8-2021

The Role of Language in Building Abstract, Generalized Conceptual Representations of One-And Two-Place Predicates: A Comparison Between Adults and Infants

Jill de Villiers
Smith College, jdevilli@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/phi_facpubs

Part of the Philosophy Commons

Recommended Citation

The role of language in building abstract, generalized conceptual representations of one- and

two-place predicates: a comparison between adults and infants.

Mohinish Shukla[¶], Jill de Villiers[§],

[¶]University of Massachusetts Boston, [§]Smith College

Abstract

Theories of relations between language and conceptual development benefit from empirical evidence for concepts available in infancy, but such evidence is comparatively scarce. Here, we examine early representations of specific concepts, namely, sets of dynamic events corresponding either to predicates involving two variables with a reversible, asymmetric relation between them (such as the set of all events that correspond to a linguistic phrase like "a dog is pushing a car,") or to comparatively simpler, one-variable predicates (such as the set of events corresponding to a phrase like "a dog is jumping."). We develop a non-linguistic, anticipatory eye-tracking task that can be administered to both infants and adults, and we use this task to gather evidence for the formation and use of such one-and two-place-predicate classes (which we refer to as *event sortals*) in 12-24-mo-old infants, and in adults with and without concurrent verbal prose shadowing. Using visually similar stimuli for both the simpler (one-place) and the more complex (reversible, asymmetric, two-place) concepts, we find that infants only show evidence for forming and generalizing one-place event sortals, and, while adults succeed with both kinds in the absence of verbal shadowing, shadowing hampers their ability to form and use the asymmetric two-place event sortals. In a subsequent experiment with adults, we find that if the shadowing material is grammatically impoverished, adults now succeed in forming and using both one- and two-place event sortals. We discuss implications of these results for theories of concept acquisition, and the role of language in this process.

TITLE: The role of language in building abstract, generalized conceptual representations of one- and two-place predicates: a comparison between adults and infants.

Most theories of conceptual development attempt to explain the ontogeny of concept acquisition as an interaction between some primitive representations of infants' perceptual inputs, and some mechanisms for their analysis. These theories differ both in their proposals for what constitutes the primitive representations; which range from sensory-motor representations (e.g., Gallese & Lakoff, 2005) to agents, objects, and numbers (e.g., Spelke & Kinzler, 2007); and what the mechanisms for analysis are, ranging from distributed, neural-network-like computations (e.g., Shultz, 2012) to rational inference (e.g., Tenenbaum, Griffiths, & Kemp, 2006). Concurrently, developmental scientists have been cataloging empirical evidence for the presence of concepts in prelinguistic infants (e.g., Carey, 2011 contains an overview of such evidence in a variety of domains).

Language has long been recognized as an important factor that could contribute to conceptual development due to its ability to create novel meanings, in part, by combining previously known words and phrases in formal algebraic, syntactic structures; so that we can immediately come to understand what it might mean if we hear someone tell us that they saw a *purple apple*. Further, such novel creations are abstract, and can refer not just to immediately perceptually available tokens, but also to *sets*, for example, of all apples that are purple. When such creations become commonplace, we can even substitute them for single words; instead of referring to 'unmarried men,' we just say *bachelor*, and if it were common in our milieu to fish with one's hands by groping under the stones or banks of a stream, then we could simply refer to

each instance of such an event with the word *guddling*. In this sense, language is ideally suited for the development of ad hoc, structured conceptual categories that are of functional relevance, but which might not be innately specified.

The question we are interested in is, to what extent are our concepts reliant on language? The development literature on prelinguistic infants' innate knowledge of numbers, objects, conspecifics, etc. alluded to above suggest that at least some concepts might be innately specified. Could it be that language is required for the development of more complex concepts?

In this work, drawing on previous literature (detailed below), we propose that sets of dynamic events involving two participants with an asymmetric relation involving both (e.g., a dog pushing a car) are relatively complex compared to sets of dynamic events involving only one participant (e.g., a jumping dog, with an irrelevant car beside it), and that the former might require the support of language to be properly represented as a category. We seek empirical evidence that can speak to this issue.

Situating this research project

The research described here falls between two existing bodies of work. On the one hand, its roots lie in the literature on *sortals* in infancy. How does an infant acquire basic object level sortals, namely the criteria for individuation (knowing how to divide and count objects) and identity (knowing what properties a kind has) (Xu, 2007)? The surprising failure of ten-month-olds' representations is revealed in the sequential presentation of objects behind a screen, where infants default to something like "physical object" to represent basic objects like cup, truck etc. (for review, see Stavans, Li, Wu & & Baillargeon, 2019). By twelve months,

infants seem to have more stable sortal criteria for many basic object level concepts, with shape among the criteria. The failure is only in *sequential* presentation, in what have been called 'event mapping paradigms': it is not that infants fail to discriminate basic object level kinds when they are presented together in space and time. Furthermore, the failures at ten months can be overcome by the objects displaying (conceptually) different kinds of motion or functions (Futó, Téglás, Csibra, & Gergely, 2010). Research also suggests that infants even at 9 months distinguish sortals from broadly distinct ontological categories such as human versus nonhuman, or animate versus inanimate (Surian and Caldi, 2010).

Events too have properties of individuation and identity, though they have not usually been considered in this light (but see, e.g., Zacks & Tversky, 2001). Yet for the purposes of making event descriptions in language, one needs to know when two events fall under the same description and when they do not. And, it is possible to count events just as it is possible to count object kinds (such as instances of *guddling*). A small but growing set of studies in this area, beginning with Gordon (2003), uses habituation procedures with infants to ask what must change for an event to be considered unlike the original. For example, Gordon asked infants at the end of the first year of life what counts as an event of *giving*. In particular, do infants recognize which aspects of the event are essential participants, namely the *arguments*, and which are not? Take the event wherein a boy hands a girl a teddy bear. The giver and the recipient are essential, but so also is the item transferred. At ten months, the infants noticed if the teddy bear was absent in a later trial once they have habituated to this event. In a control, Gordon contrasted this with an event of a boy hugging a girl in which the teddy bear was merely an adjunct to the main event. Infants did not dishabituate when the teddy bear was absent from a subsequent event of hugging.

The striking suggestion is that infants are already partitioning events into representational structures that contain certain essential parts (arguments) and inessential (adjuncts), before language even starts.

The second literature is on toddler's attempts to map events onto semantic or syntactic representations (discussed in greater detail in the next Section). In this literature the child is shown a novel action event while also being provided with a verbal label, such as "Look, gorping". One question is about syntactic bootstrapping: can toddlers tell that a transitive sentence maps to a causal event rather than a noncausal event? That is, "Look, the bear is gorping the duck" should refer to the causal event involving the bear and the duck, rather than the event in which both bear and duck are present but engaged in parallel actions like arm flapping (Naigles, 1990). The second question is whether the children can tell which event to attend to: should it be the bear doing something to the duck or the duck doing something to the bear? That is, at what age can children map the syntactic subject onto the semantic class of agent, and the syntactic object onto the semantic class of patient. In this literature, the prior conceptual analysis of a given scene into agents and patients is generally assumed (but see Perkins 2019). Golinkoff and Kerr (1978) found that toddlers aged 15 to 18 months noticed if the actors (two men) switched thematic roles after habituation, when such scenes were presented without accompanying language.

In Gordon's studies, we do not know if the infant would notice if *different* boys and girls were involved in the events, rather than the one pair shown to them. Can the infant form a general, rather than a specific representation of the event during exposure? That is, is the event

represented more generally like "Boy gives Teddy bear to Girl", or more specifically like "*This one* gives *thing* to *that one*"?

The findings by Gordon shows that prelinguistic infants can attend to the relevant dimensions of a dynamic event, namely, which are the arguments and which are adjuncts when multiple characters are involved. What is the limit on this? Do prelinguistic infants also see the *resemblance* across these action events when the characters change identity? That is, can they conceive of such events as sets, thereby allowing other events with different participants, sharing the same linguistic description ("boy giving girl a toy"), to be recognized as similar events, when they do not yet have access to the structures underlying such linguistic descriptions? Our work is designed to address these issues.

What would motivate the formation of such a class of events? We might see someone involuntarily grinding their teeth, and understand this dynamic event unfolding before us as 'someone involuntarily grinding their teeth' We might not (unless we were dentists) think of this single event as a potentially infinite *class* of events. And yet, if we *were* dentists, then we *would* think of this event as a member of a class of events, and we would even have our own special name for it – bruxism.

We would also like to note here that when we use the term "event," we recognize that these occur at different levels of abstraction. "Pushing," "giving," "guddling," occur, but also more abstract and complex dynamic scenarios such as "accidents" and "funerals" or "football games", that may not only vary in participants but also are composed, either hierarchically or sequentially, of a large number of individual actions. When these clusters occur often enough, they are nominalized.

In other words, whether or not we *should* represent a class of events would presumably depend on a host of factors such as the utility of such a class. We are interested in the underlying ability of being able to build such (novel) representational classes, and our specific question is, does the representation of such classes rely on language?

Concepts that might (not) require language.

To begin examining the relation between language and concepts, de Villiers and colleagues asked whether adults can form an implicit representation of a set of events that are alike, in the sense that they are all true under the same linguistic description: "event sortals" by analogy with object sortals. The methodology involved implicit concept formation in an eyetracker; that is, adults were simply told to attend to what was happening but no instruction was given, and no response was required. Critically, and in contrast to the experimental paradigms used to study syntactic bootstrapping, there was no language used in any of the stimuli.

In each of a series of trials, participants viewed two side-by-side pictures on a screen. The two pictures depicted a contrasting pair of events that differed in their linguistic descriptions, one of which was designated the "target" for a given participant. For example, a participant in one such experiment might have seen a pair of pictures depicting *a monster chasing a dog* and *a dog chasing a monster*. On each trial the target event for that participant (e.g., monster chasing dog) turned into an animation after three seconds. The question was, would participants confronted with the initial pair of still pictures look in anticipation at the side of the one that would animate, having extracted the rule that ties the events together?

Importantly, the issue is not about whether a participant can "parse" the scene into a monster, and chasing, and a dog. Nor is the question about whether the participant can see either event as composed, for example, that the monster is the agent and the dog the patient. Instead, it is a question about equivalence classes: can the participant come to recognize the set of such events as a class to be distinguished from the other class in which the dog is the agent and the monster the patient? Is there anything that motivates the existence of such a trained class except for its equivalence in linguistic description? Notice that in order to come to appreciate the correct event class and show successful anticipations, participants need to be able to hold in memory some representation that allows a distinction between monster-as-agent events from dog-as-agent events from previous trials.

In one such study de Villiers asked whether a participant could attend to the difference across two sets of closely matched pairs of pictures, one in which a parent kisses a child, and the second with the same participants, in which a child kisses a parent (de Villiers 2014). In short, the content to be attended to is a reversible transitive event, with the tokens (the individuals) constantly varying. Half of the participants were simultaneously engaged in a dual task, namely, shadowing an audio book after a few minutes of pre-experiment practice. The other half watched with no verbal interference. The results revealed that verbal interference disrupted the adult participants' ability to anticipate which of the two stills would animate. Comparing looking time to target versus non-target for the non-shadowing group of subjects revealed highly significant differences in looking time. In the absence of shadowing, adults easily learn to look in the direction of the still that depicted the "same" event over trials, namely "adult kiss child". However, when verbally shadowing, adults did not look consistently at the correct still image, in

fact they were reduced to chance levels. In the verbal shadowing condition, with language tied up, these participants could not track the commonality across the events captured in the propositional description. In a second experiment the stimuli were more tightly controlled stimuli, of a schematic dog and monster engaged in chasing events, and the rule that tied the targets together was "dog chase monster". The results were replicated.

In other work, Nordmeyer (2011) working in the same lab and using the same procedure with implicit anticipatory looking, tested whether adult participants could form a concept of negation, i.e., not-*p*, looking at pairs of pictures in which the "negative" example animated. For example, the pair might consist of a woman sawing wood, versus another woman standing with a saw but not sawing. Another type of negation might be where something failed to happen, for example, one set of images showed a match successfully lighting a fire, and in the negative case, the match fizzled out. The shadowing participants in the negation task failed completely to register the property that the negative stills had in common, looking equally often at target and non-target. De Villiers (2014) and Nordmeyer (2011) argued that this was because the negative pictures have nothing in common without language mediation.

Importantly, Nordmeyer (2011) also tested whether another abstract idea that is not in itself propositional might show the same pattern. In the second experiment she tested whether adults in the implicit concept formation experiment could differentiate a target that was a natural kind versus an artifact. Pairs of pictures showed stimuli quite well matched in perceptual properties: a sock versus a wooly bear caterpillar, a pile of pennies versus a pile of shiny rocks – but the target that animated was in each case the natural kind. Looking time to the target versus the non-target in the three seconds immediately prior to animation of the still constituted the measure, that is

the anticipatory gaze at the right kind of stimulus. Despite the abstract nature of the contrast, shadowers and non-shadowers alike easily distinguished the class of pictures that would animate and looked in anticipation after a very few trials.

Crucially, Nordmeyer's control can just as easily be considered a control condition for the reversible transitive condition, in that the same basic design was used, and the same methodology and participant group. This lends support to the finding that verbal shadowing selectively interferes with the recognition of equivalence classes that depend on language. It is a control against the criticism that shadowing is having an effect at some more general level like distracting attention away from the task, or getting in the way of response selection. De Villiers (2014) and Nordmeyer (2011; Nordmeyer & de Villiers, 2019) argue that some concepts have a similarity that emerges at a linguistic level, dependent on propositions (either reversible transitive events, or negation). When adults have full control of their language faculty these are easily recognized as sets. These same sets cannot be integrated by participants whose language faculty is tied up by verbal shadowing. But if the concept is not tied to linguistic formulation, such as the fundamental distinction between natural kinds and artifacts, shadowing does not disrupt the attentional processes needed to integrate across even these highly diverse events. Such concepts, then, could be considered as potential candidate primitives in infant cognition.

The present study: comparing one- and two-place predicate computations in infants and

adults

To summarize, concepts differ in their complexity, and the proposal is that simpler concepts are likely to be those for which we can find evidence in prelinguistic infants, and whose expression might not depend on language and its prodigious representational power. The critical question is, is this representational power of language *required* for the development of more complex concepts?

Consider again the reversible, two-place predicates described in the previous studies. To understand that a dog is chasing a monster requires more than just a combination of *dog*, *monster*, and *chase*. That is, even if the individual entities and the fact that it is a *chasing* event are all understood by the viewer, a memory representation that can identify a separate future event as an instance of this event must additionally include information about which kind was the chaser, if they are to correctly identify such a future event and distinguish it from one in which a monster is chasing a dog.

The previous adult studies, in particular those with two-place predicates, suggest that such complex concepts might not be ontogenetically primitive, and might require language for their development. Our primary goal here is to therefore investigate the developmental origin of the observed difficulty with reversible two-place predicates.

Towards this end, we developed an eye-tracking paradigm that allows us to gather comparable data with infants and with adults. Closely modeled on previous paradigms (e.g., de Villiers, 2014), this eye-tracking study presents participants with pairs of events that differ in the thematic role assigned to the members of each of two categories (dogs and cars), and tracks the evolution of anticipatory looks towards the designated target event (e.g., "car pushes dog") across the trials of an experimental session. We further compare the performance of infants and

adults with such two-place predicates to their performance with conceptually simpler, one-place predicates (e.g., "the dog rolls"), using very similar dynamic stimuli. We expect to find better performance with one-place, compared to two-place event generalizations.

Developmental characteristics of our infant population

We chose to examine 12-24-mo-olds. As we describe in detail in this section, this age range is of particular interest for the following reasons:

1. By this age, infants have fairly sophisticated understanding of dynamic events with two (or more) participants. As an important aside, several of these studies use computer-animated stimuli, as in our research.

2. Specifically, infants by 15 months can discriminate between agents and patients when presented with dynamic events (Golinkoff & Kerr, 1978).

3. While infants in this age range appear to have some idea about syntactic structures that can support asymmetric two-place predicates; i.e., transitives in English, they nevertheless do not show adult-like knowledge of such structures in production or comprehension. Children at the two to three word sentence stage at 18 to 24 months rarely produce reversible transitive sentences and evidence of comprehension of such sentences is weak even in two year olds, though eye-gaze reveals some sensitivity to the agent (Gertner, Fisher, & Eisengart, 2006).

That is, throughout this age range we can be fairly certain that infants are able to correctly identify individual stimulus videos as examples of dog-push-car or car-push-dog events, but the typical linguistic representations that could be used to encode such scenes are immature. If infants can nevertheless learn the contingency between a specific predicate such as

13

car-push-dog and identify novel instances as belonging to this class, it would be evidence for a prelinguistic representational system capable of representing the set of events corresponding to cars pushing dogs without relying on linguistic (transitive) structures.

From a cognitive perspective, several researchers have shown that even infants as young as 6-9 months of age appreciate the distinction between causal events (a ball striking another and making the second ball move) and non-causal events (a ball moving towards another ball, and the latter moving before the first reaches it; e.g., Leslie, 1982; Leslie & Keeble, 1987, Muentener & Carey, 2010; Kominsky et al, 2017). Recently, these findings have been extended to 3-mo-olds (Liu, Brooks, & Spelke, 2019), suggesting early developmental origins of understanding events involving multiple entities.

Older infants are sensitive to the *relation* between syntactic structures of language and the semantics of observed events (e.g., Fisher, 1996; Lidz, Gleitman & Gleitman, 2003; Abbot-Smith, Lieven, & Tomasello, 2004; Lidz & Gleitman, 2004). It has been amply demonstrated that toddlers at the end of their second year use linguistic cues – in particular, the presence of multiple noun phrases in an utterance – to infer transitive structures, and therefore map such utterances to dynamic events with a causal component. For example, Naigles (1990) showed 24-month-olds scenes with a duck and a bunny where either the duck-bunny interaction was causal (e.g., one tilting the other's leg or causing the other to adopt a bent position) or non-causal, (e.g., both the duck and the rabbit performing the same hand-flexing gesture, or both covering their eyes). Such displays were accompanied with a nonce word in either a transitive frame ("the bunny is gorping the duck) or an intransitive frame ("the bunny and duck are gorping"). Using the preferential-looking paradigm, Naigles (1990) observed that the toddlers

preferred looking at the causal scene for *gorp* in a transitive frame, and the non-causal scene for the intransitive frame.

Subsequent studies have largely confirmed that the process of learning verbs with multiple thematic roles is in place by around 24 months of age. The data from younger infants is somewhat contradictory. Gertner & Fisher (2012) found that infants treated the mere presence of two nouns in an utterance as indicative of a transitive event, even when the utterance was produced in an intransitive manner (e.g., "the boy and the girl are gorping"). In contrast, Arunachalam, Escovar, Hansen, & Waxman (2013) found that, 21-mo-olds who had merely been exposed to two interlocutors speaking using a novel verb in a transitive frame ("The lady fezzed my brother") preferred to look at a video of one person pushing another compared to a video of those two people performing the same action. However, no such preference was obtained when infants first heard a novel verb in an intransitive, two-participant frame ("The lady and my brother fezzed"). Additionally, infants are even sensitive to the *order* of participants (i.e., in English, their assignment to thematic roles) in a two-predicate sentence. Gertner, Fisher, & Eisengart (2006) found that both 24-mo-olds and 21-mo-olds, exposed to sentences like "the duck is gorping the bunny," looked longer at a video of an action where the duck was the agent, while they looked longer at a video of an action where the bunny was an agent when exposed to "the bunny is gorping the duck."

These observations support the conclusion that learning two-place predicate verbs is getting off the ground towards the end of the second year (see also, e.g., Scott & Fisher, 2009; Waxman, Lidz, Braun, & Levy, 2009; Yuan, Fisher, & Snedeker, 2012). That these competences might not be quite adult-like is suggested by, e.g., Arunachalam & Waxman (2015), who found

that even at 27 months of age, toddlers appeared to require rich semantic support to understand dynamic transitive events. They were able to infer the meaning of a novel transitive verb only when it was used with full NPs ("The boy is pilking the balloon.") but not with pronouns ("He is pilking it."). Further, in their more difficult, act-out task, 3-year-olds could not correctly place agents and patients (e.g., in "Make Big Bird dack Cookie Monster") when they had previously been exposed to only the verb in an uninformative frame ("This is called dacking"). Finally, Arunachalam & Dennis (2019) showed that when 2-yr-olds hear transitive structures (like "The girl is lorping the lady!") in the absence of any referent, their interpretation is restricted to causative events (e.g., the girl pushing the lady) but not contact events (e.g., the girl patting the lady).

We can therefore safely conclude that 12-24-mo-olds perceive and understand two-place, causal events, and are still in the process of developing adult-like assignment of such events to transitive structures in their language. But to reiterate, we are studying implicit concept formation *without* linguistic clues. There are no linguistic prompts in the experiment, and the children are at an age when they are not yet producing reversible transitive sentences themselves. We are interested in seeing if our participants can encode the events as members of the same abstract class, abstracting over different agents and patients. We are not interested in our participants merely identifying a particular scene as being causal (a car pushing a dog). Nor are we merely interested in whether participants can correctly identify the thematic roles of the pusher and the pushee across a series of reversible causative events of cars pushing dogs and dogs pushing cars. We ask instead, can they appreciate that Scooby Doo pushing the Ford is the *same as* Lassie pushing the Audi, or Pluto pushing an Aston Martin? In other words, can they

build representations of dynamic scenes depicting a causal relation between two entities (e.g., a dog pushing a car) that is abstract enough to support the identification of other scenes with different tokens of the entities performing the same action (a visually distinct dog pushing a visually distinct car in a different direction), and yet maintaining the same thematic relation (e.g., rejecting a scene portraying a car pushing a dog)? To build this representation means to encode each scene with just the right degree of abstraction that it can be recognized across trials in memory: not so particular that each event looks different, and not so abstract that the agent and object *kinds* can be switched.

We compare the identification and generalization of such reversible two-place predicates/relations/event sortals with the identification and generalization of simpler, one-place predicates such as a dog rolling (vs a car rolling). Although the two kinds of events are quite distinct, we have tried to make the visual stimuli as similar as possible. Nevertheless, given the substantial distinction between the one- and two-place predicates, we treat these as different experiments, and our analyses therefore do not also include a single factor separating one- and two-place predicate conditions. We first describe results with two-place predicates in Experiment 1, and turn to the simpler, one-place predicates in Experiment 2.

Experiment 1: Generalizing a two-place predicate in infants and adults
The dog pushes the car/ The car pushes the dog

We chose to use the general predicate "a dog pushes a car" (and vice-versa) so that category membership of the two classes would be very clear. We make no commitment about the specific categories used to encode these events. For example, these could be encoded not

17

necessarily as *dogs* and *cars*, but could be encoded as *animal* and *vehicle*, *living* and *non-living*, or as *natural* and *artificial* (see Mandler, 2008). The specific encodings don't matter, as long as participants see the two as distinct classes, and can build structured representations over these relatively more abstract classes (e.g., "a living thing pushes an artificial thing.") Indeed, given the perceptual similarities between the members of the *dog* set and the *car* set (e.g., the presence/absence of legs, wheels, eyes, or differences in shape and size), even perceptual categories over such feature sets would work. That is, while we remain committed to the idea that representing such events requires abstract symbols, we remain neutral regarding the specific symbols chosen. Nothing rests on the details of object representations beyond the findings that infants in their second year of life are capable of distinguishing our two actors as distinct kinds. Our question is, can they build structured memory representations of dynamic events involving the two kinds of actors in a way that maintains their asymmetric agency at this linguistic stage?

In de Villiers (2014) and subsequent studies, the authors used naturalistic stimuli – video recordings of the various dynamic actions. These introduce a host of irrelevant variables that separate the role-reversed videos, including the precise angles of the participants, durations of various segments of the actions, the precise manner of implementation of an action, in addition to perceptual variables like lighting and background color palettes. In order to simplify the task for infants, we decided to construct videos using animation and video editing software. The goal was to reduce the number of irrelevant differences between the videos – both within and across the "event sortals" – to give infants a leg up in learning the contingencies that predicted a reward in the task (see next section).

An important difference from the previous studies was that, instead of beginning each trial with a still image that later animates, we started with showing one complete cycle of the dynamic event in both the left and right on-screen windows, such that during the anticipatory period the participants in both (paused) videos were in neutral positions that gave no indication of their roles during the initial or subsequent dynamic video portions of each trial. Therefore, unlike in the previous experiments described above, participants could not identify the correct side based on information presented during the anticipatory period alone.

Infants and one group of adults passively watched the sequence of videos; a second group of adult participants additionally shadowed an audiobook presented over headphones.

Materials and methods

Reversible, two-place event videos and construction of individual trials

To construct the videos, we started with a set of 16 images of dog and car exemplars that were manipulated to have a variety of hues and saturations, and were cropped and scaled to fit an imaginary box, 270w × 240h pixels. All dogs were scaled to fit precisely within this box. All cars fit the width precisely, but varied in their heights, from between 80 to 128 pixels. We then created a set of animations, each of which had pairs of videos (see Fig. 1). Each pair consisted of a dog-pushing-car event loop and a car-pushing-dog event loop. Note that "pushing" in all videos consisted of the dog or car image moving along the screen towards its counterpart; when the bounding box for the pusher contacted the bounding box for the pushee, the latter began moving backward, as if pushed by the former. Then, both images returned to the original position. The original position, that is, the first and last frames of each video, were identical to allow continuous looping, and consisted of the dog and the car facing each other. Each loop began with

both the dog and the car performing a little "hop" over the first 9 frames (~375ms) – this was done to ensure that a low-level perceptual cue corresponding to *the first moving object* was not sufficient to make the correct generalization. We counterbalanced the side of the two events, and, within the events, the direction in which the pusher pushed – towards the screen center or away from it. Each video had a grayscale version and a color version. Except for the first eight trials (see Methods below), all trials had different combinations of the 16 dogs and 16 cars.

An example of one trial is shown in Figure 1 below (please see https://youtu.be/Qwo3OsTyIiA for a video demo). Briefly, each trial began with the left video animating to show two grayscale loops, followed by the right video animating to show two grayscale loops. The offset of the right video loops was followed by a '+' flashed briefly between the two video frames, which marked the onset of the anticipatory period. Following the 2.5sec anticipatory period, the target video (e.g., the car-push-dog event) animated with the figures in color, and was accompanied by a happy jingle.

Figure 1 here

Trial sequences

We prepared two lists of trials. In one, the target (i.e., the event that turned colorful and animated, accompanied with a jingle) was the dog-push-car event, while in the other the target was the car-push-dog event. Participants were randomly assigned to one of the two lists.

Participants were shown a total of up to 32 trials. The number of trials was chosen to be similar to the ones used in the previous studies (e.g., Nordmeyer & de Villiers, 2019). Given that

this is a first study of its kind, the exact number of trials was an open, empirical question. The first eight trials had the same dog and car tokens, with the side of the critical event and the direction of pushing (towards the center of the screen or towards the screen periphery) counterbalanced across the 8 trials. The next set of 24 trials had a new set of dogs and cars, and over the course of the 24 trials we varied the appearances of specific dog and car tokens. This design detail was inspired, in part, by animal behavior studies, in which a series of "shaping trials" serve to successively approach the desired behavior. Trials sequences were programmed in Psyscope X (psy.ck.sissa.it).

Experimental set up

The infant studies reported here were completed at the first author's institute. The adult studies were completed across both authors' institutes. Participants in the first location were tested using a Tobii TX300 eyetracker, which has a screen resolution of 1920×1080 pixels and samples eye data at 300Hz, and participants in the second location were tested using a Tobii T60 eyetracker with a screen resolution of 1280×1024 pixels and a sampling rate of 60Hz. The videos were designed such that the animations were identically sized on both screens, both in terms of the number of pixels and in terms of their visual extent relative to the viewer, by extending the background of the animations as appropriate to the two different screens. The analyses (see below) were designed to be insensitive to differences in the sampling rates.

Participants

Adults participants across all these studies were recruited primarily from the university undergraduate and graduate populations, and were compensated with either $5.00 or extra class credits. Further, across all adult and infant studies, an equal number of participants was assigned

to the sub-conditions – for example, the Car-push-Dog vs the Dog-push-Car sub-conditions in this experiment.

Here, we had 20 adults (13 females, mean age 22.4y ±4y) in the Verbal Shadowing condition and 20 adults (16 females, mean age 23.2y ±6.7y) in the No-shadow condition. In both conditions, the target for half the adults was Car-push-Dog, while for the other half it was Dog-push-Car. Of these, one participant in the No-shadow condition was excluded for poor data quality (see exclusion criteria described in the *Data analysis* section below).

Infant participants were recruited through a variety of methods including mailings, and online and print advertisements. We tested a total of 25 infants between 12m;9d and 24m;6d in this study (mean age 18m;12d ±125d, 14 females), of which 5 were excluded for poor data quality, and two for eye-tracker failure that resulted in no data. Caregivers were paid $20, free parking, and offered a small gift (lab magnet, sippy cup) for their participation. Each study was approved by the relevant University IRB board.

Procedure

Adults in the Verbal Suppression condition shadowed George Orwell's *Nineteen Eighty-Four*, presented over headphones at a comfortable listening level, and were given a short training session to practice shadowing, before starting the main experiment. An experimenter sat out of view of the adult participant and ensured that no participant paused for more than 2 seconds. All adult participants were instructed to simply watch the sequence of pre-constructed video trials.

Infants sat on their caregiver's lap. Caregivers wore a soft felt visor pulled low over their forehead, allowing them to monitor their baby while obscuring their view of the eyetracker

screen or cameras. Adults sat in a chair in front of the eyetracker. The eyetracker was calibrated using Tobii Studio. Subsequently, participants were shown the sequence of trial videos. Each trial video began with a small looming image of a cross accompanied by a monotonous drum. This loop continued until the participant looked at the screen, at which point the video for that trial started.

Data analysis

In order to compare results across adults and infants, and across the two eye tracking systems, we performed the following: (1) The entire data was centered by replacing the median x,y of the gaze data with the x,y of the screen center. This allows correcting for deviances due to improper calibrations and variations in gaze direction – particularly an issue with infants. (2) Since the two animation 'windows' in each trial were centered on the screen, we removed all gaze points outside the 95th percentile along the x and y dimensions. (3) Gaze data was divided into a series of epochs corresponding to the individual trials. (4) Given that the centered, percentiled data was constrained to the two on-screen animation 'windows', we ignored the y-axis and looked exclusively at the (left-right) x-axis. Finally, given that the first eight trials had identical dogs and cars, and the remaining had different combinations of dogs and cars (and directions of "pushing"), we restricted our analyses to trials 9-32.

We examined individual trials and participants for quality. First, a small fraction of trials had variant durations, most likely due to high priority housekeeping functions in the OS of the machine running Psyscope X. Therefore, any trials where the number of data frames was beyond ± 30% of the expected number of frames (i.e., ±90 frames for a 300Hz eyetracker, or ±12 frames for a 60Hz eyetracker) were rejected.

Next, trials in which the anticipatory period (the focus of the analysis) contained less than 70% of valid data were rejected. In this context, a valid data point was defined as any frame with valid gaze point coordinates, without considering whether the gaze point was from a single eye or from both eyes, and ignoring the validity score for that gaze point. A rejected data point was either a data point that the eyetracker could not provide a value for, or a point that was outside the 95th percentile along the x-dimension.

Finally, we removed participants who did not have at least two un-rejected trials when the target was on the left and two un-rejected trials when the target was on the right. Therefore, each participant contributed at least 4 trials (2 each for target on left/right), and each of these had at least 70% of valid gaze data. To reiterate, this minimum trial requirement is only considering trials 9 to 32; the first 8 trials comprise the "training" or "shaping" trials as described above.

We examined gaze data in the x-direction (the dependent variable) separately for those trials where the target animation was on the left and those where it was on the right (e.g., see Figure 2 below). It is important to note that in our design, each of the two side-by-side stills is successively animated: the left always animates before the right, and both happen before the anticipatory period for all trials. This is needed to demonstrate the order of causal action in each still. As a consequence, we expect a right-ward bias in all of our data, the magnitude of which will depend on individual and group differences in returning to the central fixation. We therefore do not compare deviations from screen center in our analyses, but only measure and report comparisons between mean gaze positions during the anticipatory period for the set of trials where the upcoming target is on the right versus the set of trials when the upcoming target in on the left.

First, we used a bootstrapping method to see if, across the entire trial time, the x-locations differed significantly at specific time-points from the beginning to the end of the trial. In particular, we employed a permutation method similar to that described in Maris & Oostenveld (2007), and Ferry et. al. (2015). For this analysis, each participant contributed two time-series: the mean x-gaze averaged across the trials 9-32 (see above) for the target-on-right trials, and the mean x-gaze averaged across the trials 9-32 for the target-on-left trials. First, frame-wise one-tailed, independent sample t-tests was computed, asking if x-locations in target-on-right trials was significantly larger (i.e., more to the right) than x-locations on target-on-left trials, for all the frames in a trial across all participants in that group. Then, all time-points where the difference was significant at p=0.05 were marked. Next, runs, i.e., continuous frames of significant points, that were at least 100ms in duration were evaluated for their significance by (1) computing the expected t-value for a significance of α=0.05 given the number of comparisons in a particular run (2) randomly shuffling the labels (target-on-left or target-on-right) of each time-series 1,000 times and computing a distribution of t-values for each run, and (3) marking as significant all the time points in a particular run where the expected t-value was in the 5% tails of their shuffled distributions. Note that, since only the labels are shuffled, we preserve any run-wise, within-trial temporal correlations. Surviving runs of significant differences are indicated on the figures for each condition (e.g., see Figures 2 and 3 from Experiment 1).

Next, we examined whether the average eye-gaze horizontal (x-)positions in the anticipatory period were significantly different for the trials where the target animation was on the left or was on the right. An examination of the shuffle analysis described above indicated that, while significant differences were found between target-on-right and target-on-left trials

across the entire anticipatory period for adults, the effect appeared to be localized to the early part of the anticipatory period for the infants (see Figure 9). We therefore used only the first half of the anticipatory window for infants for the analyses described here, but the entire anticipatory period for adults.

Each participant contributed two sets of values – the mean gaze x-location in the anticipatory periods for each target-on-left trial, and the corresponding mean gaze x-location in the anticipatory periods for each target-on-right trials, for a maximum of 48 (2x24) data points per participant. We fit this data to the simplest theoretically meaningful linear 'base' model using the R package *lme4* (Bates et al, 2015). For infants, the simplest model examined the dependent variable mean gaze x-position (MeanGazeX) as a function of whether the post-anticipatory target was to the right or to the left of center (IsTargetRight), and included random participant (Id) slopes:

model <- lmer(MeanGazeX ~ IsTargetRight + (1|Id))

For adults, the corresponding model included whether or not participants underwent simultaneous verbal suppression (IsVerbalShadow):

model <- lmer(MeanGazeX ~ IsTargetRight * IsVerbalShadow + (1|Id))

We subsequently added in a variable corresponding to the specific sub-condition the participant was exposed to, i.e., whether the target for that participant was dog-push-car or car-push-dog. We only included this variable if its inclusion significantly improved the simplest model. For infants, inclusion of the sub-category variable did not significantly improve the base model, while for adults it did ($X^2(4) = 24.14$, $p<0.0001$). We therefore upgraded the model for adults to include this additional term:

model <- lmer(MeanGazeX ~ IsTargetRight * IsVerbalShadow * Subcondition + (1|Id))

Finally, for all models, we report Tukey-adjusted *p*-values for contrasts, computed using the R

package *lsmeans* (Lenth, 2016).

Results

Adults

Figure 2 shows the eye-tracking results from the adult participants without verbal

shadowing in this two-place predicate generalization study, while Figure 3 shows this data for

adult participants with verbal shadowing. The x-axis of the figures is trial time (in seconds),

while the y-axis represents gaze location along the horizontal axis of the eye-tracker screen, with

0 representing screen center.

Figure 2 here

Figure 3 here

The figures show that, prior to the anticipatory period, participants in both groups follow the side

that animates, first looking to the left of center and then to the right (recall that this is the fixed

order of presentation of the initial videos for all trials). In the post-anticipation period too,

participants look at the side with the animation – to the right (light thin lines) in target-on-right

trials and to the left (dark thick lines) in target-on-left trials. This is to be expected. However,

critically, there is an anticipatory effect for non-shadowing participants, who look significantly

more to the right in target-on-right trials compared to target-on-left trials in the anticipatory

period, seen here in the series of dots marking bootstrapped significant stretches in the

anticipatory period.

Summarizing the results of the linear model, participants who underwent verbal

shadowing did not show any anticipatory effects, while the participants in the no-shadowing

condition showed significant anticipatory effects. These were driven primarily by the

dog-push-car sortal.

Figure 4 here

In the linear mixed model (see Figure 4), we found only a significant effect of the

particular sub-condition (i.e., whether the target sortal was dog-push-car or car-push-dog), t(85)

=-5.2, p<0.001). Additionally, the intercept was significant, t(86)=2.27, p=0.026. The

sub-condition factor showed an interaction with both the upcoming target location, t(836)=4.68,

p<0.001, and with verbal shadowing, t(82)=2.8, p=0.006, and the three-way interaction was

significant, t(836)=-2.68, p=0.008. Finally, all contrasts comparing target on left/right for the

combination of with/out verbal shadowing and sub-condition, i.e., where the target was

dog-push-car/car-push-dog, with/without shadowing, were non-significant at p>0.85, with the

exception of dog-push-car without shadowing, where the mean gaze position when the upcoming

target was on the right was significantly more right-ward than when the upcoming target was on the left, $p<0.001$.

Infants

Figure 5 shows the eye-tracking results from the 12-24 month-olds with the two-place event sortals. As with the adult data, the x-axis of the figure is trial time (in seconds), while the y-axis represents gaze x-location. Although the infants follow the side of animation, they fail to show significant anticipatory responses; while the bootstrap method identifies nearly the entire post-anticipation time period as showing significant differences between target-on-right and target-on-left trials, these do not extend to the anticipatory period.

Figure 5 here

In the linear model, we only found a significant intercept, $t(39)=4.95$, $p<0.001$, which is readily explained by the design of the experiment: because the order of presentation of the two event sortals was always left, followed by right, followed by the anticipatory period with the still images, we would expect the observed rightward bias of the mean eye-gaze position, particularly in the early part of the anticipatory period.

Discussion

In this experiment, participants had to learn to anticipate the side with a colorful, multimodal animation based on a specific, reversible two-place event sortal, like a dog pushing a car; distinguishing it from its reversed counterpart, a car pushing a dog. The results presented

29

above suggest that neither adults who have their language faculty tied up in a shadowing task, nor 12-24-mo-old infants can successfully recognize such a specific, reversible event sortal across changes in participant identities and other details such as the left/right on-screen location of the event. In contrast, adult participants in the no-shadowing condition were able to anticipate correctly, although this effect was primarily driven by participants in the dog-push-car condition. Before drawing conclusions about the observed pattern of failures and successes, we present our findings from the one-place predicate version of this experiment. Under the hypothesis that a dynamic event sortal corresponding to a one-place predicate might be cognitively easier, we expect greater success in the one-place version. Vonk (2002) asked whether great apes could form concepts of relationships, most pertinently, particular activities like grooming or playing, where the actors changed. The task used was match to sample, with real photographs. Each of the apes (a juvenile gorilla and two mature orangutans) responded well above chance in distinguishing the actions despite changes in actors, even when the species of actor varied. But notice this is less than we demand of the participants who have to notice agent type as well (car versus dog). To our knowledge no work has explored the kind of event sortal describe here in other species, even at the 1PP level.

Importantly for us, Experiment 1 used a novel method, intended to test a hypothesis not previously tested in infants, and from the data, it is not clear whether there is a failure of the method, or whether infants are cognitively incapable of forming the kinds of event sortals we test. The observation that infants generally seem to accurately track the side that animates during the entirety of the trial (see Figure 5) indicates that infants appear to be engaged in the audiovisual features of the experiment, and show a significantly distinct pattern of following the

target animation following the anticipatory period (revealed by the permutation analysis indicated at the bottom of the gaze data in Figure 5).

We again test 12-24-mo-olds and adults with and without verbal shadowing in the one-place version.

Experiment 2: Representing a one-place predicate in infants and adults

One-place event videos and construction of individual trials

In order to keep this experiment as similar as possible to the previous, we decided to keep the two participants on-screen as in the previous experiments. That is, each event always had a dog and a car facing each other as in the previous experiments, but only one of the two types was involved in the dynamic event. Indeed, the "neutral positions" at the first and last frames of each video, and at the frames during the anticipatory period were *identical* to the previous experiments.

This experiment had two different event sortals; we presented participants with pairs of animations corresponding to one-place predicates such as "dog rolling" or "car rolling" where the target event sortal for a given participant was always either the rolling dog or the rolling car. Note that by "rolling" we simply mean that we rotated through 360° and vertically translated the image of the dog or the car, while its partner remained stationary (but see the next paragraph). That is, participants were required to build classes of dynamic events corresponding to the same specific action of the same moving entity (car or dog). For example, for a participant in the dog-roll condition, neither the presence of the dog alone, nor the presence of a rolling action alone would be sufficient to identify the target correctly, since both of the side-by-side videos

have a dog and a car, and both show the rolling motion. The expected event sortal is a combination of the dog and the rolling motion. Successfully identifying such an event sortal and generalizing it to other instances of cars/dogs can then be indexed by significant anticipations to the correct, target side on each trial.

Although only one of the entities rolled or jumped across all trials, we retained the initial little "hop" by both the car and the dog (see previous Methods section) to ensure, again, that a low-level, first-moving-object cue was not sufficient to make the correct generalization (although this is less of an issue in the current design), and to keep the videos as similar to the previous study as possible. Subsequently, only the relevant entity rolled in the time it took for the car or dog to push the other object in the previous experiment. Thus, the timing and time-line of events within each trial for these videos was also identical to those from the previous experiment.

As with the previous experiment, one group of adult participants passively watched the videos, while the other group was simultaneously engaged in verbal shadowing of text.

Participants

For this study we had 17 adults in the no-shadowing condition of which 3 were excluded for poor data quality, and 19 adults in the shadowing condition of which 1 was excluded for poor data quality. We recruited 20 infants for this study (mean age 17m;18d, 12females). Of these, 3 were excluded for poor data quality. For both adults and infants, exclusion criteria and other details of recruitment were identical to the previous study.

Materials and methods

Besides the specific videos used, all other aspects of this study were identical to the previous.

Data analysis

Data analysis used the previously described analytic pipeline.

Results

Adults

Figure 6 shows the eye-tracking results from the adult participants without verbal shadowing in this one-place predicate generalization study, while Figure 7 shows these data for adult participants with verbal shadowing. As in all similar previous figures, the x-axis of the figures is trial time (in seconds), while the y-axis represents gaze location along the horizontal axis of the eye-tracker screen, with 0 representing screen center.

Figure 6 here

Figure 7 here

The figures again show that adult participants in both of these groups show a similar pre- and post-anticipatory period pattern of following the side that animates (compare with Figures 2 & 3).

Crucially, in these one-place predicate groups, there is evidence for significant anticipations towards the correct side, in both the shadowing and non-shadowing groups. This can be seen in the separation between the mean eye-gaze traces for the target-on-right (light, dotted) and the target-on-left (dark, solid) lines in the anticipatory period (box), and the significant stretches marked in the corresponding shuffle analysis results at the bottom of the graph.

Summarizing the results of the linear model for participants in this one-place predicate experiment, participants showed a significant anticipatory effect with or without shadowing, and there was further an effect of the event sortal itself, with the dog-roll event sortal showing a stronger anticipatory effect than the car-roll sortal.

Figure 8 here

In the linear mixed model for the one-place predicate group, (Figure 8), we found a significant effect of the upcoming target location on eye gaze in the anticipatory period, such that the mean gaze position was significantly more right-ward when the upcoming target was on the right, $t(763)=3.43$, $p<0.001$. There was also a significant effect of verbal shadowing, $t(77)=-2.1$, $p=0.04$, and a significant effect of the particular sub-condition (i.e., whether the target event sortal was dog-roll or car-roll), $t(76)=-3.59$, $p<0.001$). The sub-condition factor showed an interaction with both the upcoming target location, $t(759)=4.8$, $p<0.001$, and with verbal shadowing, $t(74)=3.24$, $p=0.002$, and the three-way interaction was significant, $t(758)=-3.47$, $p<0.001$. Finally, all contrasts comparing target on left/right for the combination of with/out

verbal shadowing and sub-condition (i.e., where the target was dog-roll/car-roll, with/without shadowing), were significant at *p<0.001*, with the exception of car-roll in the non-shadowing condition, which was significant at *p=0.015*.

Infants

Figure 9 shows the eye-tracking results from the infant participants. The figure indicates that infants can indeed come to correctly anticipate the target side, when the targets are classes of events that can be described by one-place predicates. The permutation analysis depicted in the figure identifies significant stretches in the first half of the anticipatory period, where the gaze x-position trace in target-on-right trials differs from the target-on-left trials.

Figure 9 here

The linear model found a significant effect of whether the upcoming target was on the right or left, with the mean gaze position more to the right on the former compared to the latter, t(255)=25.14, *p=0.004*. Figure 10 shows the mean gaze x-position during the anticipatory period for both this infant study, with one-place predicates, and the infant study described above, with two-place predicates. Note that, in contrast to the Figures for adult studies (Figures 4,8, and 14 below), Figure 10 does not show a breakdown by sub-conditions (Car-push-Dog/Dog-push-Car or Car-roll/Dog-roll), because we did not find an effect of sub-conditions in our statistical models, as we did with all the adult studies (see the *Data Analysis* section for Experiment 1 aboe).

Figure 10 here

Age-related effects in infants

Given that the age range of our infants is fairly large (12mo – 24mo), we examined the

data from both one- and two-place predicate experiments for any age-related trends. To do this,

we collapsed the anticipatory looking behavior of each infant into a single measure: the

difference between mean gaze x-position in target-on-right trials and the mean gaze x-position in

target-on-left trials. Since successful anticipation is conveyed by more positive (i.e., more

right-ward) values of the mean gaze x-position, we would expect positive values of this

'anticipation effect' metric as indicating successful anticipations.

Figure 11 here

Figure 11 shows the relation between age in months and the anticipatory effect metric. This

metric re-captures the observation that infants perform better with a one-place predicate type

(event sortals) compared to two-place predicate type – the trend line for the former is higher

– i.e., more positive – than that for the latter. However, these trend lines are not significant for

either the one-place predicate type, Pearson's $r = 0.31$, $p=0.19$, or the two-place predicate type,

Pearson's $r = 0.24$, $p=0.29$. We thus fail to find any significant developmental change in this age

range.

Discussion

In this experiment we designed materials that were visually similar to those from the previous experiment, with the exception that, while the former consisted of videos depicting dynamic events that could be described by two-place predicates, the ones in this experiment depicted dynamic events that could be described by one-place predicates. We find that adults can successfully anticipate both the rolling dog and the rolling car, whether simultaneously shadowing text or not. However, we also find a significant 3-way interaction (see Figure 8), driven by a particularly large anticipatory effect – i.e., a larger positive (rightward) mean gaze position in target-on-right trials and a larger and more negative (leftward) mean gaze position in target-on-left trials – for the dog-roll, no-shadowing condition, compared to all the other conditions. Curiously, in the two-place predicate experiment, the largest (and indeed the only significant) anticipatory effect was observed for the dog-push-car, no-shadowing group. This observation suggests the hypothesis that predicates corresponding to event sortals with an animate subject/agent might be privileged in their computation and/or generalization.

Turning to infants, in this condition we do find evidence that 12-24-mo-olds could successfully anticipate the correct upcoming target location, indicating that they can both build classes of events and generalize these over new tokens, and that they can encode and correctly anticipate the target side during the anticipatory period. These results suggest that infants' lack of correct anticipations in the previous experiment cannot easily be attributed to the novel design, nor to infants' cognitive abilities required for categorizing and generalizing dynamic event sortals, nor to making anticipatory eye movements based on their categorization abilities.

However, the nature of the target event sortal, whether with animate subject/agent or not, did not appear to have any effect on the infants' performance. Finally, while we did not find any age-related effects of infants' anticipatory prowess in this or the previous experiment, it should be acknowledged that such age-related effects might indeed exist but not be revealed by our post-hoc analyses.

Taken together, these studies reveal the possibility of an effect of language on the formation of complex concepts (event sortals) that correspond to reversible, two-place predicates: adults undergoing simultaneous verbal shadowing, and 12-24-mo-old infants, are incapable of forming and generalizing such two-place event sortals. In contrast, verbal shadowing does not prevent the formation of comparatively simpler concepts that correspond to one-place predicates, and infants too succeed in this case.

The role of verbal shadowing in these studies is not entirely clear. In part, this is because any primary task is bound to be negatively affected by a concurrent secondary, irrelevant task, just because our cognitive resources are not limitless. Our secondary task, verbal shadowing, involves cognitive elements such as articulation and attention, which can only contribute generally, alongside elements that contribute specifically to disrupt structural processing aspects of language. This secondary task has an independent effect even in the simpler, one-place predicate adult conditions (see Experiment 2, *Adult* results). Since it is the structural, grammatical elements of language that we believe might underlie the building of corresponding conceptual structures, we ran a second version of our adult verbal shadowing experiments, but we minimized structural processing by having participants repeat a sequence of letters and

numbers, rather than shadow *Nineteen Eighty-Four*. We report adults' performance in both the

one-place and the two-place versions of the task with what we call *non-grammatical shadowing*.

Experiment 3: Representing one- and two-place predicates in adults with non-grammatical

shadowing

This experiment is a replication of the verbal shadowing versions of the one- and

two-place predicate adult experiments described above, with the exception of the nature of the

shadowing task. Our hypothesis is that, since this type of shadowing involves only a minimal

engagement of the participants' grammatical system, we expect the participants to be able to

build and generalize both one- and two-place event sortals despite concurrent shadowing.

Participants

We recruited 19 participants for the one-place predicate group, of which none were

excluded. We recruited an additional 18 participants for the two-place predicate group of which 2

were excluded for poor data quality. Exclusion criteria and other details of recruitment were

identical to the previous experiments.

Materials and methods

All aspects of this study were the same as those for the one- and two-place predicate,

adult verbal shadowing conditions reported in Experiments 1 and 2, with the exception of the

shadowing itself. Instead of shadowing a novel, adults repeated the sequence

A…B…C…1…2…3 continuously. As with the previous shadowing experiments, all participants

practiced this non-grammatical shadowing just before beginning the study, and an experimenter

ensured that none of the participants paused for more than 2 seconds.

Data analysis

Data analysis used the previous analytic pipeline. The linear models used were exactly the same as those for the previous verbal shadowing conditions as described above. For the one-place predicate condition, this was also the simplest and the best model, and an improvement over the baseline model, $X^2(2)=13.83$, $p<0.001$, while for the two-place predicate, this was not statistically different from the base model, $X^2(2)=0.6$, $p=0.74$.

Results

Figure 12 shows the eye-tracking results from adult participants with concurrent non-grammatical shadowing in the one-place predicate generalization condition, while Figure 13 shows this data for adult participants in the two-place version. As in all similar previous figures, the x-axis of the figures is trial time (in seconds), while the y-axis represents gaze location along the horizontal axis of the eye-tracker screen, with 0 representing screen center, and dots at the bottom represent regions of significance in the shuffle analysis.

Figure 12 here

Figure 13 here

Figures 12 & 13 show that non-grammatical shadowing does not appear to suppress concept formation and generalization for either one- or two-place predicates. The linear model

for the one-place predicate (see Figure 14A) showed a significant effect of whether the upcoming

target was to the left or to the right, t(397)=3.07, p=0.002, of the sub-condition dog roll/car roll,

t(46)=-2.32, *p*=0.025, a significant interaction between the two, t(396)=3.76, *p*<0.001, and a

significant intercept, t(49)=-2.57, *p*=0.014. The contrasts showed that the mean gaze x-position

during the anticipatory window was more to the right for target-on-right trials compared to

target-on-left trials both for the dog-roll group, t(395)=-9.0, *p*<0.001, and for the car-roll group,

t(397)=-3.06, *p*=0.012.

Figure 14 here

A similar analysis for the two-place predicate version (Figure 14B) found only a

significant effect of whether the upcoming target was to the left or to the right, t(371)=2.24,

p=0.026. The contrasts showed that this was primarily due to the mean gaze x-position during

the anticipatory window being more to the right for target-on-right trials compared to

target-on-left trials for the dog-push-car group, t(372)=-3.12, *p*=0.011; these values for the

car-push-dog group did not reach significance, t(371)=-2.24, *p*=0.12.

To summarize, we find that non-grammatical shadowing allows for detecting and

generalizing the reinforced event sortal, whether it corresponds to the simpler, one-place

predicate, or to the more complex, two-place predicate. As in the previous adult experiments

reported above, we again find an advantage for the predicates/sortals with the animate (dog)

subject/agent. Note that the nature of the interference caused by verbal shadowing is much

disputed, and there is still much work to be done. Newton & de Villiers (2007) asked whether

adults could reason about another's false beliefs (FB) (as opposed to true beliefs) while shadowing a story, compared to an attentionally-matched condition of rhythmic shadowing. Their answer was no, but others have countered with results that suggest rhythmic shadowing can also disrupt FB reasoning (Dungan & Saxe, 2012) and Forgeot-d'Arc and Ramos (2011) found evidence that verbal shadowing disrupted physical reasoning as well as FB. All of these used shadowing of continuous verbal material, as in an audiobook. However, Samuel, Durdevic, Legg, Lurz & Clayton (2019) found verbal "shadowing" of digits did not interfere with false belief reasoning. In that study, the subjects were discouraged from speaking aloud, and judged on their memory for the digits. Thus, there are two problems: is verbal shadowing itself a continuum in terms of the involvement of the language faculty? All shadowing tasks entail decrements in attention, but within language shadowing, is some material more involving of the deeper language faculty, perhaps the grammar, than others? The present study bears out that idea, as the simple non-grammatical shadowing was disruptive, but not destructive, of the adult's ability to form a complex event sortal, but the grammatical shadowing wiped it out. Second there is the question of the extent to which the conceptual task entails language. The FB tasks in the studies above were all different in the degree of perceptual complexity and their response demands, leading to the possibility that not all were equal under interference. The current study supports that too, in that forming a class of the transitive events was not disrupted even by grammatical interference, but the class of intransitive events was not.

General Discussion and Conclusions

In this paper, we aim to extend a prior research program with adults that investigates possible candidate concepts that might be cognitively comparatively simpler and comparatively more complex– simplicity being operationally defined as the resistance to verbal suppression in an online anticipatory categorization task with adult participants. Using a modified version of those tasks that allows us to compare the performance of 12-24-mo-old infants and adults in very similar paradigms, we explore the formation of event sortals that differ in their linguistic descriptions, two-place or one-place predicates. Our results suggest that infants fail to represent event sortals corresponding to a two-place predicate ("dog pushes car") but succeed in a one-place predicate ("dog jumps"). The adult data replicates previous findings, in that two-place event sortals are more vulnerable to verbal shadowing compared to one-place event sortals.

From a technical and design viewpoint, this study presents a unique opportunity to understand both the similarities and differences of the infant and the adult mind, and the role of language in extracting event descriptions that tie together sets of dynamic events despite variations in surface features – what we refer to as event sortals. Several features of the study and the results give us confidence in the pattern of results obtained, and which lead us to our interpretations. First, we have taken great care to ensure that low-level features cannot easily explain the pattern of results. All videos were created using software that allowed us to depict events that were spatially and temporally very well controlled. Thus, event sortals corresponding to dogs pushing cars and those corresponding to cars pushing dogs cannot be readily distinguished based on low-level features. We included a small, simultaneous "hop" by both the pusher and the pushee in each animation to further ensure that the first moving object did not contain information that could be used to distinguish the target side. Between the two-place and

one-place experiments, we kept the timing and gross spatial layout the same, such that the neutral positions of all videos (a dog facing a car) were *identical* across the two experiments.

Second, we find ample evidence that both adults and infants are engaged in our experiments, and that they spontaneously learn to predict the side of the upcoming target event sortal. Evidence for engagement comes from examining the pattern of looking behavior during the pre- and post-anticipatory phases: during these phases, both infants and adults track the side that animates, as seen in the eye-gaze traces (Figures 2, 3, 5, 6, 7, 9, 12, 13). This is particularly clear in the post-anticipatory period, where all groups show significant differences between the target-on-left traces and the target-on-right traces throughout this period, as can be seen by the results of the shuffle analyses in these figures. Evidence for predictions are of course the primary focus of these analyses, and we find that participants in all the conditions with event sortals that can be represented by one-place predicates succeed in spontaneously grasping and generalizing the regularity that underlies the target event sortals for their group.

Third, although not a primary focus of our analysis, and although we do not characterize this in detail, we find that, for both one- and two-place predicate groups across the adult experiments, there is a better anticipatory response for those sub-groups for whom the event sortal had the dog as the subject/agent. Interestingly, previous studies have shown that 2-year-olds were inconsistent in their preference for animate entities in agent roles in acting out reversible transitive sentences (Chapman & Kohn, 1978). In fact, Gelman & Koenig (2001) propose that children only link animacy and agency between the ages of three and five years. Animacy is used as a probabilistic cue to agentivity in adult sentence processing (e.g. Primus,

2012), though Lowder & Gordon (2015) have recently questioned whether it is really animacy or rather more general natural forces.

Fourth, by comparing verbal shadowing of an audiobook, and "shadowing" of simple sequences of letters and numbers, we find that only the former and not the latter appears to impact concept formation and generalization of a two-place predicate event sortal. Given the various shared resources between the two types of shadowing, including presumably various attentional and memory systems like the phonological loop, the results point to a role specifically of the structure-building, grammatical properties of language, which we hypothesize to be critically required for representing such complex concepts. At the very least, a comparison of the one- and two-place predicate adult groups very convincingly demonstrates the much greater ease by which adults grasp and generalize the former, compared to the latter.

However, we must also acknowledge certain shortcomings of these experiments. The most crucial is identifying the period during a trial when participants, whether infants or adults, might make correct anticipatory looks towards the side of the upcoming target. Theoretically, once the participant has figured out the event sortal, then there is sufficient evidence in the very first animation to indicate the correct side. That is, since we always animate the left video first, participants can already judge whether that video matches the target event sortal description; if yes, then they know that the target will be on the left, and if not, on the right. Therefore, the time period during a trial for when anticipatory looks could be made is anywhere from the middle of the first video to the end of the anticipatory period. Indeed, the choice of the duration of the anticipatory period was designed to be on the higher side for infant studies because of the novelty of the paradigm. Similarly, given the novelty of our paradigm, we justify our use of the

entire (for adults) or the first half of (for infants) of the anticipatory period based on the shuffle analysis. Note that for a given trial, the shuffle analysis uses all the data points, while the linear models use only the mean, and in that sense the two analyses do not use the same data twice, although temporal correlations between eye gaze positions across a trial remain a potential concern in terms of the independence of the two analyses.

Finally, we turn to qualifications of the claim that the task entails representation of a reversible transitive proposition to capture the event. The stimulus design rules out most low-level perceptual alternatives. But even with this contrast, the "event sortal" being formed is ambiguous. We took care to note that we do not know how a participant might code the event: is it dog push car, or animal push vehicle, or natural kind do something to artifact? Within this design we could not arrange a contrast where say, a dog pushed a train, or a horse pushed a car instead, or a dog kissed a car, unless we continued for many more trials. Even then, a novel participant would undoubtedly perturb looking times. However, we do know that a subject must register what is encoded in the syntax of all natural languages, namely the causal direction of an event between two possible agents. Similarly, in the intransitive case, the event could be e.g., "dog rolling", or "animal doing action", or less. We did not have contrasts where the action was other than rolling, or the dog was something else. Note also that intransitives can be thought of as modifiers of the subject: a dog rolling is a rolling dog, that is, a subset of dogs. Is it more complex than say, a brown dog? The true scope of both generalizations remains for future work to explore.

To sum up: we now know that infants in the second year of life can recognize individual object sortals on the basic of shape and function, and also recognize kinds of objects (Xu, 2007;

Futó et al, 2010). Furthermore, they understand prelinguistically which objects in a scene are primary and which are inessential to a given event (Gordon, 2003). They can discriminate a change in agency in a transitive action (Golinkoff & Kerr, 1978). Our study shows that they can form a category of *intransitive* events where the actors change but remain members of the same kind. What infants of this age cannot do is form a category of *transitive* events in which the agents change but remain members of a kind.

Many studies remain to clarify this finding. For example, if the task was match-to-sample, with no requirements on memory, would infants or shadowing adults easily match the original transitive event despite the agents changing? If language accompanied the stimuli, would infants succeed in maintaining attention across trials to the correct event?

In a series of papers, Hinzen argues that only a human brain comes equipped with the machinery necessary to create syntactic structures, and hence propositional assertions. He describes how the syntactic operations of external and internal Merge create a new kind of meaning, e.g. (DOG (PUSH CAR)). In other words,

"..there is a kind of meaning that only arises with and within grammatical structures –

grammar is *transformative* for what kinds of meaning there are." (Hinzen, 2017, p. 17)

We argue that the evidence from infants and adults is compatible with a radical claim about the role of language in the formation of a similarity class of events. Reversible transitive events fall under a propositional description, such that the similarity across them is captured by a syntactic string rather than a label. The results of the experiment reported here suggest that participants viewing such events can only see the resemblance that ties them together if they have access to

grammar, both because they are old enough to have such grammatical distinctions (i.e. beyond

24 months old), and without grammatical interference from a secondary task.

References

Abbot-Smith, K., Lieven, E., and Tomasello, M. (2004). Training 2;6-year-olds to produce the transitive construction: The role of frequency, semantic similarity and shared syntactic distribution. *Developmental Science*, *7*(1), 48–55.

Arunachalam S., and Dennis, S. (2018). Semantic detail in the developing verb lexicon: An extension of Naigles and Kako (1993). *Developmental Science*, *22*(1), e12697.

Arunachalam, S. and Waxman, S. R. (2015). Let's see a boy and a balloon: Argument labels and syntactic frame in verb learning. *Language Acquisition*, *22*(2), 117–131.

Arunachalam, S., Escovar, E., Hansen, M. A., and Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes*, *28*(4), 417–425.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences*, *34*(3), 113.

Chapman, R. S., & Kohn, L. L. (1978). Comprehension strategies in two and three year olds: Animate agents or probable events? *Journal of Speech and Hearing Research*, *21*(4), 746-761

de Villiers, J. (2014). What kind of concepts need language? *Language Sciences*, *46*, 100–114.

Dungan, J. and Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cognitive Science, 36*(6), 1148–1156.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81.

Forgeot d'Arc, B. and Ramus, F. (2011). Belief attribution despite verbal interference. The *Quarterly Journal of Experimental Psychology*, *64*(5), 975–990.

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, *117*(1), 1-8.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3-4), 455-479

Gelman, S. A., & Koenig, M. A. (2001). The role of animacy in children's understanding of 'move'. *Journal of Child Language*, *28*(3), 683-701.

Gertner, Y. and Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*(1), 85–94.

Gertner, Y., Fisher, C., and Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691.

Golinkoff, R. M. (1975). Semantic development in infants: The concepts of agent and recipient. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 181-193.

Golinkoff, R. M. and Kerr, J. L. (1978). Infants' Perception of Semantically Defined Action Role Changes in Filmed Events. *Merrill-Palmer Quarterly of Behavior and Development*, *24*(1), 53–61.

Gordon, P. (2003). The origin of argument structure in infant event representations. In *Proceedings of the 28th Annual Boston University Conference on Language Development* (Vol. 1, pp. 189-198). Cascadilla Press.

Hinzen, W. (2017). Reference across pathologies: a new linguistic lens on disorders of thought. *Theoretical Linguistics*, *43*(3-4), 169-232.

.

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., and Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological science*, *28*(11), 1649–1662.

Lenth, R. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33

Leslie, A. M. (1982). The perception of causality in infants. *Perception*, *11*(2), 173–186.

Leslie, A. M. and Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288.

Lidz, J. and Gleitman, L. R. (2004). Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences*, *8*(4), 157–161.

Lidz, J., Gleitman, H., and Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, *87*(3), 151–178.

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85-90.

Mandler, J. (2008) On the birth and growth of concepts. *Philosophical Psychology*, *21*,2, 207-230.

Marcus, G. (2001). *The algebraic mind.* Cambridge, MA: MIT Press.

Muentener, P. and Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374.

Newton, A. M. and de Villiers, J. G. (2007). Thinking while talking: Adults fail nonverbal false-belief reasoning. *Psychological Science*, *18*(7), 574–579.

Nordmeyer, A. E. (2011). To be or not to be: examining the role of language in a concept of negation. BA honors thesis, Smith College.

Nordmeyer, A.E. & de Villiers, J.G. (2019). To be or not to be: Examining the role of language in a concept of negation. In A.K. Goel, C.M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2468-2474). Montreal, QB: Cognitive Science Society.

Perkins, L. (2019). *How Grammars Grow: Argument Structure and the Acquisition of Non-Basic Syntax* (University of Maryland, Doctoral dissertation).

Primus, B. (2012). Animacy, generalized semantic roles, and differential object marking. In *Case, word order and prominence* (pp. 65-90). Springer, Dordrecht.

Samuel, S., Durdevic, K., Legg, E. W., Lurz, R., and Clayton, N. S. (2019). Is Language Required to Represent Others' Mental States? Evidence From Beliefs and Other Representations. *Cognitive Science*, *43*(1), 10.1111/cogs.12710. https://doi.org/10.1111/cogs.12710.

Scott, R. M. and Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, *24*(6), 777–803.

Shultz, T.R. (2012). A constructive neural-network approach to modeling psychological development. *Cognitive Development*, *27*(4), 383–400.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96.

Stavans, M., Li,Y., Wu,D., & Baillargeon, R. (2019) Catastrophic individuation failures in infancy: a new model and predictions. *Psychological Review* 126(2), 196-225.

Surian, L., and Caldi, S. (2010). Infants' individuation of agents and inert objects. *Developmental Science, 13*(1), 143–150.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309-318.

Vonk, J. (2002). Can orangutans and gorillas acquire concepts for social relationships? *International Journal of Comparative Psychology*, *15*(4).

Waxman, S. R., Lidz, J. L., Braun, I. E., and Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95.

Xu, F. (2007) Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, *11*(9), 400-406.

Yuan, S., Fisher, C., and Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399.

Zacks, J.M., and Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21

Figure 1
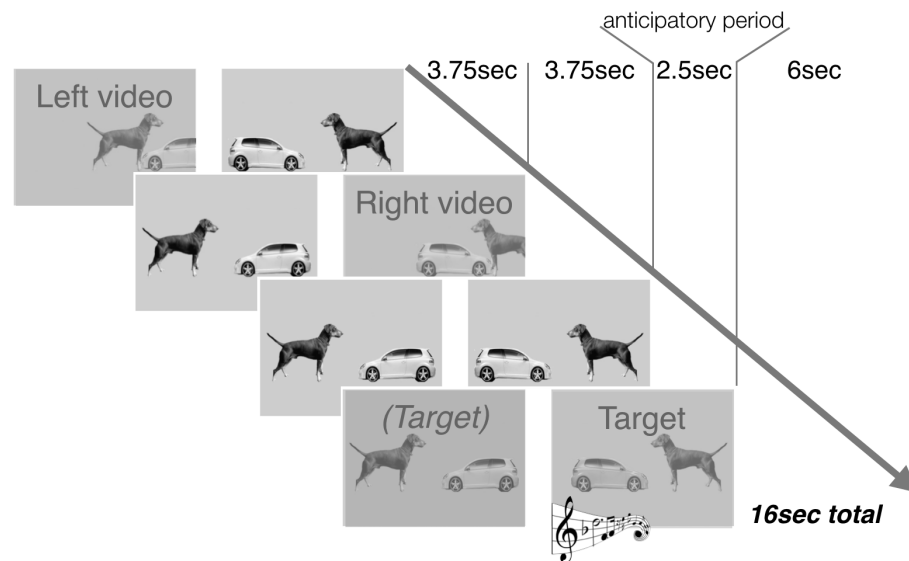
Figure 1. The time-line of an individual trial (a video version of the first trial is available at https://youtu.be/Qwo3OsTyIiA). The four pairs of images are screenshots at different moments during a single trial video. In all trials the left gray-scaled video animates first, followed by the right gray-scaled video. After an anticipatory period, the video depicting the target predicate for that participant animates in color, accompanied by a happy jingle.

Figure 2
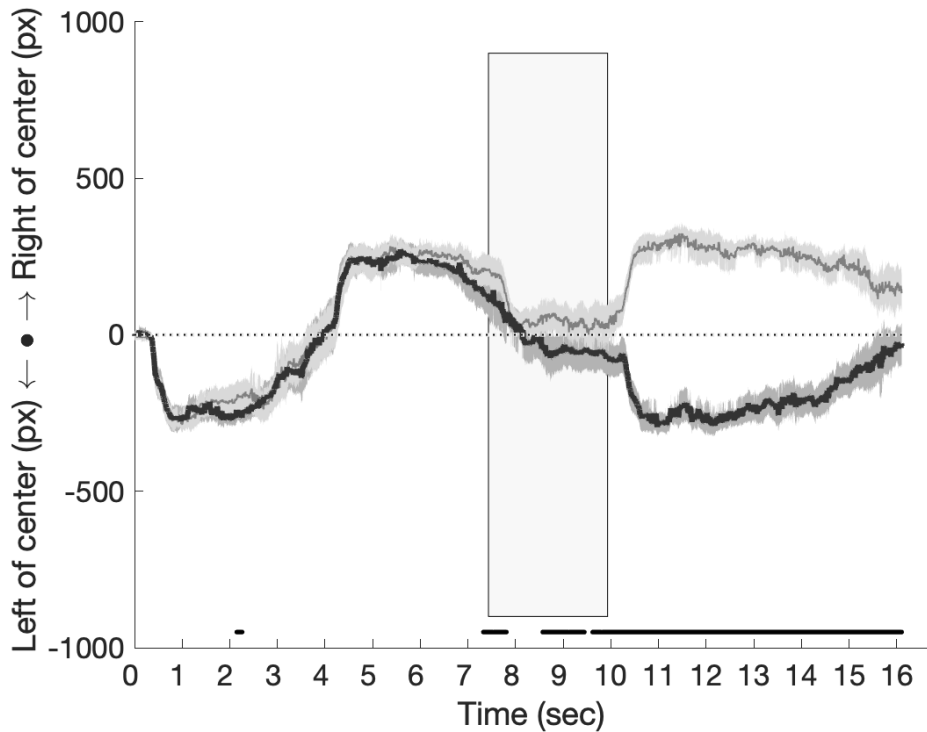
Figure 2. Adult participants (N=18) succeed in generalizing a reversible 2-place predicate in the absence of verbal shadowing. X-axis is trial time (sec) and y-axis is distance of the gaze point from center along the horizontal eyetracker axis, with 0 representing screen center. The lighter, thinner line (the line entirely above zero, i.e., to the right of center, in the post-anticipatory period) is the mean gaze x-position for trials where, following the anticipatory period, the target was on the right, while the darker, thicker line is the corresponding mean gaze x-position for the trials where the target was on the left. The x-extent of the rectangle between ~7sec-10sec marks the anticipatory period. The row of dots (running together into several line segments) at the bottom of the graph mark stretches of time that are significant by the bootstrap method (see *Data analysis* section for Experiment 1).

Figure 3
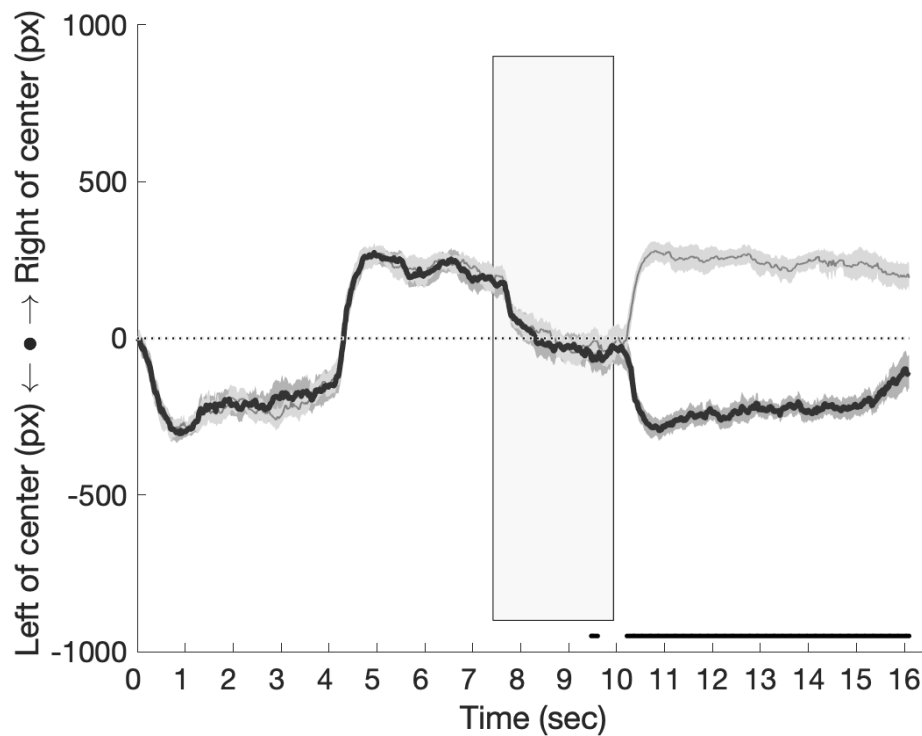
Figure 3: Adult participants(N=20) fail to generalize a reversible 2-place when simultaneously verbal shadowing an audiobook. As in Figure 2, x-axis is trial time (seconds) and y-axis is mean gaze position on screen, with 0 representing screen center. The gaze pattern for target-on-left and target-on-right trials is only significantly different in the post-anticipatory period.
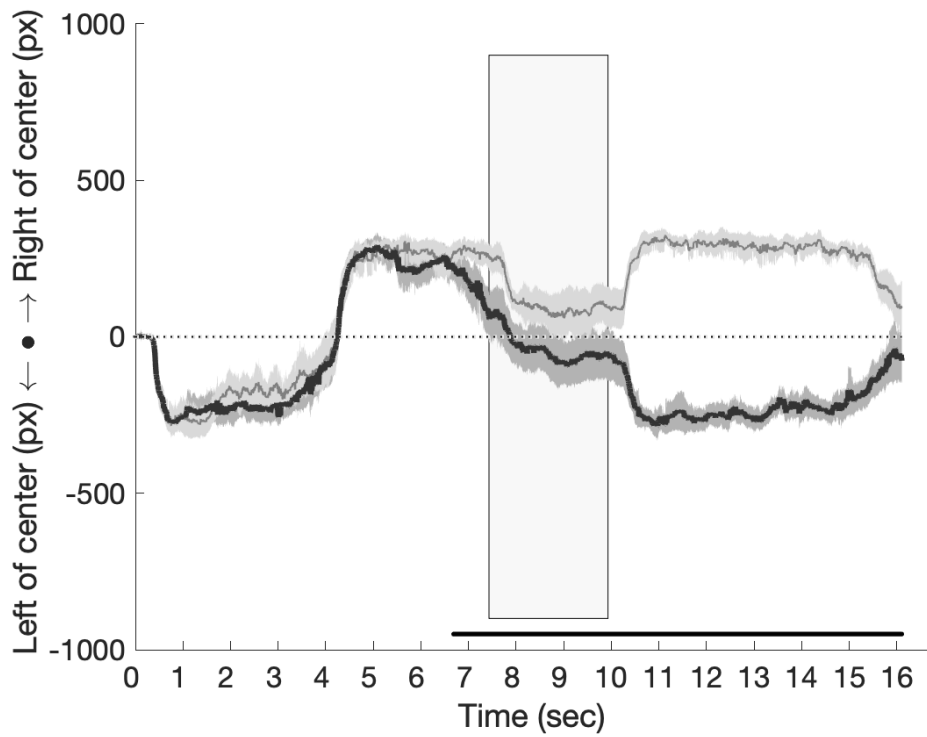
Figure 4
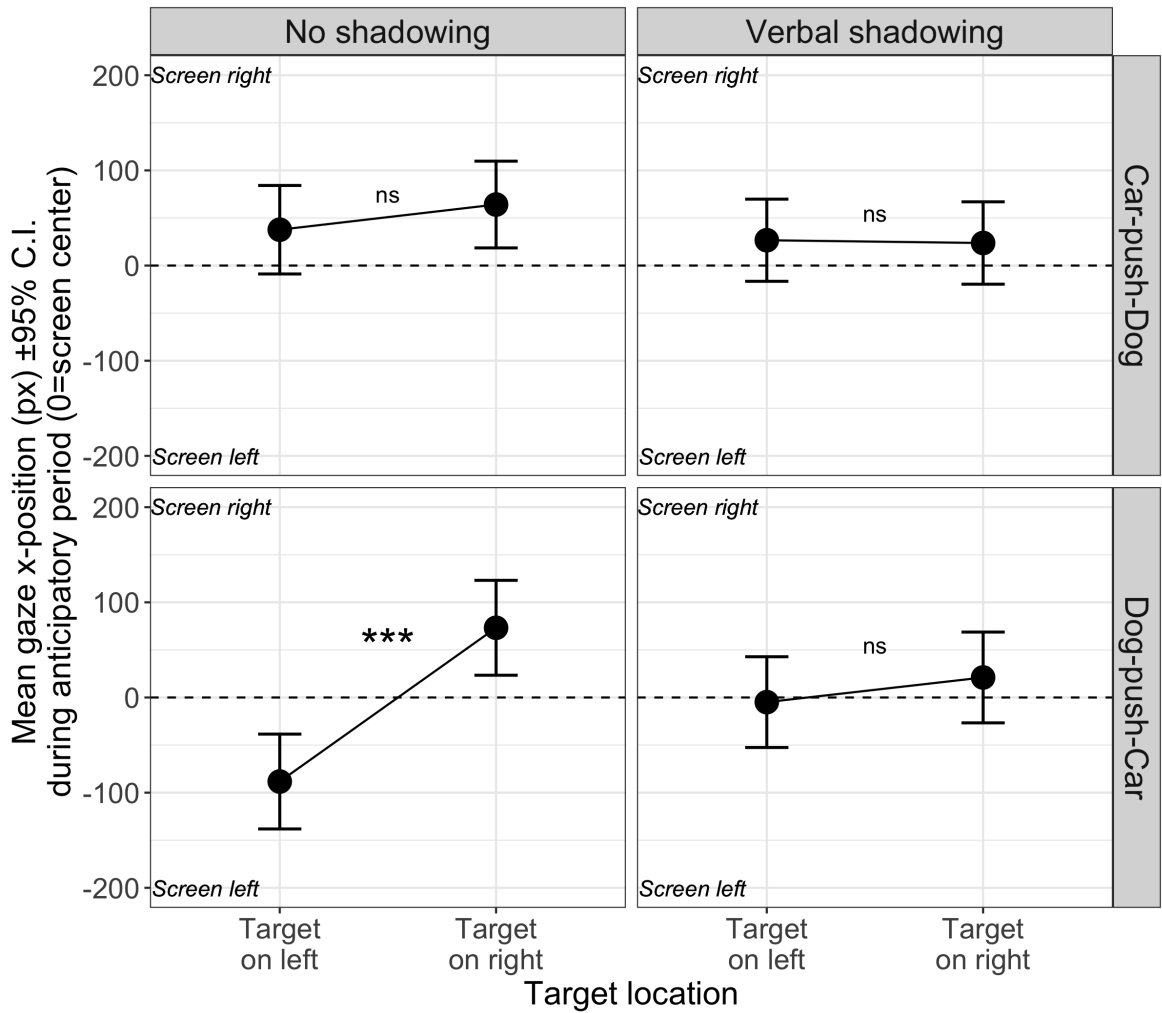
Figure 4. Group-wise results for the two-place predicate experiment with adult participants,

showing mean gaze position (estimated marginal means ±95% Confidence Intervals) during the

anticipatory period for trials when the upcoming target location is to the left or the right on the

screen, for each combination of the specific predicate used (rows) and the presence/absence of

verbal shadowing (columns), An "ideal" anticipation pattern is seen in the Dog-push-Car, No

shadowing group, where the mean gaze x-position during the anticipatory period is significantly

lower (i.e., significantly more left-wards) for the Target on left trials compared to the mean gaze

position during the anticipatory period for the Target on right trials.

Figure 5
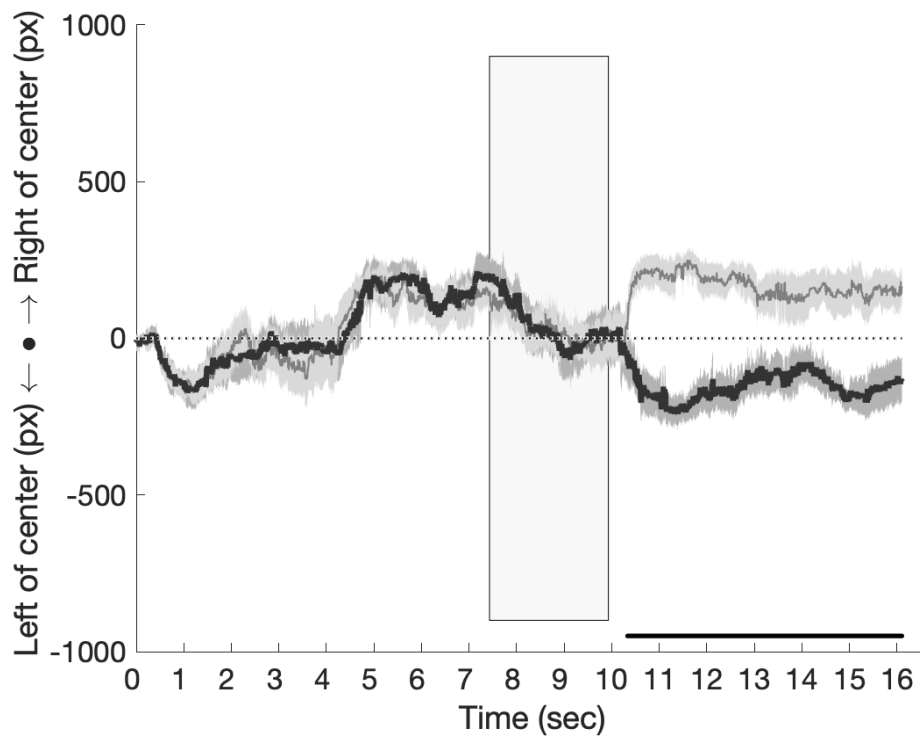
Figure 5: 12-24-mo-old infants (N=20) fail to generalize a reversible two-place predicate. As in the previous figures for adult participants, x-axis is trial time (seconds) and y-axis is mean gaze position on screen, with 0 representing screen center. The gaze pattern for target-on-left and target-on-right trials is only significantly different in the post-anticipatory period.

Figure 6

Figure 6. Adult participants (N=14) succeed in generalizing a one-place predicate in the absence

of verbal shadowing. The shuffle analysis indicates significant difference between the

target-on-left and target-on-right eye gaze patterns starting before the onset of the anticipatory

period (box).

Figure 7
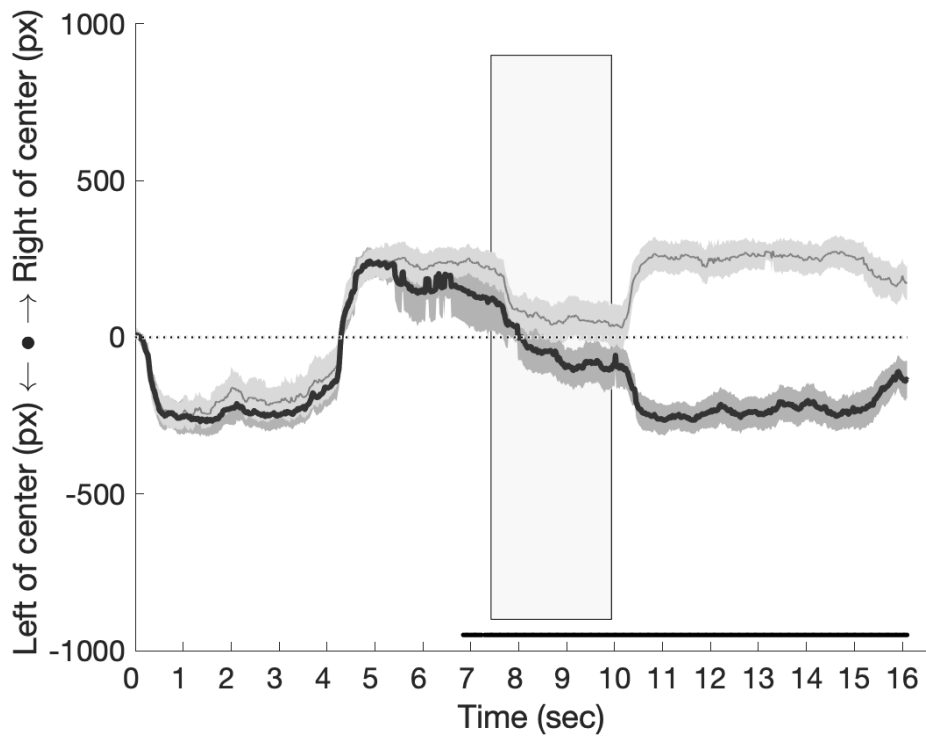
Figure 7: Adult participants (N=18) succeed in generalizing a one-place predicate even when verbally shadowing an audiobook. The shuffle analysis indicates significant difference between the target-on-left and target-on-right eye gaze patterns starting before the onset of the anticipatory period (box), similar to the no-shadowing version of this experiment.

Figure 8
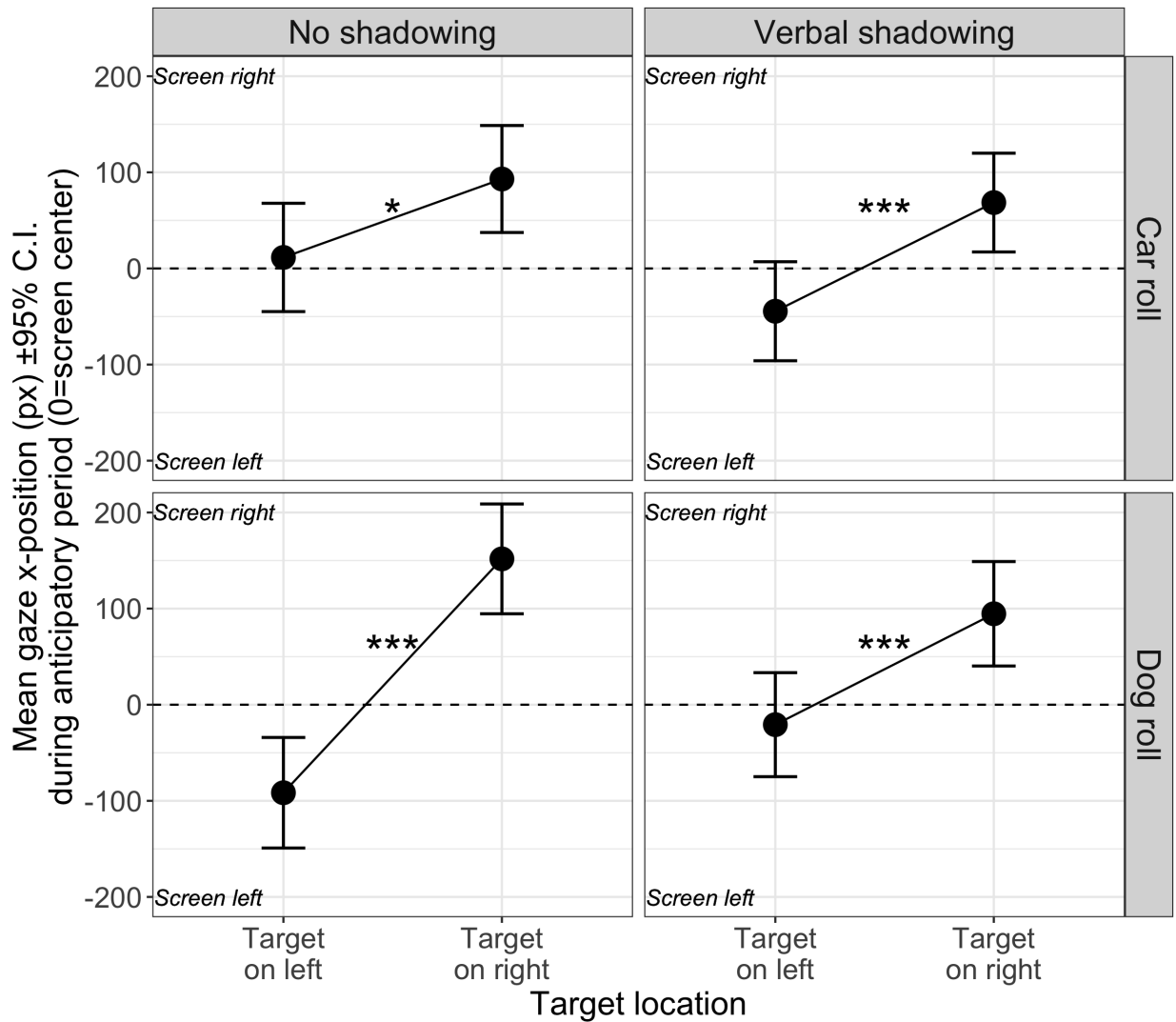
Figure 8. Group-wise results for the one-place predicate experiment with adult participants, showing mean gaze position (estimated marginal means ±95% Confidence Intervals) during the anticipatory period for trials when the upcoming target location is to the left or the right on the screen, for each combination of the specific predicate used (rows) and the presence/absence of verbal shadowing (columns). All cells show significant, correct anticipatory looking patterns: significantly higher mean gaze x-positions during anticipation for the Target on right trials.

Figure 9
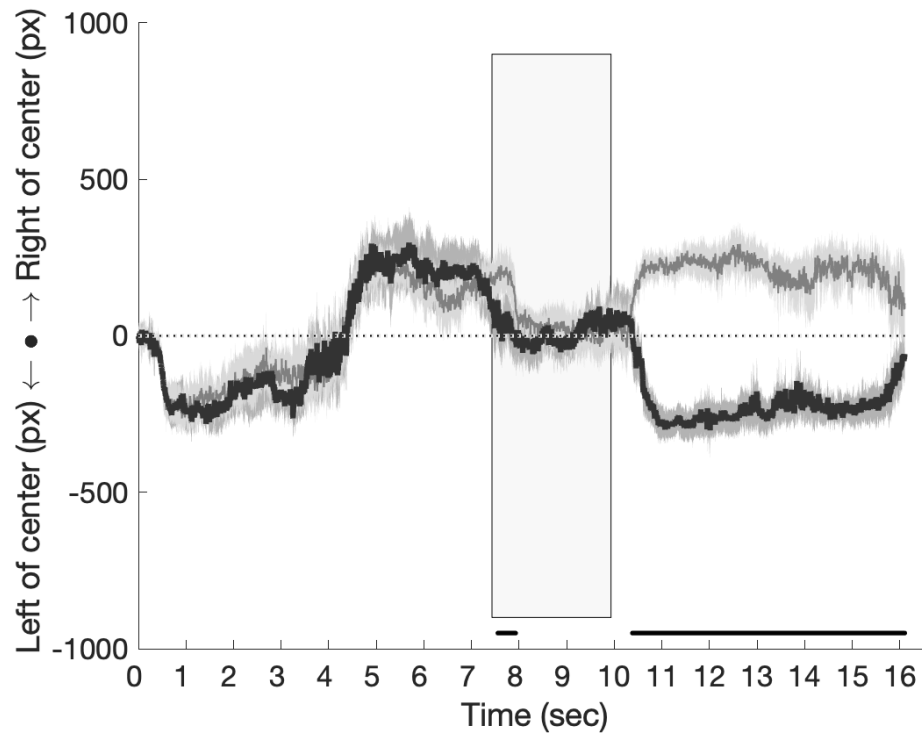
Figure 9. Infants (12-24-mo-olds, N=17) show significant correct anticipations, signaling successful generalization of a one-place predicate. As in comparable figures from the previous experiment, the light and dark lines are the mean gaze x-positions for target-on-right and target-on-left trials respectively, the box marks the anticipatory period, and the row of dots at the bottom mark stretches of time that are significant by the bootstrap method. The anticipatory effect is primarily in the first half of the anticipatory period.

Figure 10
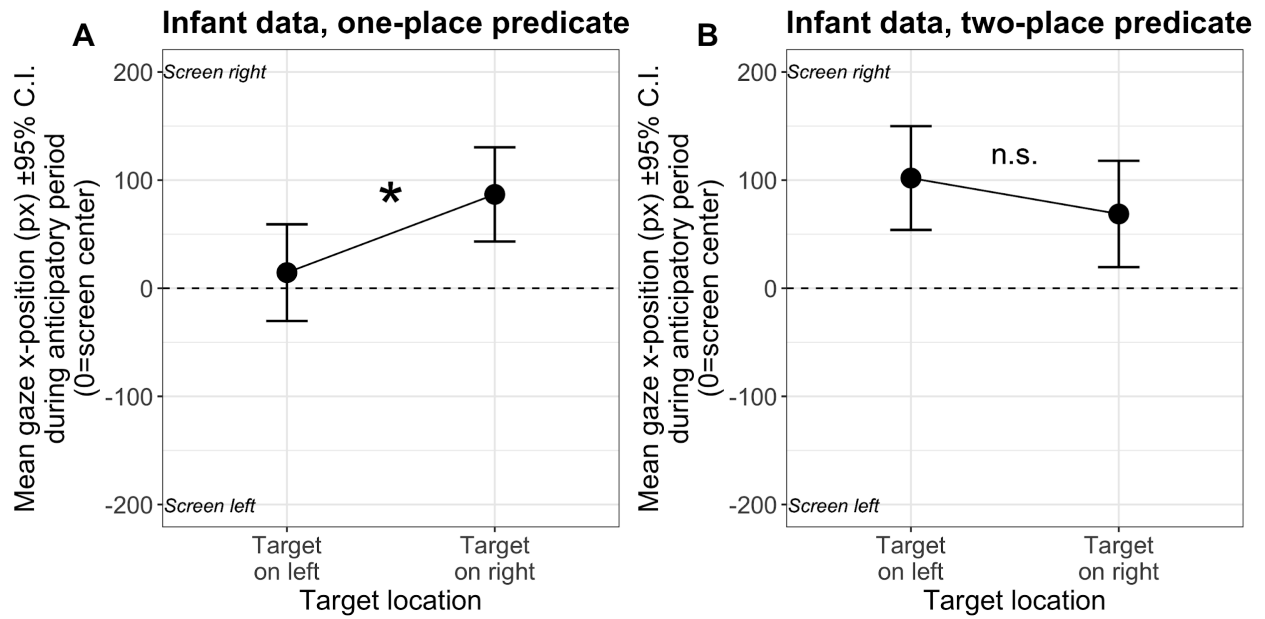
Figure 10. Infant data for the one-place predicate experiment (A) and the two-place predicate experiment (B), shown side-by-side. The mean gaze position during the anticipatory period is significantly higher (i.e., more to the right of the screen) when the upcoming target is on the right, compared to trials when the upcoming target is to the left, only for the "simpler," one-place predicate experiment (A).
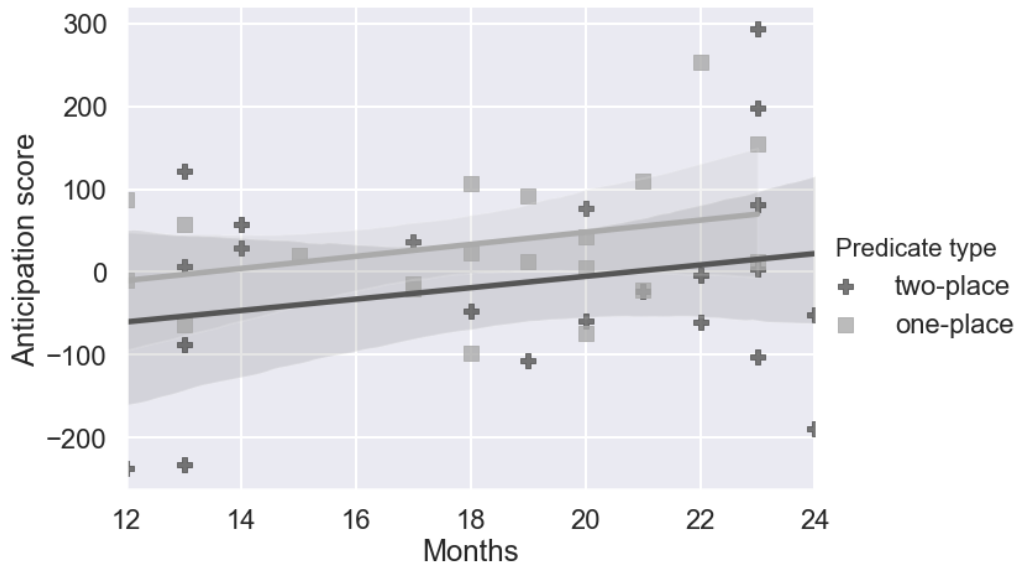
Figure 11

Figure 11. Correlations between age (in months) and an 'Anticipation score' for infants in the one- and two-place predicate experiments. Although both trend lines slope upwards, neither is significant
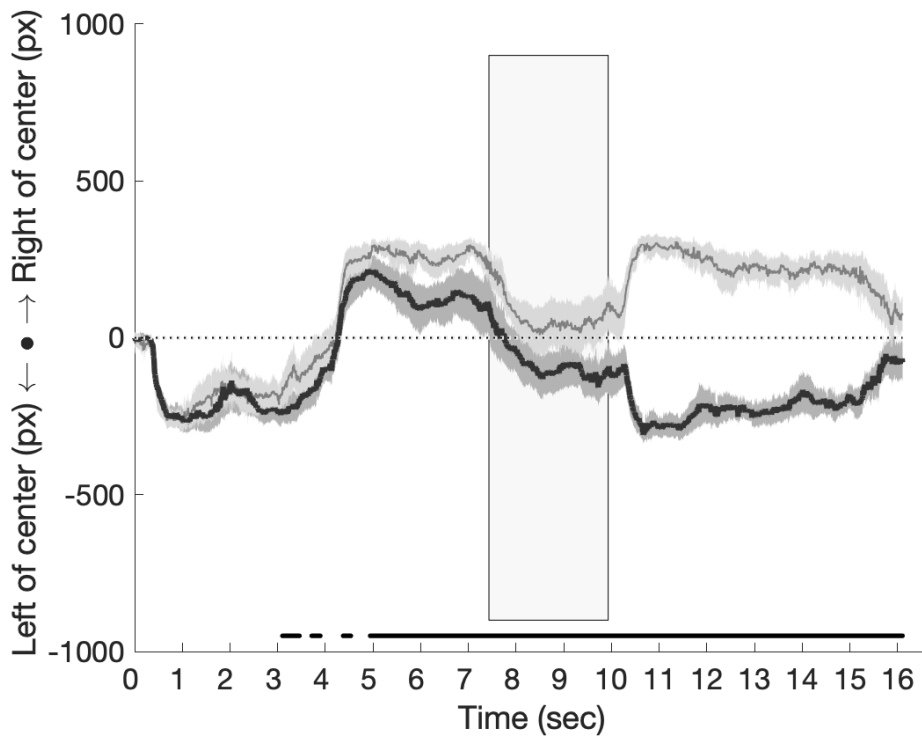
Figure 12

Figure 12. Adult participants (N=19) succeed in generalizing a one-place predicate when simultaneously engaged in non-grammatical shadowing. The shuffle analysis indicates significant difference between the target-on-left and target-on-right eye gaze patterns starting well before the onset of the anticipatory period (box), before the second video even begins to animate.
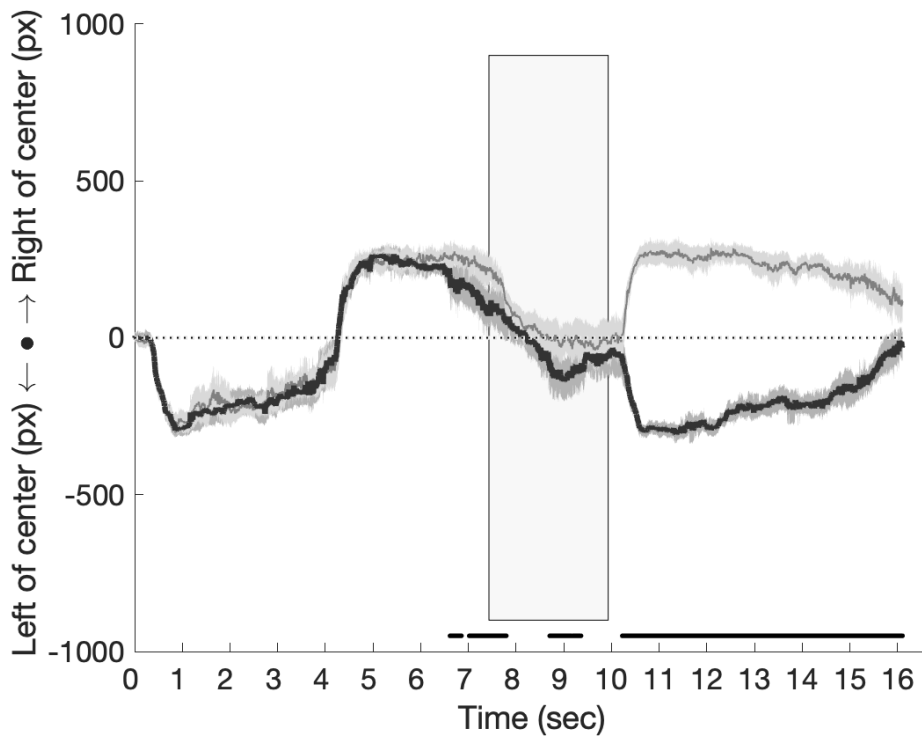
Figure 13

Figure 13 Adult participants (N=16) also succeed in generalizing a two-place predicate when simultaneously engaged in non-grammatical shadowing, although the shuffle analysis indicates a relatively smaller effect, compared to the one-place predicate group (see Figure 11).
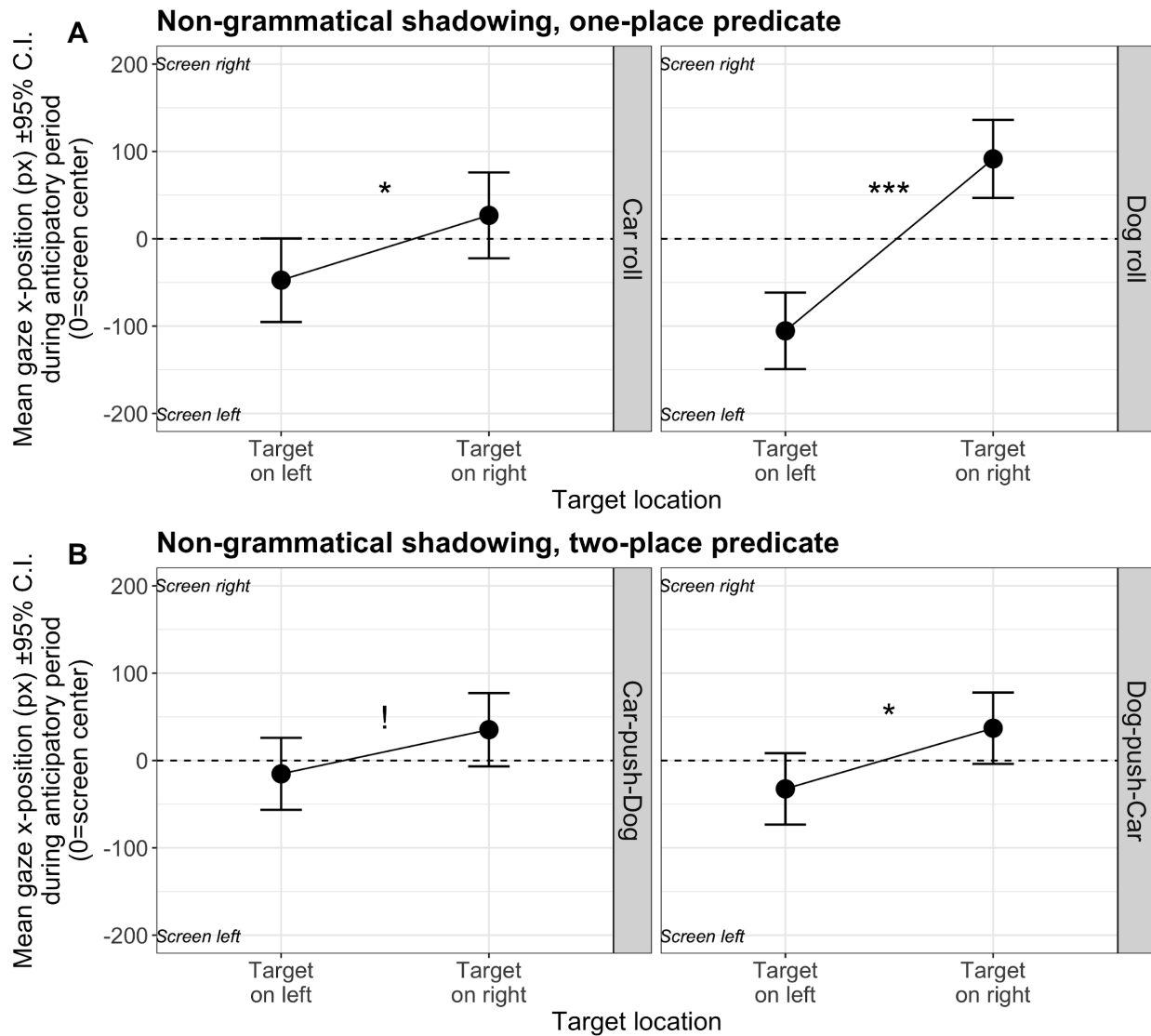
Figure 14

Figure 14. Group-wise results of two-way contrasts, for the one-place (A) and two-place (B) predicate experiments respectively, where adult participants underwent non-grammatical shadowing, showing estimated marginal means (±95% Confidence Intervals) of gaze x-location during the anticipatory period. Both (A) and (B) show an overall anticipation effect, with the combined mean gaze x-positions significantly more to the right for Target on right trials, compared to Target on left trials.