Universitat Pompeu Fabra Barcelona

Disjunctive Syllogism in Infants: a Multimodal Study

Kinga Anna Bohus

CBC Infant's Reasoning and Cognition Group

Brain and Cognition Master's Project

Director: Luca L. Bonatti

Barcelona, September 2015

Abstract

A study investigated prelinguistic infants' looking behaviour in a multimodal object tracking scene. Participants were presented with a task where correct assessment of the location of objects required association of auditory and visual cues and drawing an inference. The inference was a Disjunctive Syllogism (i.e. Object A *or* Object B, *not* Object A, *therefore* Object B). The results indicate a preference for tracking conspecifics rather than artefacts. The findings are compatible with the hypothesis that mental logic can be present before usage of logical operators in natural language.

Introduction

Newborns face a complicated problem deciphering the blooming, buzzing confusion that surrounds them. They need to organize the incoming information from all their senses to a coherent unity. One important aspect of this is representing the intermodal relationships between incoming information, by detecting temporal cooccurrence of stimulus characteristics such as duration, rhythm and intensity. Many intermodal events give both amodal and arbitrary information. For example, when a person speaks their mouth and voice gives synchronous amodal information. At the same time, the pairing of their face and voice is arbitrary.

Multimodal Object Tracking

The integration of information from several perceptual sources is undoubtedly fundamental for any organism. We have evolved in a way that we integrate singular sensations into multimodally unified objects and events. Perhaps the most important is the integration of correlating visual and auditory information. Newborns already show a capacity to bind visual and auditory information. By 4 months infants perceive speech bimodally (Kuhl & Meltzoff, 1982), and by 5-7 months they match faces with voices (Bahrick, Netto, & Hernandez-Reif, 1998). By 6 months infants can perceive objects bimodally and use auditory information to identify and disambiguate visual events (e.g., Scheier, Lewkowicz, & Shimojo, 2003) In an experiment by Clifton and colleagues 6.5 month old infants were presented with two sounding toys of different sizes, then presented with one of the sounds in the dark (Clifton et al. 1991). Their reaching was in preparation to the size of the object, which indicates that they associated the sound to the toy.

Infants also need to acquire the ability to individuate and track objects. Object based attention is carried out by object-files - representational units of the visual system, stored in working memory. They open automatically for visual objects to which spatio-temporal indexes are assigned (Pylyshyn, 2001). It has been shown that the visual system can track the locations of several objects as they move through space (Scholl & Pylyshyn, 1999) and the encoding of location information is highly automatic (Nissen, 1985). This preattentive mechanism of the visual system tracks multiple objects at the same time by attaching an index to each object that serves as link between objects in the world and the mental representation of its properties (Pylyshyn, 2001). Investigating the existence and flexibility of spatial indexing Richardson and Kirkham (2004) observed that dynamic spatial indexing tends to multimodal events and studied its dynamics.

Indexing is a strategy to save working memory, by only storing location information, and using this information to "look-up" properties as they are needed (Ballard et al., 1997). Interestingly, it has been shown that subjects systematically fixate on an empty space when they have to recall properties of an object that previously occupied that location (Spivey & Geng, 2001). Furthermore, Spivey and Richardson extended this result by showing that not just visual properties, but auditory, semantic information are recalled by looking at empty regions of space that were associated with the information.

The adult experiment by Richardson and Spivey (2000) showed four ports on a screen. Each port had a visual stimulus (first a person speaking, and later just a spinning cross) presented and at the same time subjects heard a factual statement. In the test phase the ports were cleared and subjects were asked about one of the facts. While recalling the information, subjects increased looking to the critical port (where the fact was presented) even though there was no visual cue there. These results were extended to dynamic events: subjects increased looking to the critical port, even when the ports were moved before the memory task (Richardson & Kirkham, 2004).

Richardson and Kirkham (2004) extended these results by designing a simplified version of the experiment to test if the indexing system is already present at 6 months. Infants were shown two ports on a screen. 6 month olds were shown in one port a toy that moved and made a specific sound, then in the other port another toy with a specific movement and sound. The familiarization showed the two toys in sequence three times (so a total of 6 familiarization trial). The two test trials were conducted with two empty ports and one of the sounds playing. Infants gazing pattern was recorded and they found the same indexing strategy as with the adults. This result held up even when the ports were moved before the test (two ports on the left and right side were moved clockwise or counterclockwise to vertical alignment). These results show that as young as 6 months, infants look at the location of a specific object even if the object cannot be seen and the location moves after hearing the associated sound.

The indexing strategy seems to be aiding working memory capacity. What is surprising, is that adults do not only associate locations to semantic facts but they launch an eye-movement to that location while remembering that fact. It seems that we have a system in place that uses indexes to conserve working memory. The index (or pointer) is associated with some kind of coarse semantic information, and the spatial pointer can be triggered when sensory input activates that semantic information (Richardson, Dale & Spivey, 2007). The pointers allow the organism to perceptually access relevant properties of the external world when they are needed without storing all the information. This system sometimes indexes information that is no longer there - as in the case of the adults in the fact-remembering experiment. These experiments describe a system that seems to treat auditory and semantic information as if they were physical objects. This could be because over the course of evolution our visual system had to encode objects and locations and language came later. This is supported by the fact that the system is already in place at 6 months, when linguistic capacities have yet to develop.

Logical reasoning

By the time infants reach 12 months hey show the ability to reason about several aspects of physical events. Gergely and colleagues showed that they can form expectations about future states based on agents' rationality (Gergely et al. 1995). They expect agents' actions to be goal-oriented which suggests that they form a hypothesis based on some kind of rational reasoning.

It is an important question when logical reasoning abilities start to form. We have reason to believe that language plays an important role in the foundation of concept development. For example, developing a notion of exact numerosity, and counting is linked to the availability of names for the numbers (Pica et al., 2004). Similarly, spatial orientation has been shown to be linked to linguistic development (Hermer & Spelke, 1994). These studies suggest that the emergence of natural language plays a crucial role in merging information to formulate logically coherent representations.

It is under debate whether preverbal logical concepts exist. Formal logic is by its nature difficult to study without linguistic testing. Studies in language acquisition provide some insights on this matter. For infants, every new word can have infinite possible meanings. It is a problem of induction, infants need to reach a conclusion about what a certain word refers to considering countless alternatives. Children need to narrow the possibilities using certain constraining principles. Children's tendency to map novel labels to novel objects might be a strategy to constrain these possibilities. The principle underlying this tendency is termed Mutual Exclusivity.

Mutual Exclusivity (ME) is the assumption that every object has just one name (Merriman & Bowman, 1989). A novel label in conjunction with two objects, one of which has a known name, must refer to the novel object. Halberda (2006) showed subjects two objects, one known and one novel. The subjects were asked to look at the "dax" (novel name). Adults and pre-schoolers tested with the same stimuli showed a particular gazing strategy: after hearing the novel label, they first looked to the known object and then finally fixated on the unknown object. This gazing pattern reflects a process of elimination: it is not the name of A, therefore it must be the name of B (Halberda, 2006). Halberda (2003) tested infants to investigate whether they use the Mutual Exclusivity strategy. He found that infants at 17 months show a tendency to fixate on the novel object after hearing the novel label.

The ME strategy is best explained with the use of the logical argument called disjunctive syllogism (although there are other possible explanations). Disjunctive Syllogism is any argument of the form: A *or* B, *Not* A, *Therefore* B. It is a process of elimination, in which a conclusion is reached by systematically rejecting all other possibilities. The process does not need to be in the form of explicit verbal reasoning. It can possibly appear at any level of computation, as a conscious and explicit reasoning, or unconscious, encapsulated process.

Whether such hypothesis-testing is possible before the acquisition of language is not clear. Linguistic behaviour provides insight into the development of logical concepts and representations. Using natural language words such as "not" and "either … or" are indicative of representations of logical operators of *negation* and *disjunction*, respectively. Studies show that the emergence of the comprehension of negation is after 24 months (e.g. Austin et al, 2014). Whether logical concepts representation is a result of linguistic development or some logical concepts are more innate, and might even help language acquisition is an open question. We tried to examine this question by investigating whether Disjunctive Syllogism is present in prelinguistic infants.

Previous experiments

An experimental procedure has been developed by Cesana-Arlotti and colleagues (in prep.) to investigate infants' deductive abilities. Subjects were shown videos of events involving two objects, a cup and an occluder, first in a familiarization phase, where the basic properties and actions of the scene were introduced, then a test phase where infants' looking behaviour was recorded.

The Scooped Object Task is designed to test the ability to perform a disjunctive syllogism in population with limited linguistic capacity. A logical problem is presented in an animated movie. The movie shows a scene with two objects, Object A and Object B, and a cup. The two objects have different shape, colours and belong to different categories (e.g. human, non-human). However, the upper part of the objects are identical, thus when the lower part is covered the two objects are visually undistinguishable (Figure 1.). Participants view the two objects enter the scene then get covered by an occluder. The empty cup rises and goes behind the occluder (from above at exactly

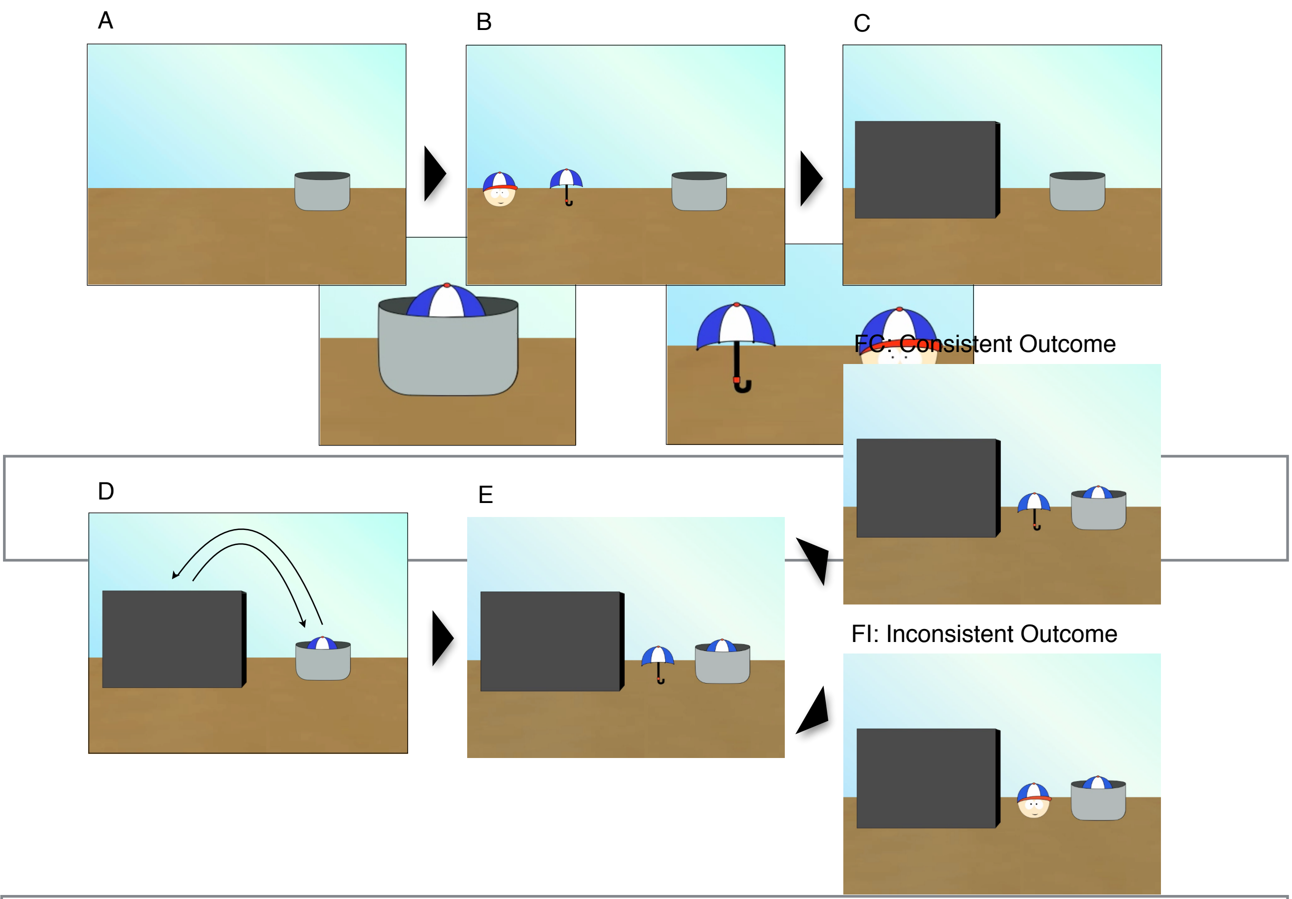

Figure 2. Sequence of events in Cesana-Arlotti's experiment

the centre of the upper edge) and then reemerges with one object inside and settles in its original position. The contained object's identity is ambiguous: spatiotemporal and features information is insufficient to determine which object is inside the cup. Thus, there are two alternatives, the object in the cup is *either* Object A *or* Object B.

In the test movies subjects view the following sequence of events (Figure 2.):

1. Entering. The scene begins with a clear stage and an empty cup on the right side (Figure 2A). Two objects enter the scene by falling from above and then make a distinctive movement (Figure 2B). The upper part of the objects are identical.

2. Occlusion. The two objects get partially and then fully covered by an occluder rising from the ground (Figure 2C).

3. Scooping. The cup gets up from the ground, moves behind the occluder and comes out with an object inside (Figure 2D). The upper part of the object is visible, which indicates that there is an object there, but not enough for identifying which of the two objects it is (Object A *or* Object B is in the cup).

4. Object exit. One of the objects exits from the right side of the occluder for a short period of time (Figure 2E). At this moment, the identity of the object in the cup can be inferred by disjunctive syllogism (*it is not* Object A, *therefore it must be* Object B). The object goes back behind the occluder.

This sequence of events constitutes the basis of several experiments. The main goal is to find out whether prelinguistic infants can infer the identity of the object in the cup by excluding the other possibility. In experiments designed by Cesana-Arlotti, the author used the Violation of Expectancy paradigm, where the different final scenes were either compatible or incompatible with the logical expectation regarding the location of the objects. The last scene showed another object exiting from the right side of the occluder. In the Consistent Outcome (Figure 2FC), this object was visually identical to the one that exited before (Object A), while in the Inconsistent Outcome (Figure 2FI) the object was identical to the other object (Object B). In the current experiment we tried a new approach that does not rely on the Violation of Expectancy paradigm, but instead uses multimodal features to investigate object tracking and the presence of prelinguistic logical inference. The question is whether multimodal processes access inferential processes and guide complex actions through integration of information.

Experiment 1.

The experiment was designed to elicit a particular ocular movement based on participants' belief about the contained object's identity. We added sounds to the target objects, which made it possible to attract attention to them without revealing their location. Our hypothesis was that if infants hear the sound associated with the object they will visually attend to the location of where they believed the object is located. So we predicted that if infants draw the inference, it should be reflected in the visual pattern that is elicited by the sound of the (partially or fully) covered object.

Adding particular sounds to the viewed objects can be benefitting. Providing a redundancy of information (visual and auditory features) can facilitate the formation of multimodal object representation. Several studies show that infants are more likely to attend to modality-specific information when they are allowed multimodal experiences with the objects (e.g. Bahrick & Lickliter, 2000). Experiments by Wilcox and colleagues provided evidence that multimodal (visual and tactile), but not unimodal (only visual) exploration of objects increased 10.5 month olds sensitivity to colour differences, even though object individuation based on colour information has been shown to arise only at about 11.5 months (Wilcox et al., 2007). Thus, the multimodal aspect of our experiment might be advantageous. On the other hand, the task is complicated and integration of the information is necessary which can hinder performance.

Method

Participants

We tested two age groups. The younger participants were sixteen healthy 12-month-old infants (10 female, between 11 month 19 days - 12 month 14 days) from Spanish or Catalan-speaking families. Three participants had to be excluded due to technical difficulties or inattentiveness.

The older participants were thirteen healthy 19-month-old infants (10 female, between 19 month 10 days - 20 month 2 days) from Spanish or Catalan-speaking families. Four participants had to be excluded due to technical difficulties or inattentiveness.

Stimuli

Visual stimuli consisted of 6 familiarization movie and 8 test movie. The movies were animated in Keynote 5, exported at 60 fps and subsequently edited in Quicktime 7 Pro. The experiment was run on an Apple Mac Pro computer controlled by Psyscope X. The stimuli appeared on a 24-inch TFT screen. Eye movements were recorded with a Tobii T60 XL Eye Tracker. The stimuli movie size was 1450x830 pixels. For the auditory stimuli we recorded a native Spanish/Catalan speaker say various sentences.

The familiarization movies were constructed to introduce all aspects of the objects and actions of the test (see Appendix A). In these movies participants viewed all the the basic elements of the scene except for the final sound. After the familiarization each participant viewed 8 test movies. The test movies were always about the same pair of objects: the boy (F = Face) and the umbrella (U = Umbrella). Their upper part was visually identical. The movies depicted the same scene as in Cesana-Arlotti's experiment with one added feature, namely that objects now had specific sounds. The test movies have shown the following scene:

1. Entering. Two objects enter the scene by falling from above and then make a distinctive greeting (Figure 3A). The umbrella makes a movement by rhythmically moving its handle aligned with a distinctive sound. The boy greets the subjects by saying one of two possible phrases: "Hi! Do you want to play with me?" or "Hi! Let's play!" in Spanish or Catalan (Figure 3B).
2. Occlusion. The two objects get partially and then fully covered by an occluder rising from the ground (Figure 3C).
3. Scooping. The cup gets up from the ground, moves behind the occluder and comes out with an object inside (Figure 3D-3E). The upper part of the object is visible, which indicates that there is an object there, but not enough for identifying which of the two objects it is (Object A *or* Object B is in the cup).
4. Object exit. One of the objects exits from the occluder for a short period of time (Figure 3F). At this moment, the identity of the object in the cup can be inferred by disjunctive syllogism (*it is not* Object A, *therefore it must be* Object B). The object goes back behind the occluder.
5. Baseline. The scene of the occluder and the cup stays still for 3 seconds (Figure 3G).
6. Test phase. Keeping the same visual scene, one of the objects' sound is played (Figure 3H). For the umbrella it is the same sound as in the beginning. The boy's call is one of the sentences "Hi! Can you find me?" or "Hi! Where am I hiding?" in Spanish or Catalan. The sound is played two times, first it takes 2 s then after 0.5 s silence, again for 1 (the boy says the other sentence) or 2 s (the umbrella makes the same sound). (So, in the case of the boy subjects heard e.g. "Hi! Can you find me?" then after half second silence "Where am I hiding?") After the second sound plays there is again 3 s silence.

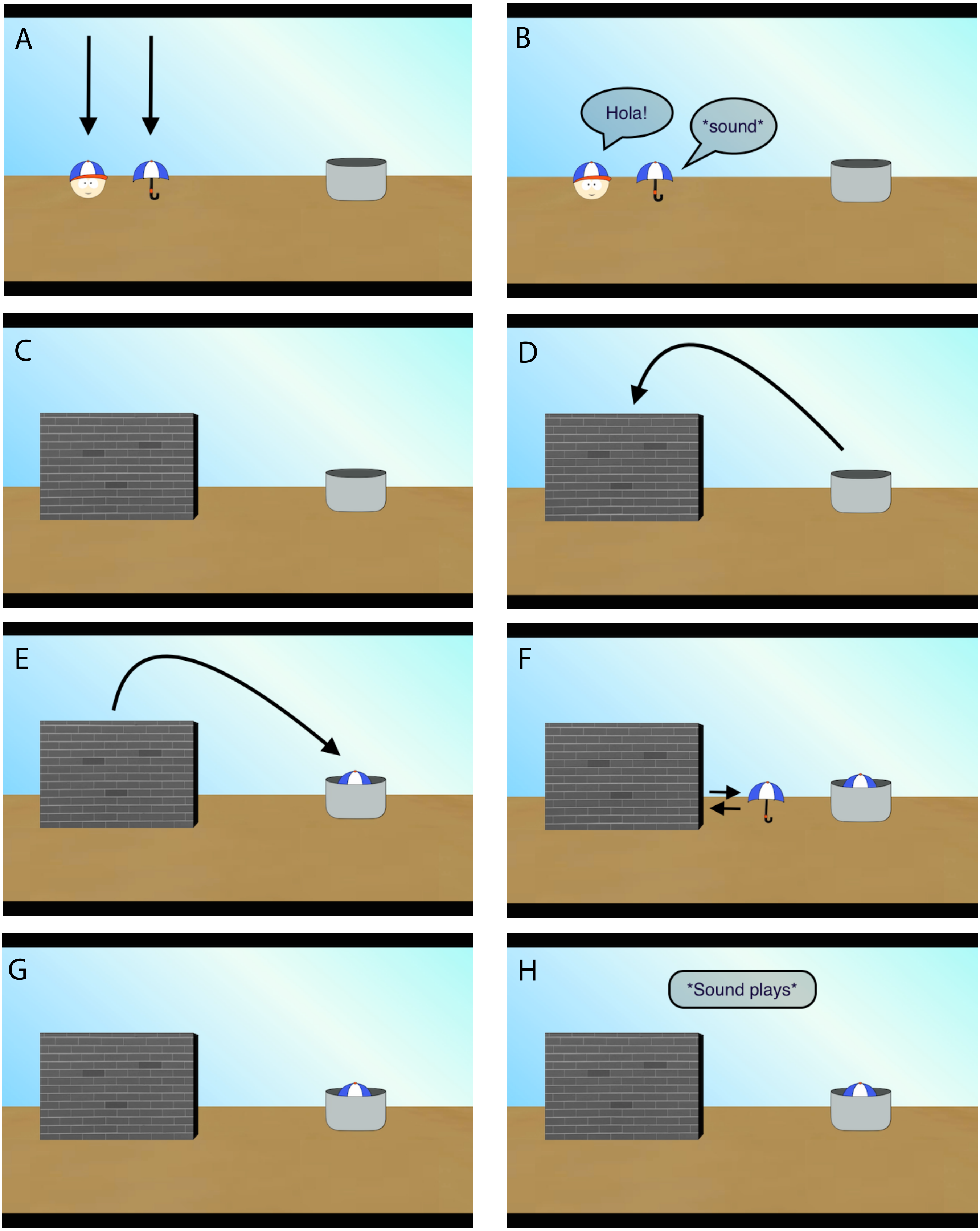

Figure 3. Sequence of events in Experiment I.

The object that was seen exiting from the occluder and the played sound defined where the target was. If the sound was the same that corresponded to the occluded object, the target was the occluder (TO, Target = Occluder), and if the sound was the other object's sound, the target was the cup (TC, Target = Cup). After the test phase there was a reward movie, where the target object appeared in the middle of the screen with a colourful background and danced to cheerful music.

Procedure

Subjects were tested in a quiet room, approximately 60 cm from the computer screen. They were sat on their caregiver's lap. The caregiver wore opaque glasses so they could not see the stimuli and influence the child's response. They were instructed not to interact with the infant verbally or non-verbally. The experimenter monitored infants via a camera placed above the screen from a separate room. Before the baseline the movies stopped playing when the infants' gaze was not recorded by the eye-tracker, which ensured they see every movie in its entirety.

Before starting the familiarization, the eye-tracker was calibrated. The calibration procedure consisted of five objects successively appearing in the four corners and in the middle of the screen. In the Familiarization Phase participants were shown 6 movies. The first two movies introduced the cup function without the occluder: the cup picks up an object, carries it then releases it. The third and fourth movie introduced the cup function with the occluder: an object gets occluded, the cup picks it up from behind the occluder then releases it. These two movies also introduced the test objects, the Face and the Umbrella. The last two familiarization movie showed the cup function when two objects were behind the occluder.

The Test Phase consisted of eight movies. Each participant saw all possible combinations of target location twice (Occluded Object: Face or Umbrella, Played Sound: Voice or Sound).

Results

To analyze the eye-tracking data first I had to define Areas of Interests (AOIs). Because of individual differences in the calibration and head movement the recorded gaze coordinates do not correspond to the exact same location on the screen. Thus the individual AOIs had to be extracted from the data. I defined the AOIs using the recorded gaze points at a certain time slice of the test movies: when the occluded object exits from behind the occluder (I averaged the recorded coordinates of a short time period, when fixation was almost certainly on the object). This provided for each subject a point of reference from which I defined the AOIs for the cup and the occluder based on the original coordinates of the animation. For each time point (measured at a 60 Hz frequency, that is 60 gaze points per second) I coded if the gaze was inside either of the AOIs.

The main dependent variable used in the analysis was looking proportion. In the analysis looking proportion was calculated as time spent looking to the Cup divided by the total time spent looking to either AOI (Cup or Occluder). This measure does not take into account recorded gazes that were outside of the AOIs, such as looking to the side of the screen or elsewhere in the room. Fixations to locations other than the AOIs could be signs of attention lapses (fixating to areas without any visual content or outside of the screen) or missing the point of the task (searching for

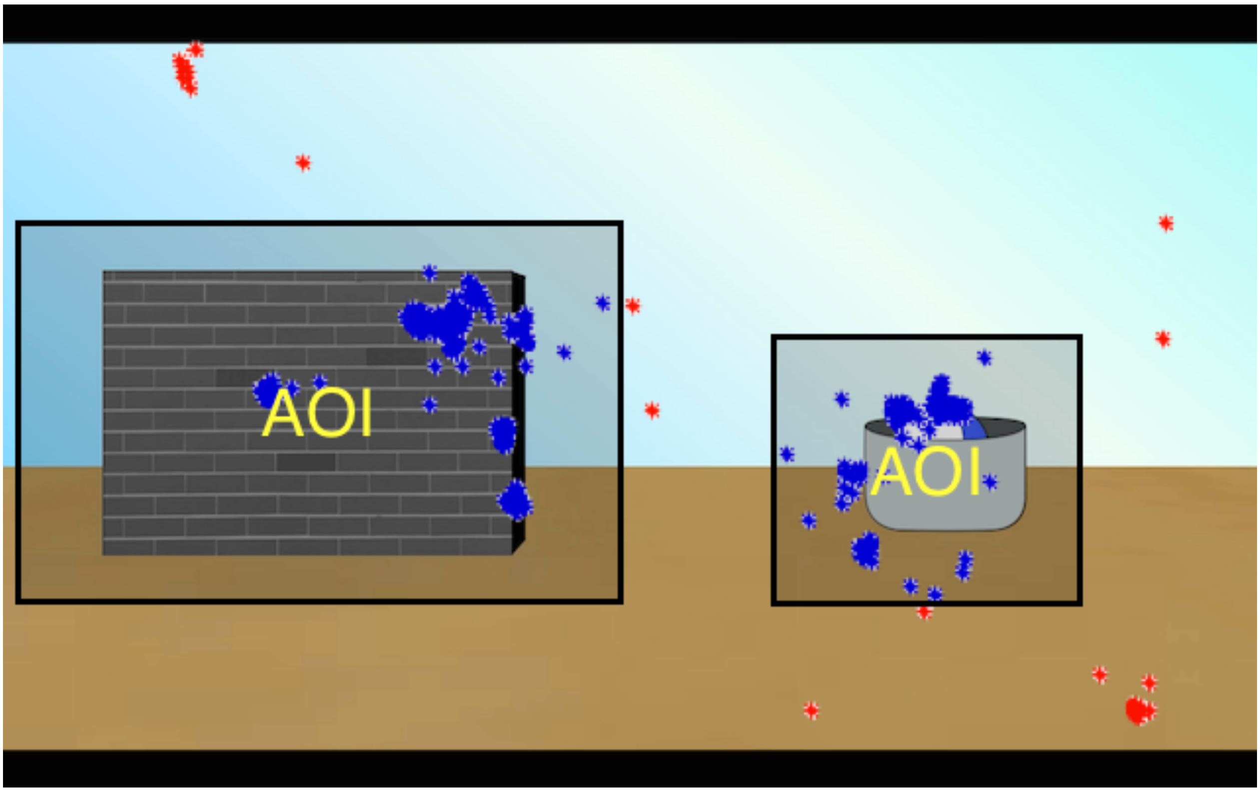

Figure 4. Example of AOIs. In a certain trial, the recorded gaze points are either inside one of the AOIs (blue points) or outside (red points).

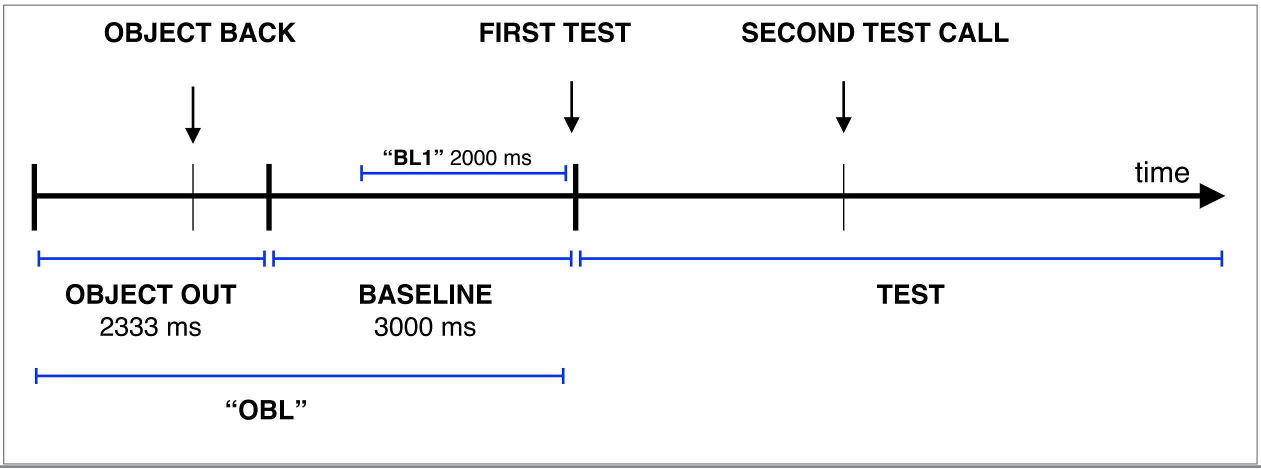

Figure 5. Time course of the test phase.

the objects around the room, behind the screen etc.). Thus, the looking proportion to the cup measure has the same content as the looking proportion to the occluder (LP cup = 1 - LP occluder).

Figure 5. depicts the time course of the test movies, starting from the point, when the object exiting from the occluder is completely visible (Figure 3F). The Object Out Phase is when the object is visible and then moves back behind the occluder. The Baseline Phase is the three seconds without any movement showing only the scene with the two target locations: the Cup and the Occluder. The Test Phase starts when the first test call is played and ends three seconds after the second test call offset. The Object Out and Baseline phases make up the "OBL" Phase. The last two seconds of the Baseline is the "BL1" Phase. Looking proportion in the Object Out Phase was total looking at the cup divided by total looking at either the cup, the visible object or the occluder.

We first consider the "OBL" Phase. Before this phase, the identity of the object in the cup is ambiguous. In the "OBL" Phase all information is available to draw the inference: seeing Object A exit and return behind the occluder specifies that the object in the cup must be Object B. For each subject, looking times were collapsed by Object type (object in the cup), creating an overall looking proportion for each object in the OBL Phase. Looking proportions for each subject in this period were entered into 2 (Age) $\times$ 2 (Object) repeated measures ANOVA to test for differences in looking proportion between trials when the cup's content was the Face and when it was the Umbrella. There was a main effect of Object ($F(1,23) = 10.45$, $P < 0.01$) and no effect of Age ($F(1,23) = 0.166$, $P = 0.69$). There was no interaction effect detected between Object and Age ($F(1,23) = 1.079$, $P = 0.31$). The mean looking proportion to the cup for the 19 month olds were 21.1 % when the cup contained the Umbrella, and 31.75 % when the cup contained the Face. The mean looking

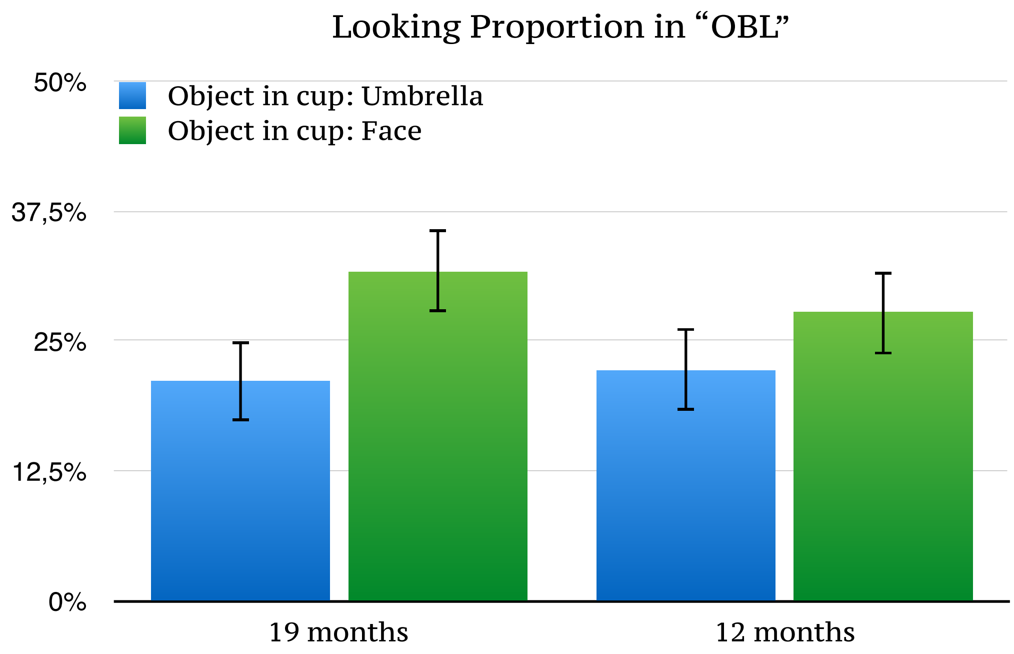

Figure 6. Mean looking proportions in the OBL phase.

proportions for the 12 month olds were 22.24 % when the cup contained the Umbrella and 27.71 % when the cup contained the Face (Figure 6.).

This result indicates that infants in both age groups were more likely to look at the cup in this phase when the cup contained the Face than when it contained the Umbrella. Participants showed an interest in the human-like object, and looked proportionally more to the location where it was hidden. For the 19 month olds this effect was significant in both the Object Out and the Baseline phases. In the Object Out phase, where an object was moving participants had a higher tendency to shift their gaze to the cup when the visible object was the Umbrella. In the Baseline phase participants looked more to the cup, after they saw the Umbrella move back behind the occluder. These findings show that infants are more inclined to track the boy's location than that of the other object.

Investigating whether looking proportions increased or decreased upon hearing the test sound can provide valuable information about how infants track multimodal objects. The baseline looking proportions we consider here are the looking proportions calculated in the "BL1" Phase. The reason for this is that in the beginning of the baseline the gazing is still influenced by the object going back behind the occluder, as the movement generally attracts visual attention. In the "BL1" Phase participants looking proportion is likely to reflect the actual attention to the two AOIs without any visual bias. The looking proportions in "BL1" Phase were entered into 2 (Age) × 2 (Object) repeated measures ANOVA. Again, there was an effect of Object ($F(1,27) = 6.438$, $P <$

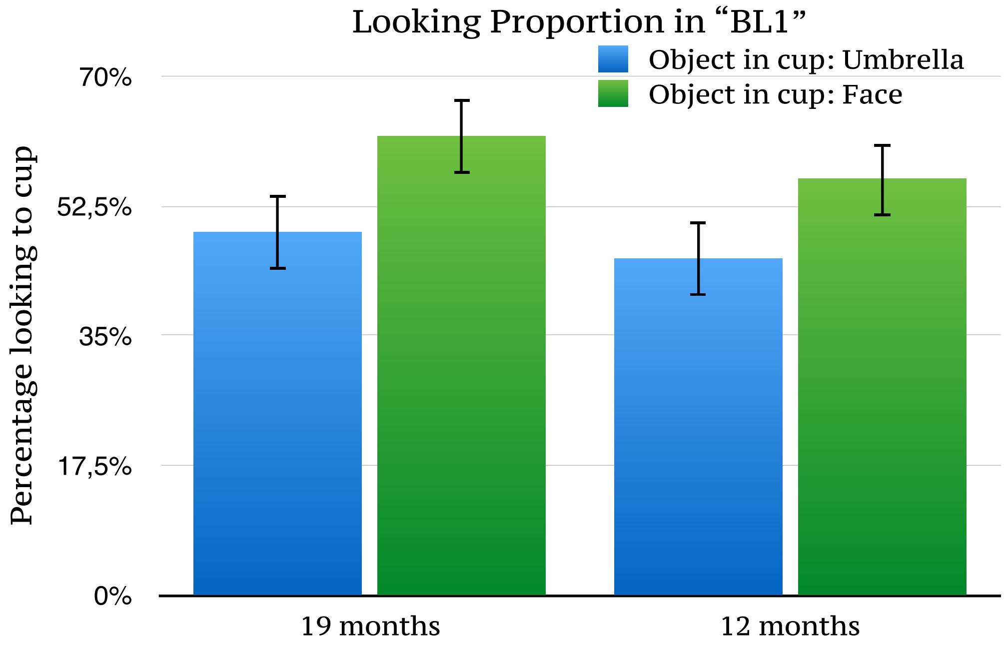

Figure 7. Mean looking proportions in the "BL1" phase.

0.05), and no interaction. The mean looking proportions in the 19 month olds case were 48.78 % when they previously saw the Face and 61.98 % when they saw the Umbrella. The mean looking proportions in the 12 month olds case were 45.46 % when they previously saw the Face and 55.97 % when they saw the Umbrella (Figure 7.).

Increases in looking proportion were calculated by subtracting the "BL1" looking proportions from the looking proportions of the Test Phase. As we have already seen there are already differences between looking proportions in the baseline based on which object came out from behind the occluder. Therefore, we analyze the effect of the played sound separately in the two conditions. Because we had four conditions (two objects and two sounds), some subjects had to be excluded from the analysis because of insufficient data. Two subjects were rejected from the 19 month group and seven subjects from the 12 month group, because they did not have sufficient data for all four combinations (recorded gazes were under 50 % in the test phase).

First we analyze how looking proportions changed when subjects had seen the Face emerging from the occluder (and the cup contained the Umbrella). For each subject, looking times were collapsed by trial type (Played Sound = Voice or Sound) to create the mean differences in looking proportion. The difference scores were compared with a paired t-test. For the 19-months the effect of Played Sound was not significant ($t(10) = 1.3$, $P = 0.28$). Neither sound elicited a significant increase or decrease in the looking proportion. Similarly, for the 12-months the effect of Played

Sound was not significant (t(8) = 0.345, P = 0.57). Neither sound produced an increase or decrease in the looking proportion compared to the baseline.

The same analysis was conducted for the trials when subjects saw the Umbrella (and the cup contained the Face). For the 12 months paired t-test revealed no effect for Played Sound (t(8) = 0.224, P = 0.65). Neither sound changed the looking proportion significantly. For the 19 months there was an effect of Played Sound (t(10) = 9.452, P < 0.05). The sound of the Umbrella did not change the looking proportions significantly. The Voice decreased the looking proportion to the cup on average by 23.24 percentage points (t(10) = -3.384, P < 0.01, Figure 8.).

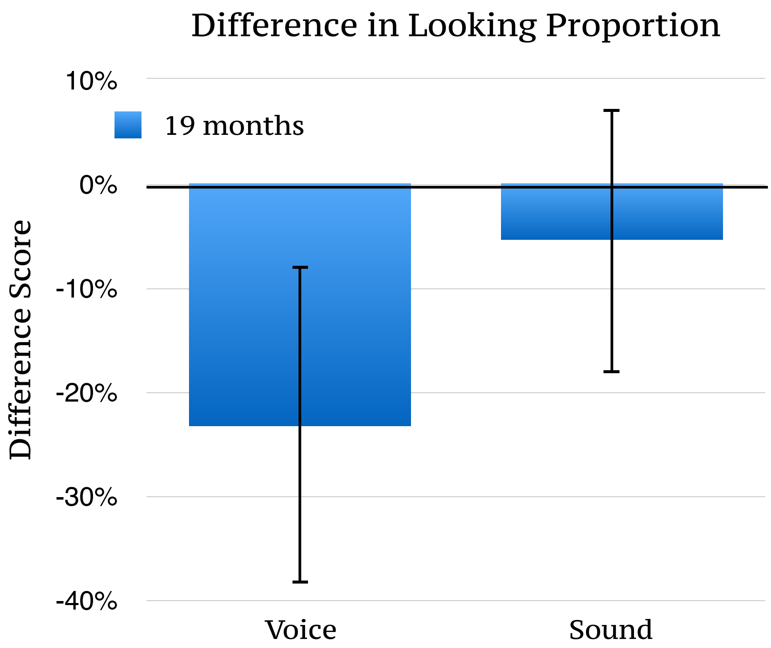

Figure 8. Difference scores when the occluded object was the Umbrella for the 19 month olds. Listening to the Voice decreased looking proportion to the cup.

Discussion

Our main finding was that infants showed a specific interest in tracking the Face rather than the Umbrella. 19 month olds shifted their gaze to the cup when it contained the Face even when the other object was still visible and moving. This might indicate that they inferred that the boy was in the cup when they saw the Umbrella. Infants also kept looking more to the cup when the umbrella went back behind the occluder. This result is consistent with the hypothesis that they inferred the content of the cup based on the process of elimination i.e. disjunctive syllogism. However, it is a possibility that they did not make the inference, but rather just kept fixating to the boy more when it came out from behind the occluder, and kept looking more to the occluder, after they saw the boy return behind it (compared to the other object). Even though this possibility cannot be excluded, we believe that an aspect of the data can lead us to favour the first possibility. Indeed, had participants not inferred that the Face was in the cup after seeing the Umbrella, they would have looked for it behind the occluder. Thus, they should have fixates on the occluder more, and this should decrease their looking to the cup. The fact that we did not observe this pattern of results could be interpreted in favor of the inference hypothesis.

After hearing the sound, looking proportions did not show significant changes in all but one case. When 19 month olds heard the Voice after seeing the Umbrella, they decreased looking to the cup and increased looking to the occluder. Our original assumption was that the looking proportion would increase to the location to where the object is located when we played the sound associated with that object. However, even in the cases when participants listened to the sound of the object behind the occluder which they saw recently and the location of which was known, they did not increase looking to that location. This suggests that playing the sound does not necessarily motivate them to fixate on the right location. In other visual search tasks, participants are rewarded by the target object being revealed (usually dancing to music). In these and other tasks there are visual rewards for anticipatory gazes: infants look to a certain location because they expect something to happen there. Our paradigm was different from almost all other eye tracking tasks. We did not reveal the final location of the objects, only the target object itself in the middle of the screen. Our reason not to reveal the final locations of the objects was to avoid that infants would develop a learning strategy to look to the cup without actually making the inference.

Based on our results it is likely that subjects were mostly interested in tracking the Face - rather than looking to the correct location. When they saw the Face behind the occluder, they kept their original looking pattern which is consistent with the hypothesis that from the beginning of the experiment they were mostly interested in the hiding place of the Face.

When 19 month olds saw the Umbrella go behind the occluder and heard the voice of the boy they increased looking to the occluder in proportion to the cup. This seems to be indicative of some kind of search pattern related to the inference. The increased proportion to the occluder might be related to the earlier mentioned multimodal spatial indexing experiment of Richardson & Kirkham (2004). There experiment is one of the very few that records gazing in response to auditory stimuli without any visual reward. An explanation to our findings could be that when participants heard the voice of the boy, they looked to the occluder as an automatic response. The auditory speech stimulus activated the spatial pointer previously associated with that location. This resulted in an automatic oculomotor response. Indeed, subjects sometimes fixated to the location between the occluder and the cup, where they have previously seen Object A. This indicates that they might be looking at certain locations to recall the events they saw.

Another possibility to explain why infants increased looking to the occluder might be found in a strategy similar to the one participants were shown to use in the Mutual Exclusivity experiment (Halberda, 2006). In this study, when hearing the novel label subjects first looked to the known object and then shifted back to the novel object. This pattern indicates a process of elimination strategy. In our experiment subjects might have recalled the identity of the recently seen object and looked to the occluder as part of an elimination process.

Young infants benefit from paying special attention to their conspecifics. Infants between 2 and 7 months have been shown to prefer speech compared to structurally similar complex non-speech sounds, because it is important for separating and selecting speech out of the auditory information for language learning (Vouloumanos & Werker, 2004). Similarly, newborns orient to face-like stimuli (Farroni et al., 2005), which helps the development of face recognition (Johnson, Dziurawiec, Ellis & Morton, 1991). Bonatti and colleagues showed that 10 month old infants possess information about their conspecifics and use it to identify objects (Bonatti et al. 2002). Mental representation of individuals or agents seem to be different than object representation (Murez & Smortchkova, 2014).

Further study

The analysis of this data did not consider the time course of the Test Phase. We might find valuable information if we investigate further the time course of the searching behaviour.

Unfortunately there were not enough subjects and too many conditions to find general rules guiding the looking patterns.

Participants seemed to care much more about finding the boy than the umbrella. The boy acts more like an agent, and his call might be viewed as a goal-oriented action (Sommerville et al., 2004), which might motivate infants to search for his location (especially because he asks “Where am I?”, “Can you find me?”). Thus, a possible future experiment could use only humans, rather than contrasts between humans and artefacts.

It may be noticed that in our design the plausible areas of interest are sharply different. One presents a cup, and the other presents an occluder. We opted for this decision for compatibility with previous experiments. However a cleaner version of the experiment which we have already started testing would present two identical areas of interest. This would be obtained by adding another occluder in the final scene covering the cup. This would make the visual scene completely symmetrical and might help us by reducing the visual bias towards the cup in the test phase. Unfortunately, we have only tested very few subjects with this condition, so the results could not be discussed here.

Another possibility of finding out more about the inference is incorporating the data gazes that were recorded but were not in the AOIs. This might help us show that subjects look to the cup with the same conviction as they look to the occluder, instead of looking around in a scattered fashion. Prolonged, intentional fixation to the cup should support the hypothesis that they inferred what subject is contained within.

Conclusion

Infants bind multimodal information and use spatial indexing to form object representations and track objects from a very early age. They show a higher interest in tracking their conspecifics than other multimodal objects. In a Scooped Object task we found that infants at 12 months and 19 months tracked a speaking, human-like agent with more interest than a sounding, moving object. They seemed to have inferred where the boy was hidden by performing a disjunctive syllogism. This could provide evidence that mental logic concepts are already present before the formal linguistic operators, such as “not” (negation) and “either … or” (disjunction) are used.

Bibliography

Austin, K., Theakston, A., Lieven, E. & Tomasello, M. (2014). Young children’s understanding of denial. *Developmental Psychology*, 50(8), pp. 2061-70.

Bahrick, L. E., Netto, D. & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, 69, pp. 1263–1275.

Bahrick, L.E. & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36, pp. 190- 201.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, pp. 723–767.

Bonatti, L., Frot, E., Zangl, R. & Mehler, J. (2002). The Human First Hypothesis: Identification of Conspecifics and Individuation of Objects in the Young Infant. *Cognitive Psychology*, 44(4), pp. 388-426.

Clifton, R. K., Muir, D. W., Ashmead, D. H. & Clarkson, M. G. (1993). Is visually guided reaching in early infancy a myth? *Child Development*, 61, pp. 1098–1110.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D. & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *PNAS,* 102 (47), pp. 17245-17250

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, pp. 165–193.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87, pp. B23- B34.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), pp. 310-344.

Hermer, L. and Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature,* 370, pp. 57-59.

Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2), pp. 1-19.

Kuhl, P. K. & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218, pp. 1138–1140.

Merriman, W.E. & Bowman, L.L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54, pp. i+iii+v+1-129

Murez, M. & Smortchkova, J. (2014). Singular Thought: Object-Files, Person-Files, and the Sortal PERSON. *Topics in Cognitive Science,* 6(4), pp. 632-646.

Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. Posner & O. Marin (Eds.), *Attention and performance XI: Mechanisms of attention.* pp. 205–220. Hillsdale, NJ: Erlbaum.

Pica, P., Lemer, C., Izard, V. & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science,* 306, pp. 499-503.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80, pp. 127–158.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, 76, pp. 269–295.

Richardson, D. C. & Kirkham N. Z. (2004). Multimodal Events and Moving Locations: Eye Movements of Adults and 6-Month-Olds Reveal Dynamic Spatial Indexing. *Journal of Experimental Psychology,* 133(1), pp. 46–62.

Richardson, D., Dale, R. & Spivey, M. (2007). Eye movements in language and cognition: A brief introduction. In Gonzalez-Marquez, C. & Mittelberg, S. (eds.) *Methods in Cognitive Linguistics*. pp. 323-344. John Benjamins.

Scheier, C., Lewkowicz, D. J. & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, 6, pp. 233–244.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple objects through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38, pp. 259–280.

Sommerville, J. A., Woodward, A. L. & Needham, A. (2005).Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96, pp. B1-B11

Spivey, M. J., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65, pp. 235–241.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, 7(3), pp. 270–276.

Wilcox, T., Woods, R., Chapa, C. & McCurry, S. (2007). Multisensory exploration and object individuation in infants. *Developmental Psychology*, 43(2), pp. 479-495.

Appendix A: Familiarization Movies

Familiarization	Object in the scene	Actions
1	Bear	Empty stage with an empty cup on the right. The object falls down from above and makes a distinctive movement by moving its arms paired and aligned with a distinctive sound. The occluder rises and covers the object then falls back down. The cup rises and takes the object to the cups initial position. The cup drops and reveals the object. The object makes a movement paired with a sound.
2	Star	Empty stage with an empty cup on the right. The object falls down from above and makes a distinctive movement by spinning around paired and aligned with a distinctive sound. The occluder rises and covers the object then falls back down. The cup rises and takes the object to the cups initial position. The cup drops and reveals the object. The object makes a movement paired with a sound.
3	Umbrella	Empty stage with an empty cup on the right. The object falls down from above and makes a distinctive movement by rhythmically moving its handle paired and aligned with a distinctive sound. The occluder rises and covers the object. The cup rises and takes the object to the cups initial position. The cup drops and reveals the object. The object makes the movement paired with the distinct sound.
4	Boy (Face)	Empty stage with an empty cup on the right. The object falls down from above and greats the subject by saying “Hola peque” in Spanish or Catalan. The occluder rises and covers the object. The cup rises and takes the object to the cups initial position. The cup drops and reveals the object. The boy says “I'm here!” in Spanish or Catalan.
5	bear and star	Empty stage with an empty cup on the right. The objects fall down from above and make their distinctive movements paired with their sound. The occluder rises halfway then goes back down, then rises completely. The cup takes an object. One of the objects comes out from behind the occluder to the middle then goes back.
6	bear and star	Empty stage with an empty cup on the right. The objects fall down from above and make their distinctive movements paired with their sound. The occluder rises halfway then goes back down, then rises completely. The cup takes an object. The other one of the objects comes out from behind the occluder to the middle then goes back.