Visual-Unfamiliar	shredded yellow squash next to a shredder	shred	close
Indirect Visual-Unfamiliar	a slice of avocado with a hole in it	punch	wipe
Indirect Visual-Unfamiliar	carton torn into four pieces	tear apart	close
Indirect Visual-Unfamiliar	mashed butternut squash next to a masher	mash	stack
Indirect Visual-Unfamiliar	a perfume bottle taken apart into pieces	take apart	wash
Indirect Visual-Unfamiliar	two halves of a walnut	crack	hit
Indirect Visual-Unfamiliar	a paper bag on its side with cocktail umbrellas falling out	drop	wipe
Indirect Visual-Unfamiliar	pieces of wood spread around	knock down	wash
Indirect Visual-Unfamiliar	a crushed carton	crush	bite
Indirect Visual-Unfamiliar	two halves of a turnip next to a knife	cut	hit