It’s happening!

1. Introduction
2. Approach/trial structure
3. Adults
   1. Predictions
   2. Methods
   3. Results! 🡪 Talk with Jesse about which ones to include; put the others in supplemental materials?
4. Kids
   1. Adaptations to method
   2. Predictions
   3. Methods yey
   4. Results
   5. Discussion
5. Conclusions!
   1. It cool.
   2. Limitations
   3. Future work.
      1. Other languages
      2. Younger kids

Language breaks events into pieces: a sentence like ***Jane***agent***danced***manner ***across***path ***the floor***ground factors a directed motion event into components that can’t easily be separated in space and time. This is accomplished with both words and syntactic structures that pattern across similar events, grouping event components into categories and highlighting connections between them. By understanding the nature of these linguistic concepts we gain a clearer understanding of the thoughts which language encodes, the underlying structure of this system, and the degree to which they might be universal (cf. (Pinker, 1989, 2007). What kinds of event components are encoded in language? How do they account for differences (and similarities) in verb meanings and the syntactic contexts in which verbs appear? These questions have been at the center of linguistic theory and the study of language development (Croft, 2012; Goldberg, 2006; Hale & Keyser, 1993; Halle & Marantz, 1993; Jackendoff, 1983; Levin & Rappaport Hovav, 2005; Van Valin & LaPolla, 1997). Here, we aim to understand how adults and children represent events for language by looking at generalizations that they form in experimental verb learning paradigms. The inspiration for these studies is the parallel work on how children learn nouns, and the role that abstract conceptual and linguistic biases play in this process (cf. Bloom, 2000; Goodman, 1955). Novel noun generalization studies have been central in understanding the scope and development of learning biases (e.g. the shape bias, mutual exclusivity, Landau, Smith, & Jones, 1988; Markman, 1990; Smith et al., 2002; Soja, Carey, & Spelke, 1991; Xu et al., 2009). Initial studies of verb learning tried to import these biases to the new domain, with mixed success (Golinkoff et al. 1996; Kersten & Smith, 2002). Subsequent work has relied on analyses of verb meanings that are rooted in linguistic theory (cf. Fisher, Gertner, Scott, & Yuan, 2010). This proposal unites these two strands.

Here, we test theories of event representations that correspond to abstract generalizations in language, which may both play a role in the processing of individual words and sentences, and act as over-hypotheses to guide the learning of new verbs (cf. Perfors, Tenenbaum, & Wonnacott, 2010). These generalizations capture syntactic regularities both within specific event classes (e.g., manner of motion - MoM – verbs like swimming/rambling- and path of motion – verbs like ascending/crossing) and cutting across event classes (e.g., manner and result as they apply to both motion events and events of object alteration (e.g. *hit/push* vs. *break/melt*). These generalizations can be described at both levels of abstractness – but which of these generalizations actually support human language processing? We build on an existing line of studies from our lab (Kline, de Rechteren van Hemert, & Snedeker, 2016; Shafto, Havasi, & Snedeker, 2014) suggesting the presence of ***highly abstract generalization*** by adults and older children in three directions. First, we extend our work to speakers of Romance languages (which have different patterns of verb lexicalization) to explore the degree to which our findings reflect universal and language-specific expectations about verb meaning. Second, we map the development of these generalizations in early toddlerhood. Finally, we test the impact of higher-level generalizations on verb vocabularies, by providing the first training study to measure the impact of teaching a semantic bias on children’s subsequent learning of verbs in their own language. Together, these studies will provide significant advances in our understanding of the basic processes that facilitate rapid first-language acquisition at the sentence level, and inform theories of semantic structure that aim to explain the richly structured nature of human communication.

## 1. Background: The Linguistic Representation of Events

Over the last 50 years, work in linguistics has revealed a rich system of correspondences between the syntax of an utterance and the conceptualization of the event it conveys. The label under which this work is done varies (e.g., argument structure, distributed morphology or lexical decomposition), and different research communities approach the problem in radically different ways. Jackendoff (1983) conceptualizes the correspondences as mappings between independent semantic structures and syntactic structures (see also Pinker, 1989). In contrast, Hale and Keyser (1993), analyze the same phenomena as arising from an initial syntactic structure that more transparently captures meaning which is then shaped by other constraints on the way to externalization (Halle & Marantz, 1993). Like Jackendoff, theories rooted in the functionalist tradition propose mappings across independent levels of representation, but they typically place a greater emphasis of discourse structure, cross-linguistic variation, imagistic representations, and probabilistic constraints (cf. Croft 2012; Goldberg 2006; van Valin & LaPolla, 1997).

While these differences in perspective are important, and the individual theories make distinct empirical predictions, there is also a remarkable degree of convergence across these traditions (Levin & Rappaport Hovav, 2005). (1) The different theories acknowledge many of the same data patterns and describe them using categories of similar scope (e.g., *“languages syntactically differentiate causal agents from affected entities”*). (2) Most theories break verb meaning, implicitly or explicitly, into a component that is abstract and can predict argument realization and another component (the root) which is more concrete (more closely linked to perception or knowledge specific to that conceptual domain) and which captures differences between the predicates that have the same syntactic behavior (*boil* vs. *bake*). (3) Most theories break the abstract portion of meaning into smaller predicates, and the list of predicates invoked by each camp is remarkably similar (e.g., cause, have, become). All theories with feature (2) must wrestle with the questions of why some verbal roots are more compatible with some predicate structures than others. Are there conceptual classes of roots that are compatible with different structural positions? Or is the perception of fit based solely on prior experience as shaped by communicative needs and the storage of frequently used chunks of structure?

Our work at the interface of language and event concepts focuses on these common properties. We will refer to these event conceptualizations as semantic structures, following Jackendoff, but nothing in the current proposal rides on this description. Our ultimate goals are to look for psychological evidence for the kinds of categories and primitives that are shared across theories of semantic structure, to refine our understanding of these primitives using experimental methods; and, eventually, to understand the relationship of these structures to prelinguistic thought. Many of the features proposed in these analyses (e.g. agent, manner, goal) have also been studied in infants’ understanding of events, especially agents acting to reach an outcome (Leslie, 1984; Luo, Kaufman, & Baillargeon, 2009; Phillips & Wellman, 2005; Woodward, 1998). This raises the intriguing possibility that language development is guided by an early abstract understanding of events (Goksun, Hirsh-Pasek, & Golinkoff, 2010; Hartshorne, Pogue, & Snedeker, 2015; Kline, Snedeker, & Schulz, 2016; Lakusta et al., 2007). The present experiments focus on how these conceptual categories may shape language learning.

The evidence for semantic structures (and for primitives like those above) comes primarily from work in theoretical linguistics that takes as its raw data observed patterns of meaning and grammaticality within and across languages (cf. Dowty, 1991; Jackendoff, 1983; Levin & Rappaport Hovav, 2005; Talmy, 1985). Much of this work has focused on events of one kind--motion events Talmy (1991) breaks motion events into four components: the moving figure, the ground it moves against, the path of motion with respect to that ground and the manner in which the motion occurs (MoM). Speakers, he argues, can encode the path in a particle (*into* or *up*) or in a verb (*enter* or *ascend*). Languages vary in which option they choose. ‘Manner’ languages (e.g., English, Mandarin & Russian) typically pack MoM into the verb, leaving path for an optional prepositional phrase (*He ran into the store*). In contrast, ‘Path’ languages (e.g., Spanish, Greek & Japanese) typically encode path in the verb and fob off manner on an optional gerund (*Él entró en la tienda corriendo*). (But see Beavers, Levin, & Tham (2010) for additional typological variation). This cross-linguistic difference is apparent in corpus analyses and production studies with both children and adults (Aske, 1989; Berman & Slobin, 1994; Jackendoff, 1990).

By breaking up events into primitives like manner and path, Talmy develops a more satisfying theory of event description suggesting that these primitives are "real", in the sense that they, or something like them, exert a force that shapes language. But this linguistic theory leaves two questions unanswered: (1) Are these representations active in the minds of language users? Or do the patterns result from historical or communicative pressures that affect how new lexical items and constructions are generated, spread and die out? Not every property of language is necessarily a property of the individual human mind (cf. Christiansen & Chater, 2015). (2) What form do these representations take? Theorists generally favor descriptions at the broadest level of granularity consistent with the data, but it is unclear if breadth should be our default hypothesis for the psychological constructs underlying patterns of use. In fact many developmental psychologists have argued that learners begin by making narrow generalizations and psycholinguists have proposed that narrower, lexically based categories continue guiding comprehension through the lifespan (cf. Goldberg, 2006; MacDonald, Pearlmutter, & Seidenberg, 1994; Tomasello, 2000). To answer these questions, we must look at how actual language users employ constructs like manner and path during language learning and use.

There is a considerable experimental work on the first of these questions (Are these representations active in the minds of language users?). The central paradigm is one in which participants see a single instance of a motion event (a man running across a road), hear a novel verb (*He is glipping*), and are asked whether the verb applies to two new instances that vary on the critical dimensions (e.g., a man walking across a bridge and a man running around a table). Examples of this generalization task are shown in Fig. 1. Verb extension in adults is shaped by language: speakers of Manner languages are more likely to choose MoMs consistently than speakers of Path languages (Cifuentes-Férez & Gentner, 2006; Maguire et al., 2010; Naigles & Terrazas, 1998; Papafragou & Selimis, 2010). These cross-linguistic differences in lexicalization biases are clearly established by age 5 but appear to be absent in 2 and 3 year olds (ibid. and Hohenstein, 2005; Shafto, Havasi, & Snedeker, 2014).

These methods, however, cannot directly address the second question (What form do these representations take?). Testing single instances rather than the category that contains them, they cannot define the *scope* of the mental category. Maybe English speaking children and adults generalize a verb from running-across to running-around because they classify it as a manner of motion, or maybe they make somewhat narrower generalizations (one about vehicular and one about biological motion) or much narrower ones (one about running, one about walking, etc.). Alternatively, learners might work with even broader concepts than manner of motion (MoM) and path. Talmy notes that particle construction patterns extend beyond motion, suggesting that the relevant distinction is between the manner in which an action is carried out (uniting MoM, manner of harming, etc.) and the result of an action (path, change of state, etc.). This broad distinction has spurred considerable theoretical work (Beavers & Koontz-Garboden, 2012; Jackendoff, 1983; Kiparsky, 1997; Levin & Rappaport Hovav, 2005; Pinker, 1989; Rappaport Hovav & Levin, 1998, 2010; Talmy, 1985). Rappaport Hovav & Levin (2010) propose the manner/result divide reflects a constraint on the kinds of verbal roots that can be lexicalized and the positions they can take in the semantic structure. Specifically, they argue that a single root may refer to the manner (*I kicked the lamp*) or result of an action (*I broke the lamp*), but not both. The experimental work on this question has had the same limitation as the work on MoM and path: it tests people's interpretation of instances of the category but does not require generalization across instances, and thus cannot define the breadth of the mental construct (see e.g., Gropen, Pinker, Hollander, & Goldberg, 1991).

(A) (B)

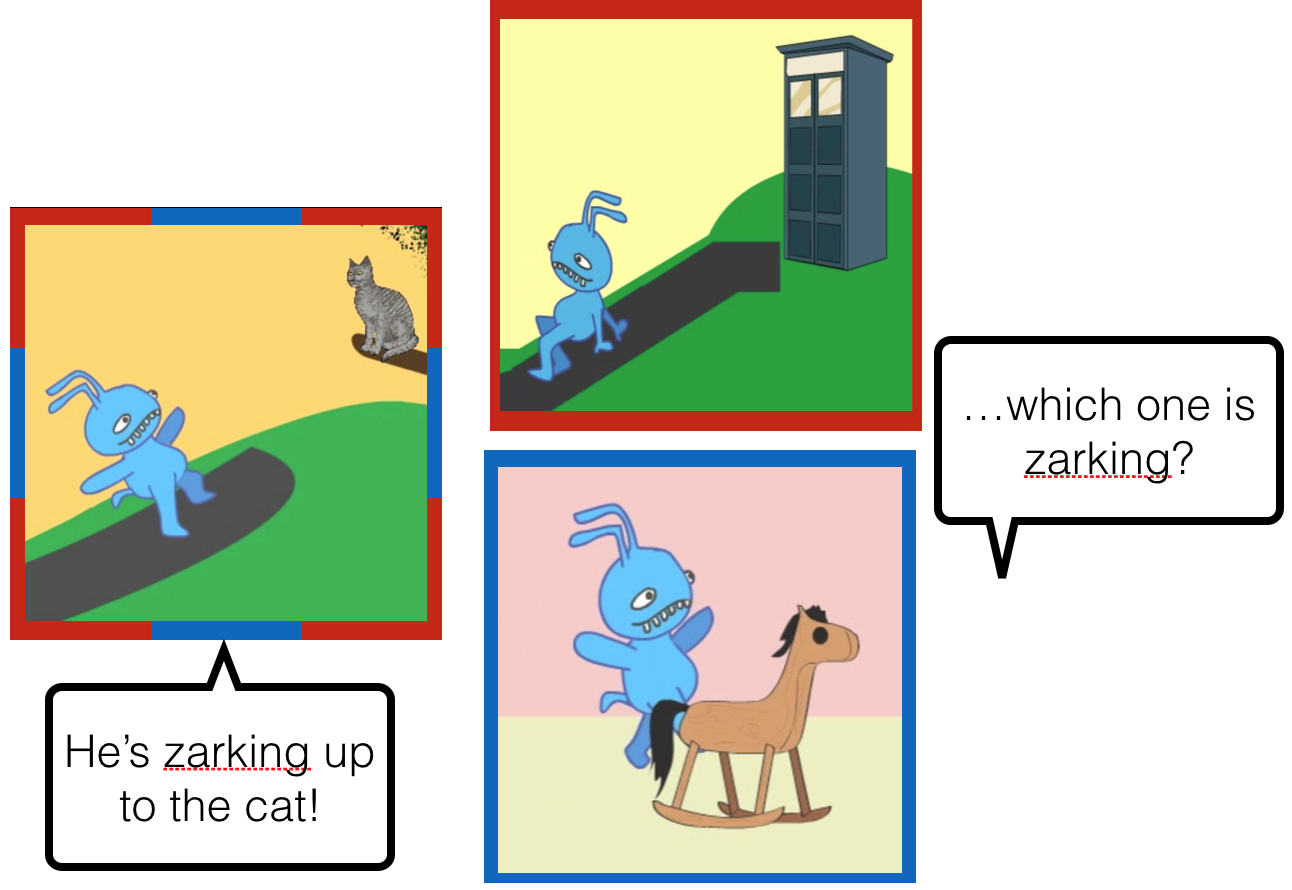
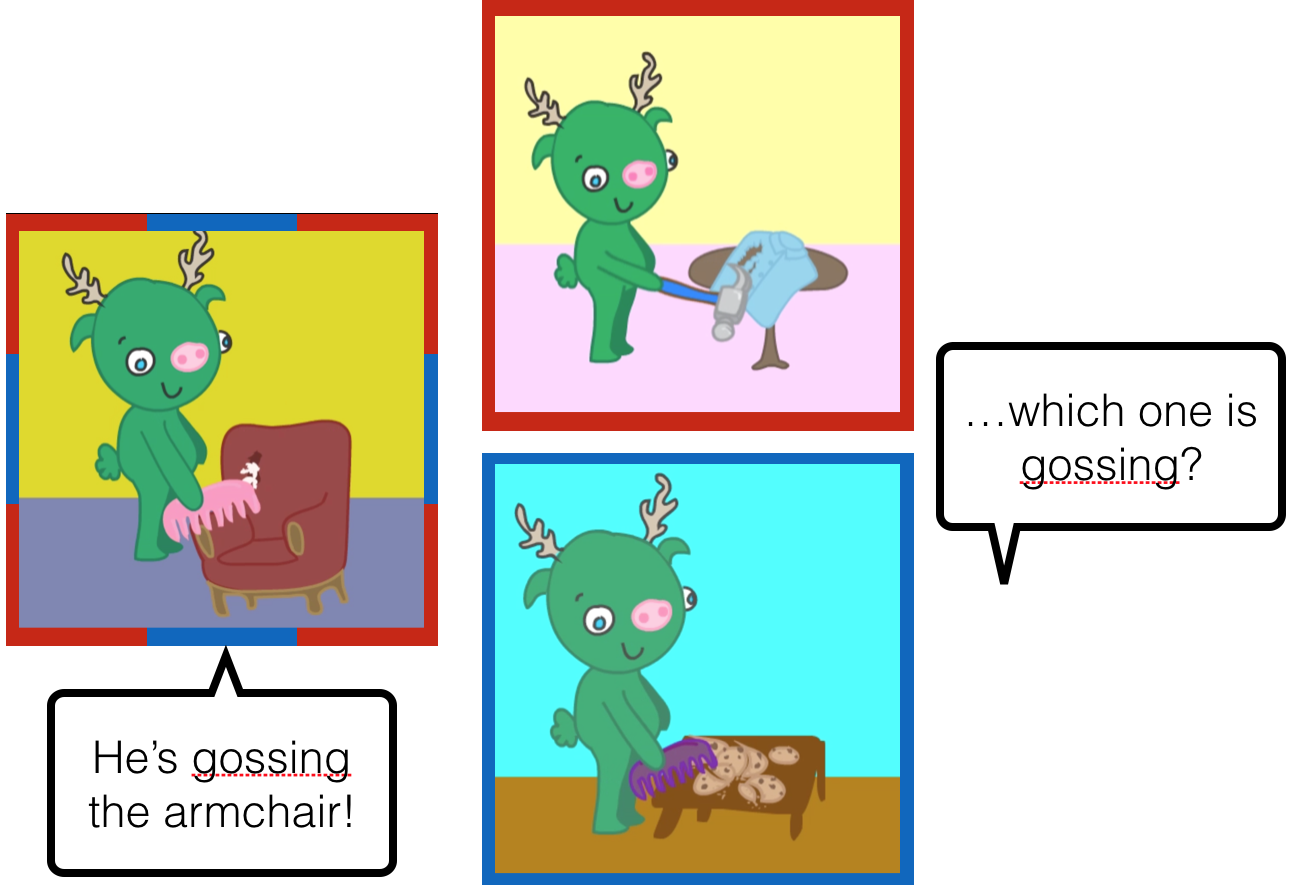
 

Figure 1: Novel verb sets used to study generalization biases. Set A: directed motion scene, ambiguous between ‘ascending’ and ‘arm-flapping’, with path (red) and MoM (blue) extensions. Set B: object-change scene with ‘ripping’ (effect) and ‘combing’ (MoA) extensions. We test whether biases acquired in one domain extend to the other, i.e. generalization from MoM to MoA (blue) or effect to path (red).

Recently we have developed a simple method for exploring the scope of lexicalization biases, and thus studying the grain-size of the concepts encoded in event structures. We present input that leads participants to shift their lexicalization bias, then we assess just how far they will generalize on the basis of this input. Our assumption is that their hypotheses about the nature of the new bias (Is it specific to walking events, a bias in manner of motion or a bias in manner of action?) will reveal the semantic categories that guide verb learning. Our initial studies show that English speaking adults and older children (like theorists) generalize at the broadest level, developing biases for manner and result which they extend readily from one semantic field (motion events) to another (object change events).

Our studies to date have been limited in two ways: they have all been conducted with English speakers from relatively homogenous populations, and we have not explored the development of generalizations in people under the age of four. The first Aim of this proposal is to test speakers of a path language in our bias creation paradigm to determine whether our findings generalize: Will a shift in bias toward MoM, (for Spanish and Portuguese speakers), generalize across event classes in the same way as a shift toward path does for English speakers? We test both adults and 4-5 year olds to explore cross-linguistic differences in development of these constructs, aiming to establish whether these biases reflect deep generalizations in the linguistic system. The remaining Aims trace the development of abstract event categories for language in two ways. First, we test for verb bias learning in young toddlers, using two different methods, in order to establish the trajectory of development. Second, we ask what role these biases may play in language development more broadly: unlike the shape bias for noun learning, verb biases, while a powerful tool for predicting verb meaning in one’s own language, are (notoriously) different across languages. Theorists have proposed that the development of these language specific biases allows for more efficient language acquisition (Gentner & Boroditsky, 2001; Slobin, 1997) but we know of no direct evidence supporting this contention. The proposed work would provide the first experimental test of whether learning English verb biases can affect verb learning in the wild.

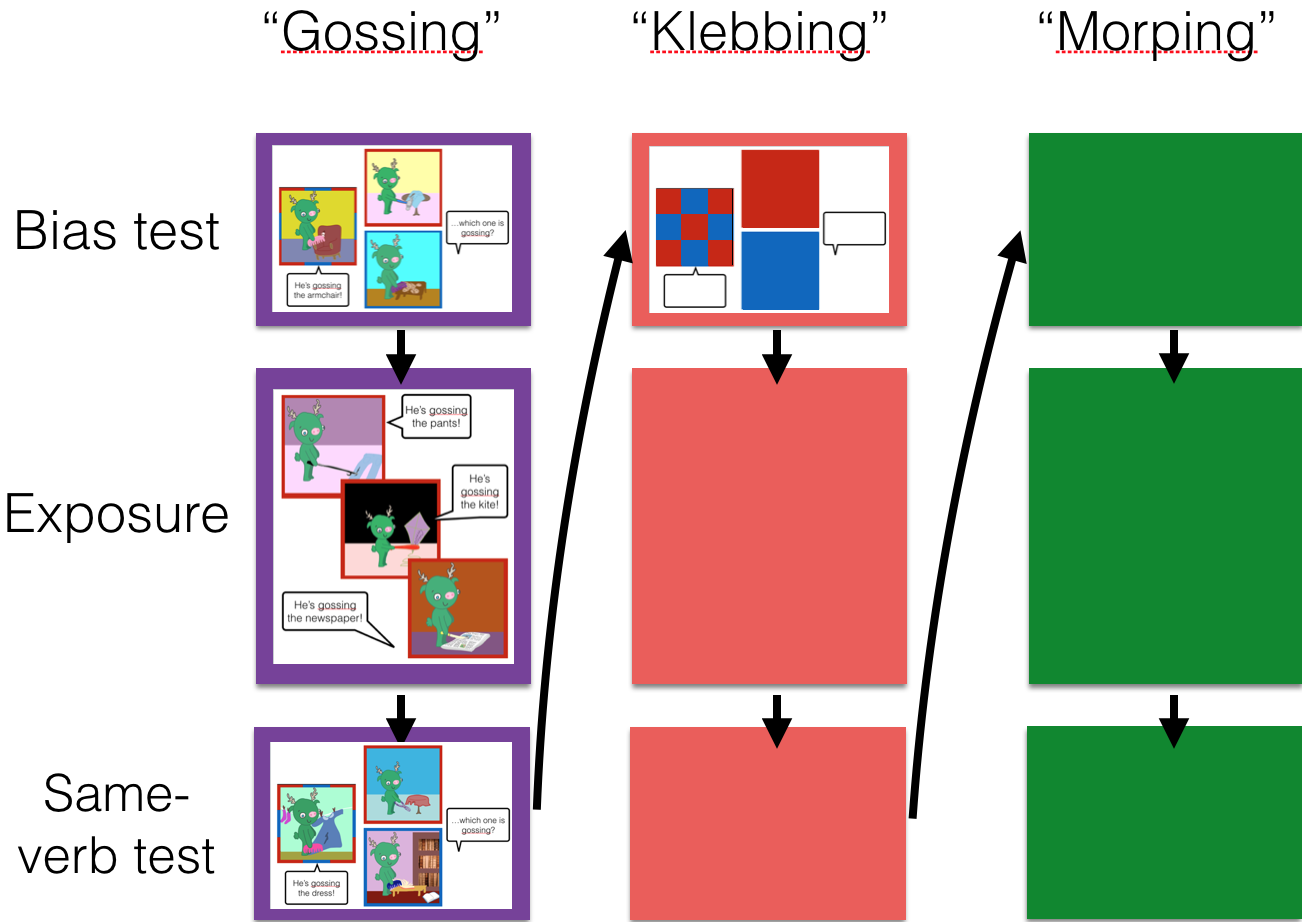
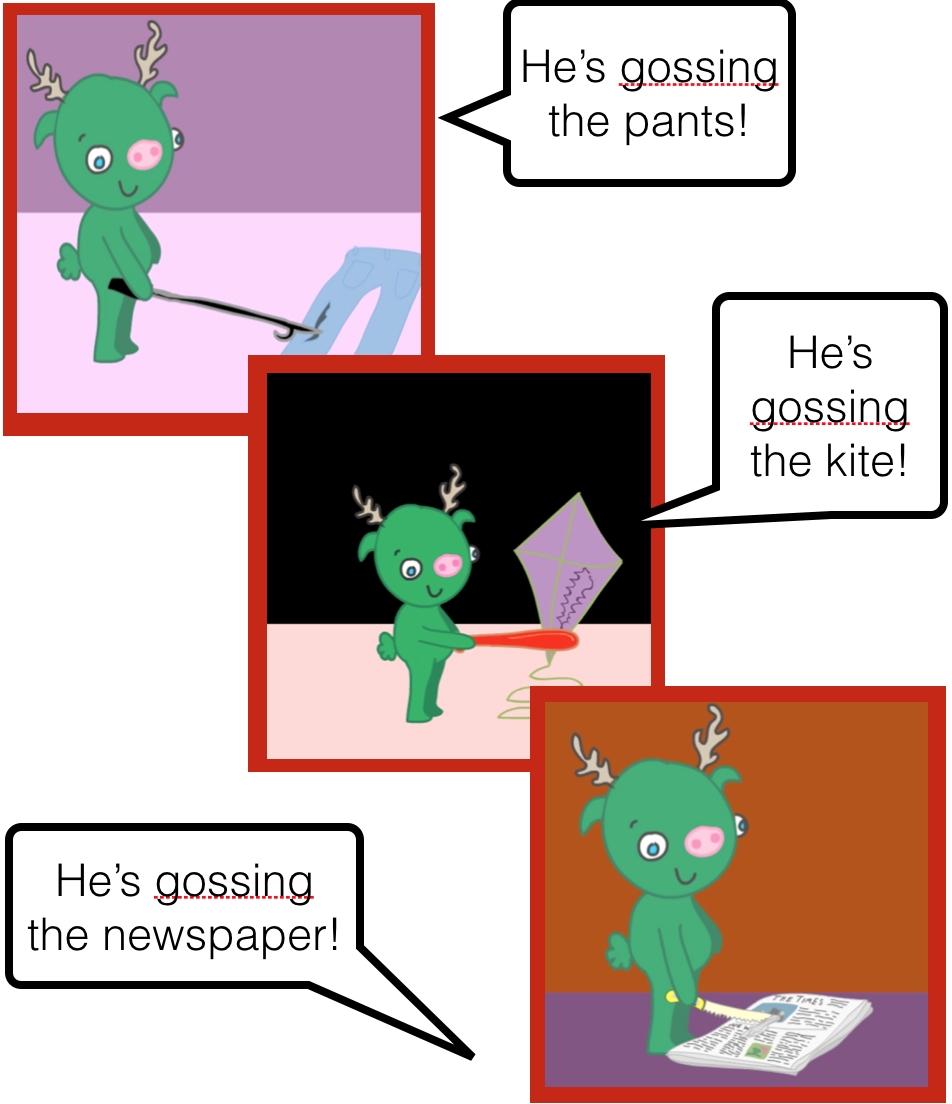


Figure 2: Phase 1 of the experimental design. During Exposure participants learn either manner or result meanings for each verb. We ask whether this affects subsequent Bias test responses.

**2. Approach**

All of the studies in this proposal rely on a common experimental structure, based around multiple instances of single-verb triads like those shown in Fig.1. The presentation and testing of these triads is implemented differently depending on the study population and experimental design, but the basic logic is the same. Here we describe the experiment as it is conducted with adults and older children (the modifications for toddlers are described in Aim 2). The study is composed of a within-domain phase and a cross-domain phase. In the first phase, participants are exposed to eight novel verbs. Each verb is initially introduced as shown in Fig. 1 (reproduced as the top panel of Fig. 2): a single exemplar is presented, and then participants are asked to select between two possible extensions (the *Bias test*). Next, participants see three new exemplars paired with the same verb (*Exposure*, see detail in Fig. 3). This is the only component that varies between conditions: half the participants see videos with a consistent result (e.g., “gossing” = ripping), and the other half see videos with a consistent manner ("gossing" = combing). Finally, participants see a second triad and are asked about the meaning of the verb (*Same-verb test*). This is designed as a manipulation check: older child and adult participants should select the extension of the verb that matched what they have just seen. However, our primary interest is not the learning of individual verbs, but the biases they develop across the study, displayed during the Bias test of each trial. After learning several path verbs (*across, down*), do they guess that the next novel event will be described in terms of its path (e.g. *around*)? In the second half of the experiment, when participants switch to a new domain, we use shorter trials consisting of just the bias test (Fig. 4) so we can measure participants’ expectations about novel verbs in the new semantic field without providing any evidence about individual verb meanings. Next, we describe initial results using this paradigm that test the nature of these second-level generalizations.



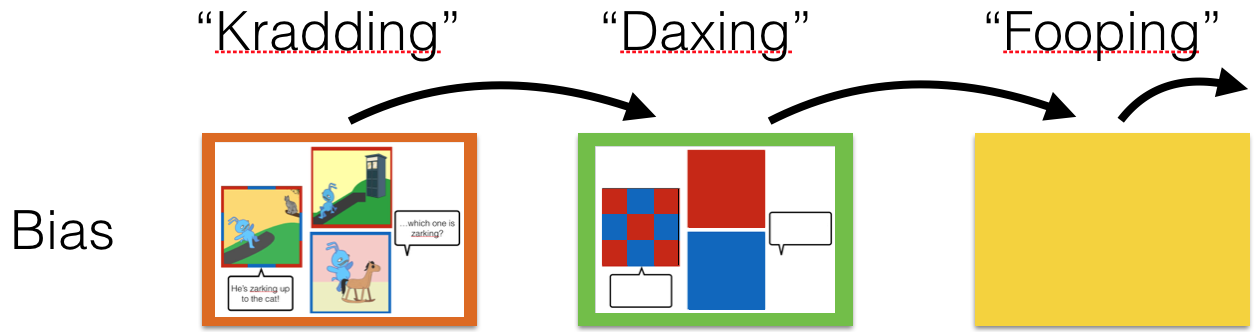


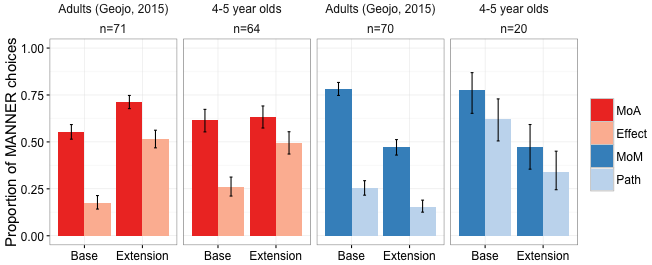
Figure 4: In Phase 2, participants see Bias tests in a new domain (e.g. directed motion); we ask whether biases persist to this domain.

Figure 3: Training sequences that disambiguate the meaning of the novel verb (Object-change actions, Effect condition)

**3. Prior work – The role of manner/path and means/result in verb learning**

In two previous projects, we have used the sequences just described for studying the scope of semantic generalizations that guide verb learning. As the relationship between the initial verbs and the new ones changes from the first phase to the second, we can learn how far a certain population of language users extends this lexicalization bias. This method has produced robust findings in both adults and preschool aged children. Shafto, Havasi & Snedeker (2014) applied this paradigm within the motion domain: participants were taught a series of novel verbs which either turned out to have path meanings (*crossing*, *entering*) or manner meanings (*crawling*, *hopping)*. After learning just a few of these, they began to spontaneously generalize to new verbs (e.g., guessing that a novel word for an event that could be described as either *running* or *ascending* referred to *ascending* in the path condition but to *running* in the MoM condition). Critically, Shafto et al. found that 4-6 year old children were also able to learn these biases within just 3 trials. This required that they not only represent individual manners and paths, but also recognize them as parts of common categories.

In a recent dissertation from our lab, Geojo (2015) explored the hypothesis that lexicalization biases are rooted in the broad distinction between manner and result, rather than the narrower distinctions that are specific to a semantic field, such as MoM or path. To test this she developed the stimulus sets described in Section 2: one vaying the manner and path of motion events; the other varying the manner of action (MoA) and the effect of an object-change event. We reasoned that, if MoM and MoA were part of a single broad category (manner), and path and effect were similarly instances of the broader category, result, then adults who were taught biases within one semantic field should spontaneously generalize them to verbs in the other field, without any additional training. This is precisely what she found. We have preregistered and are extending this finding to 4- and 5-year old English speaking children (Kline, 2016; Kline et al., 2016). This experiment has a 2 x 2 between subjects design. Half of the children learn a series of eight novel object-change verbs and are then asked about novel motion verbs, and half learn motion verbs and then see the object-change verbs. Within each of these conditions, half of the children initially learn verbs that encode manners and half learn verbs that encode results. 32 participants will be included in each of four conditions (initial exposure to MoA, effect, MoM or path, verb meanings) for a total of 128 participants (based on a power calculation from a separate pilot experiment with children).

 Figure 5: Prior work with adult and 4-5 year old English speakers shows more MANNER extensions to new verbs after training on novel manner verbs (manner of action, manner of motion) than result verbs (path, effect). Biases are seen both in the original domain (Base, i.e. additional object-change verbs after MoA/Effect training) and across domains (Extension, i.e. verbs of motion after MoA/Effect training).

Currently, we have finished running participants in the first two conditions (object-change verb learning) and are beginning work on the second (motion). Fig. 5 shows the adult experiments from Geojo (2015) and from ongoing work with 4- and 5-year-old English speakers. Our results suggest abstract generalization at both levels: during the first phase (within-domain) having learned several effect verbs (*rip, open*), children guess that a subsequent entirely novel event will be described in terms of effect (e.g. *bend*). In the second phase (across domain), children-like adults-spontaneously generalize this bias to motion verbs, suggesting that they readily employ broad semantic categories such as manner or result. Pilot data 20 participants (9 MoM, 11 path) suggests we will find similar results in the second half of the study, replicating and extending the findings of Shafto et al., (2014). If this pattern holds up, it would suggest that biases for these two types of events are based on the same underlying conceptual distinction (manner/result). In contrast, if we find an asymmetry (e.g. training on MoA/effect leads to expectations about motion events, but not vice versa), it might suggest that these two domains are hierarchically organized (i.e. that one kind of manner/result representation is a subset of the other).

***Open Questions****:* Our prior work demonstrates that adult and child English speakers quickly adopt new verb lexicalization biases in two semantic fields (motion and object change). These biases transfer readily across domains, indicating that the semantic generalizations are not specific to a given field and lending support to Rappaport Hovav and Levin’s hypothesis that roots come in two basic ontological types (manner and result). These findings raise more questions than they answer. We observe asymmetries in both adults and children between manner and result training conditions: are these the result of stimuli properties, native language effects, or task strategies? Are the generalizations a relatively constrained experimental effect, or a reflection of the basic representational structure of language? Do the biases in our native language affect our ability to acquire new biases or scope of our semantic generalizations? More fundamentally, do children begin with more narrow semantic relationships rooted in the properties of specific predicates, gradually forming generalizations at the level of semantic fields before developing the more abstract adult conceptualization of events? Or are these broad semantic categories an early emerging property of human conceptual structure that guides language acquisition from a young age (cf. Ambridge et al., 2009; Fisher et al., 2010; Pinker, 1989; Tomasello, 2000)?

**Project working title: MannerPathPriming**

**Preregistration June 9, 2016**

**Authors: *Melissa Kline, Annelot de Rechteren van Hemert, Jesse Snedeker***

**Affiliation: *Jesse Snedeker***

This repository contains (so far) information relating to 3 related experiments, referred to throughout at MP-NoExtend, AE-Extend, and MP-Extend. MP-NoExtend was finished as of this preregistration – having made the plans for the second two experiments, we found ourselves in a great position to preregister analyses & methods for the next two studies. AE-Extend data collection has already begun, although at n significantly lower than our intended sample size (see methods). This is therefore a qualified preregistration of AE-Extend and a truer preregistration of MP-Extend.

**A. Hypotheses**

There are two domains, Change-of-state (CoS) and Motion (MOT). Within each domain, there is a pair of abstract perspectives/features of events: Action & Effect (AE) within CoS, and Manner & Path (MP) within MOT. We are asking (1) whether participants (4-5yo typically developing English speakers) can be trained to attend to one of these dimensions within a domain (e.g. either M or P within MOT) and (2) whether the training will extend across domains to the analogous perspectives (Manner & Action – both means of action - predicted to pattern together, Path & Effect – both outcomes of action - predicted to pattern together). Answers are always given as a forced-choice between two verb extensions, maintaining 1 of the 2 features in that domain (e.g. M vs P or A vs E).

MP-NoExtend (already conducted)

This study was designed to replicate Havasi Shafto & Snedeker (2013) Experiment 3, with new stimuli (designed along the same lines, but animated rather than live-action stimuli).

Hypothesis 1: Within the MOT domain, P’s trained to attend to Manner will give more Manner answers than those trained to attend to Path.

This hypothesis was loosely supported by E1: there was a trend in the correct direction with an apparent effect size in line with previous findings, but result not significant and the study was severely underpowered. (See Analysis/MPP\_Analysis.R for all calculations)

AE-Extend (Referred to as E2 in the data sheets)

Hypothesis 2: Within the CoS domain, P’s trained to attend to Action will give more Action answers than Ps trained to attend to Effect.

Hypothesis 3: In phase 2, within the MOT domain, P’s trained to attend to Action will go on to give more Manner answers than P’s trained to attend to Effect.

MP-Extend (Planned, will be E3 in the data sheets)

Hypothesis 4: Within the MOT domain, P’s trained to attend to Manner will give more Manner answers than those trained to attend to Path. (Identical to Hypothesis 1 because the 1st phase of MP-Extend is identical to MP-NoExtend!)

Hypothesis 5: In phase 2, within the CoS domain, P’s trained to attend to Manner will go on to give more Action answers than P’s trained to attend to Path.

Manipulation Checks

All experiments are structured as a series of verbs which are presented and then learned. We are actually interested in the guesses made on **presentation** (ie whether they are affected by learning/exposure to *previous* verbs). However, successful learning of the individual verbs (ie. above-chance performance on selecting the trained answer at the *end of* each verb trial) can serve as a manipulation check that Ps are learning something during the experiment.

Alternate Hypotheses

Previous research and our own experiment so far suggest that 4-5yos are able to generalize within domain – they can learn to generalize from exemplars of paths (e.g. crossing, entering, ascending) to an abstract category of Path, measured by an expectation that a new verb will also refer to the manner of motion (i.e. descending, not skipping.) If this is in fact the case, we expect Hypotheses 2 and 4 to be supported. If the previous results are misleading, we expect to find no differences of training on meaning selection in either experiment.

The next question is whether the categories children learn are domain specific, or if, like adults (Geojo, 2015), they extend them to very different verbs. If they do, we expect H3 and H5 to be supported. However, if children’s categories are abstract but also narrower than adults, we would see H2 and H4 confirmed, but find no differences of training in the second phase (i.e. no difference in conditions for H3 and H5).

**B. Methods**

**Design**

Independent variables

AE-Extend: Training condition (A or E) between-subjects

MP-NoExtend and MP-Extend: Training condition (M or P) between-subjects

Dependent variables

AE-Extend: Binary choice of A-generalizing movie or E-generalizing movie on the bias phase question of verbs 2-8; Binary choice of M-generalizing or P-generalizing movie on verbs 9-16

MP-NoExtend/MP-Extend: Binary choice of M or P generalizing movie on the bias phase question of verbs 2-8; Binary choice of A-generalizing or E-generalizing movie on verbs 9-16

No covariates or moderators

Planned sample

Data will be collected from 4-5yos recruited in Snedeker lab in the usual way.

Details of power analysis are in MPP\_Analysis.R

We used MP-NoExtend to calculate an estimate of the effect size (we have no a priori reason to believe effect size will differ across MP and AE domains.) Both MP-NoExtend and Havasi et al. 2013 (on which it was based) are quite underpowered (16/cell).

We know ourselves to be very underpowered; we don’t have the resources to test the recommended 99/cell, so we will have to hope our estimate of effect size is low. The planned sample size is 32/cell, with an option to extend to 64/cell if results are trending but not conclusive.

Exclusion criteria

Participants will be included so long as they are able to complete the task, we will not attempt to make judgments about whether the child was paying sufficient attention on a given trial. Verb learning (i.e. responses at *test* phase) will serve as a manipulation check at the group level, but we will not use it as a screener for individual inclusion. Decisions to exclude data from analysis will be made at the subject level by MK, as soon as possible after testing and prior to seeing the data from that participant (if MK ran the participant, JS or another lab member will make the judgment), though see below for cases where this judgment may be applied later.

Three conditions warrant exclusion from analysis: accidentally recruited a participant outside our intended sample (e.g. significant developmental delay reported during the session), the child refuses to give answers or otherwise participate in the whole study, the experimenter makes an error that gives the child information about either the novel verb or the movies they are seeing, or the parent/guardian give the child information about the novel verb/which movie to pick. (Parents are asked not to help during the task introduction).

Procedure

The stimuli, script used by the experimenter during the task and the actual code used to present stimuli, including all randomization and counterbalancing, are available in the repository.

Repository/Docs & Info for running/MannerPath\_script.txt

Repository/MPP\_Stim\_and\_Data/MPP.m (Matlab script)

Experimenters are not blind to condition because they must be present in the room to help the child understand the pointing task. To check for fidelity and possible experimenter cueing, we video-record all sessions and will have a blind-to-condition coder (accomplished by putting post-its over the part of the screen showing the laptop that presented the stimuli) rate 10% of the participants’ videos for (1) child ‘checking in’ with the experimenter and (2) experimenter showing a preference for one video before the childs’ choice. If we discover possible cueing problems, we will then blind code the remainder of the sessions and exclude/replace data from sessions with cueing problems.

**C. Analysis plan**

**Confirmatory analyses**

Our planned confirmatory analyses are all included (though not all are implemented) in Analysis/MPP\_Analysis.R

To summarize, we have one between-participants variable, Condition; participants give 7 binary responses in Phase 1 and 8 binary responses in Phase 2 (A/E or M/P). We will use mixed-effects logistic regression with random effects of items in two models. First, we can ask whether Condition affects the choices participants made (ie. whether training caused them to make more M choices rather than P choices), and second we will ask if there is an effect of the trial number; that is whether differences increase over time.

Missing data is handled automatically by this method because we enter individual responses rather than scores into the regression – a child might not have an observation at trial 6, but that’s ok.

We do not plan to do any data transformations other than calculating ‘average correct’ scores based on the responses children give at the *end* of each trial. To try and deal with our low power this score will provide a secondary way to select a subset of participants who are especially attentive/on the ball and see if the expected effects described above hold in this group.

These analyses assume only (?) that the binary data will not turn out to be totally at floor/ceiling, which based on pilot data is very very unlikely to happen!

**Answer the following final questions:**

Has data collection begun for this project?

* Yes, data collection is underway. In the MPP\_Data.csv spreadsheet, E1 represents the initial study (MP-NoExtend) that was used to calculate effect size estimates, and E2 represents the ongoing AE-Extend study (described above) which is the 1st of the two preregistered experiments begun here.

If data collection has begun, have you looked at the data?

* Yes. During MP-NoExtend (replication of Havasi, Shafto & Snedeker 2013) we realized that larger sample sizes than expected would be needed, so we decided to move directly to the longer and more interesting ‘extension’ studies rather than beginning with the within-domain version. We therefore developed the code for the extension version and began testing children on the new version as soon as we could (interleaved with collecting the rest of the informally planned 32 total MP-NoExtend participants). There were no designated ‘pilot’ participants for the beginning of AE-Extend, but there were several participants who did not complete the study because of bugs in our code.
* Because the effect size was smaller than expected, we knew that the intended subject size for AE-Extend would be larger than 32 at least. Power analysis and ‘pre’registration were not conducted until after subject 19 was collected, thanks to the researchers (MK) not getting to it until then (she acknowledges that this is not a good reason).
* The planned analyses were decided on during the MP-NoExtend pilot phase and after ~12 AE-Extend participants had been run. We did not try any additional models because the glmer with max. random effects -> anova comparsion of models is a standard analysis formula I am using at this point. We can now preregister, though, that we intend to continue using this method of analysis for AE-Extend and MP-Extend
* Because of these issues, MK would consider this preregistration to certify against analysis/garden-of-forking-path issues, but it does NOT qualify as a preregistration preventing the ‘file drawer’ problem, because we already have some indications that AE-Extend is panning out the way we expected, and can’t certify we’d be doing this preregistration in the alternate world where that wasn’t the case.

The (estimated) start and end dates for this project are: Aiming to finish data collection by the end of Summer 2016