Gavagai is as gavagai does:

Learning nouns and verbs from cross-situational statistics

Padraic Monaghan[1], Karen Mattock[2], Rob Davies[1], & Alastair C. Smith[3]

[1]Centre for Research in Human Development and Learning, Department of Psychology, Lancaster University, UK

[2]The MARCS Institute, University of Western Sydney. Australia

[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Key words: language acquisition, cross-situational learning, noun learning, verb learning, symbol grounding

Running head: cross-situational noun and verb learning

Corresponding author:

Padraic Monaghan

Department of Psychology

Lancaster University

Lancaster, UK

E-mail: p.monaghan@lancaster.ac.uk

Tel: +44 1524 593813

Fax: +44 1524 593744

Abstract

Learning to map words onto their referents is difficult, because there are multiple possibilities for forming these mappings. Cross-situational learning studies have shown that word-object mappings can be learned across multiple situations, as can verbs when presented in a syntactic context. However, these previous studies have presented either nouns or verbs in ambiguous contexts and thus bypass much of the complexity of multiple grammatical categories in speech. We show that noun word-learning in adults is robust when objects are moving, and that verbs can also be learned from similar scenes without additional syntactic information. Furthermore, we show that both nouns and verbs can be acquired simultaneously, thus resolving category-level as well as individual word level ambiguity. However, nouns were learned more quickly than verbs, and we discuss this in light of previous studies investigating the noun advantage in word learning.

1. Introduction

One of the great difficulties for learning words is the potentially infinite number of possibilities for mapping between a word and potential referents, which has become known as the "Gavagai" problem (Quine, 1960). Child-directed speech generally comprises utterances containing multiple words, along with multiple potential objects in the child's environment to which any of those words may refer (Yu & Ballard, 2007). Some of the difficulty of resolving the word-referent problem was shown to be at least partially resolved by use of "cross-situational" statistics by the learner (Horst, Scott, & Pollard, 2010; Siskind, 1996; Smith, Smith, & Blythe, 2011; Smith & Yu, 2008; Yu & Smith, 2007). For example, when several words and several objects were simultaneously presented to a language learner, it was not possible to learn the mapping between individual words and objects in a single learning instance, but over multiple learning instances both adults (Yu & Smith, 2007) and infants (Smith & Yu, 2008) were able to determine that certain words and objects always co-occurred. Cross-situational statistics applies not only to learning nouns. Childers and Paik (2009) tested 2-3 year old children's ability to learn extension for novel verbs. They found that multiple presentations of events with different objects preserving the action or outcome of the action, resulted in effective learning. Relatedly, Scott and Fisher (2012), in a paradigm similar to that of Smith and Yu (2008) where two verbs and two actions were always present, found that 2.5 year old children were able to learn by exploiting cross-situational statistics.

However, these previous studies have addressed only one aspect of the

complexity of mapping speech onto referents. In the noun-learning studies (e.g., Smith and Yu, 2008), only objects are in view, and in the verb-learning studies, verbs are presented in a syntactic context which disambiguates the category of the referent to be identified (Childers, 2011; Gillette et al., 1999; Scott & Fisher, 2012). But child-directed speech is replete with words from multiple grammatical categories (Mintz, 2006; Monaghan & Mattock, 2012; Yu & Ballard, 2007), and this may have a substantial effect on cross-situational learning, not least because constraints such as mutual exclusivity no longer apply (Monaghan & Mattock, 2012), but also because the possibilities for mappings increase rapidly as more modalities of referent are considered. Hence, in the "Gavagai" problem, an object's motion needs to be disregarded in order to learn the object's name. The first aim of our study was to test whether the learning of word-object mappings from cross-situational statistics remains robust to the introduction of the additional complexity of multiple grammatical categories tested in an adult population.

Verbs are less frequent than nouns in early language production across all studied languages (Bornstein et al., 2004; Childers & Tomasello, 2002; Gentner, 1982; Goldin-Meadow, Seligman, & Gelman, 1976; Imai, Li, Haryu, Okada, Hirsh-Pasek, Golinkoff et al. (2008); Oviatt, 1980; Schwartz & Leonard, 1984; Tomasello & Akhtar, 1995). One contribution to this delay may be that verb-motion mappings cannot be learned with the same cross-situational statistics that apply to acquisition of noun-object mappings (Fisher, Hall, Rakowitz, & Gleitman, 1994; Gillette, Gleitman, Gleitman, & Lederer, 1999; Golinkoff, Hirsh-Pasek, & Mervis, 1995; Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996). For instance, Gleitman

(1990) noted that cross-situational statistics are not able to disambiguate many verb-motion mappings, because it is not possible to distinguish the act of "giving" from "receiving" just from the motion itself (though this is not an issue for intransitive verbs). Verb-referent mappings may thus require additional knowledge about syntactic information within the utterance (Childers, Heard, Ring, Pai, & Sallquist, 2012; Gleitman, Cassidy, Nappa, Papagragou, & Trueswell, 2005), potentially in the form of prior acquisition of nouns and function words to constrain and define the syntactic frame within which the verb is presented (Fisher et al., 1994; Gillette et al., 1999).

Previous cross-linguistic studies of verb learning have presented verbs embedded within syntactic contexts, e.g., "look I'm going to <verb> it" (Childers, 2011), or "she's <verb>ing" (Scott & Fisher, 2012). Such syntactic cues are beneficial for adults learning verbs from scenes (Snedeker & Gleitman, 2004). They also reduce the referential uncertainty – as the frame constrains the mapping from the novel word to apply only to the motion, whereas if the noun and the verb are both unknown then the construction of a mapping becomes manifoldly more complicated. The second aim of our study was thus to determine whether cross-situational statistics are sufficient alone for learning verb-motion mappings. In this regard, we wanted to test whether the conditions that support noun and verb learning are similar (Waxman, Lidz, Braun, & Lavin, 2009), or whether precursors to verb learning, such as syntactic contextual information, are required before cross-situational constraints can be applied (Gentner, 1982; Kersten & Smith, 2002). Relatedly, the third aim of our study was to test whether nouns and verbs could be

acquired simultaneously, or whether learning of nouns in the utterance is a prerequisite to constrain verb mappings.

Other explanations for why verbs tend to be learned later than nouns draw on suggestions that, conceptually, referents for verbs are less coherent than nouns (Gentner, 1982; Gentner & Boroditsky, 2001; Gillette et al., 1999; Golinkoff et al., 1996; Tomasello & Kruger, 1992), or that verbs tend not to be uttered simultaneously with the motion being observed (Gillette et al., 1999), or that the distributional information present within language for learning verbs is less reliable than that for nouns (Gleitman, 1990; Monaghan, Christiansen, & Chater, 2007). Previous studies that directly compared the acquisition of noun-object and verb-motion mappings have tended not to perfectly equalise the information present in the learning environment for directly testing how names for objects and motions are learned (Oviatt, 1980; Schwartz & Leonard, 1984; Tomasello & Akhtar, 1995).

An exception was Childers and Tomasello (2002, 2003), who found a noun learning advantage over verb learning for 2 ½ year old children, when the variability of objects and actions, frequency and distributional properties were controlled for nouns and verbs, though individual tokens of verbs were less frequent than for the nouns (frequency was controlled for lemmas rather than wordforms). In our experimental design, we control a number of factors that have been suggested to contribute to the noun learning advantage, for instance, the frequency, distributional information, simultaneity of occurrence of word and referent, and the number of potential referents in the environment (both objects and motions). We also, critically, tested adults' learning to minimise the proposed effect of conceptual

acquisition of nouns and verbs on the noun learning advantage (see Gillette et al., 1999, for a similar rationale). A final aim of our study was to test whether the noun learning advantage is still observed for adults when several of these key conditions have been controlled.

In our study, we directly compared learning of noun-object pairings, verb-motion pairings, and learning of both noun and verb pairings simultaneously, using an identical cross-situational learning task and environment in each case. We predicted that, consistent with previous studies of noun learning, word-object mappings would be learned from cross-situational statistics even when the utterances and the observed dynamic scenes were more complex than in previous studies. We also predicted that it would be possible to learn verbs from the same cross-situational statistics applied to nouns even without prior learning of nouns, and further that when nouns and verbs could be learned from the same utterance then cross-situational statistics would be sufficiently powerful to demonstrate learning for both these grammatical categories. Finally, we hypothesised that, if the noun learning advantage was due to external informational factors such as distributional complexity, frequency, or simultaneity of occurrence, then there should be no difference in learning mappings from nouns or verbs to their referents. If, however, other factors make verbs are harder to learn, then we should still observe an advantage for learning nouns over verbs.

2. Method

2.1. Participants

48 students at Lancaster University volunteered to participate for course credit or for payment of £3. There were 23 males and 25 females with mean age 20.8 years (SD = 3.4), and all reported speaking English as their first language and had normal vision and hearing.

2.2. Materials

For the objects we used 8 geometric shapes in black on a grey background, as used in Fiser and Aslin's (2002) study. Each shape was viewed in a path of motion. The motion paths were programmed in a form of visual basic within e-prime and represented 8 distinct motions. In a pilot study we presented 19 different motions to 28 participants who did not take part in the main study. We selected 8 of the motions that were classified using coherent terms by the majority of participants and that did not overlap with descriptions of any other motions by the whole group. The motions selected were: *bouncing*, *growing*, *hiding*, *rising*, *shaking*, *shrinking*, *spinning,* and *swinging*. The words were recorded by a female native-British English speaker instructed to produce the words in a monotone. There were two categories of words: "function words" which were monosyllabic, and "content words" which were bisyllabic. The words are shown in Appendix 1. The monosyllabic words were 500ms in duration, and the bisyllabic words were 900ms. The function words and content words were distinguished in length to make them analogous to general

differences observed in natural languages. However, we took care to avoid the function words being directly related to any particular function word in participants' native language – for instance, for the "tha" function words, the "th" was an unvoiced dental fricative as in "thin", a sound that does not occur in the onset of English function words (as in "the"). Ten of the 16 bisyllabic words had first syllable stress, and none contained morphological cues.

For each condition, the heard utterance referred to one of the viewed scenes. In the noun-only condition, eight object pictures were paired with a bisyllabic word. This meant that whenever the object word was heard by the participant its referent object was one of the two moving objects on the screen. The remaining eight bisyllabic words were randomly permuted over the learning situations, and so did not occur reliably with any of the objects or the motions. In the verb-only condition, eight of the words were paired with each of the motions, such that when the word was heard the target motion was one of the two motions observed on the screen, and the remaining eight bisyllabic words did not relate to any of the objects or motions. In the noun-and-verb condition, eight of the bisyllabic words were paired with an object each and the remaining eight bisyllabic words were paired with a motion. The occurrence of particular objects and motions were randomly permuted, so that there was no association between any object and motion (nor between any noun and verb). The word-referent pairings were randomly generated for each participant to avoid any potential preferences in linking certain sounds to shapes or motions. The probabilities of co-occurrences of nouns, verbs, non-referring words, objects, and motions are shown in Table 1.

---------- Insert Table 1 about here ……

2.3. Procedure

At each learning trial, participants observed two different objects undergoing a different motion. After three seconds, the participant heard a sentence over headphones, composed of a function word followed by a bisyllabic word followed by the other function word followed by another bisyllabic word. The function words provided distributional information in the task such that they indicated the role of the word they preceded. Thus, in the noun-only condition, one of the function words always preceded the object referring word and the other function word preceded the non-referring words. In the verb-only condition, one of the function words preceded the motion-referring word and the other preceded the non-referring word. In the noun-and-verb condition, one of the function words preceded the object referring word and the other function word preceded the motion referring word. We selected function words to precede both nouns and verbs in the speech in order to control the distributional information for nouns and verbs. Though function words (determiners) only reliably precede nouns in English, other high-frequency words reliably precede verbs in English (e.g., pronouns, prepositions, conjunctions), and so expectations about high-frequency words marking context are likely to apply equally to both nouns and verbs in the language (see e.g., Monaghan et al., 2007).

After the sentence had finished, the participants were instructed to indicate whether the scene on the left or the right of the screen was described by the sentence by pressing either the “1” or the “2” key. After a pause of 500ms, the next trial began. An example trial is shown in Fig. 1.

---------- Insert Figure 1 about here ……

There were 12 blocks of training each containing 24 learning trials. Within each block, each motion occurred six times, and each word occurred three times. Pairings of bisyllabic words were balanced for frequency across the experiment, and as far as possible within each block. Pairings of particular motions with particular words were also balanced for frequency across the experiment and within each block. The order of the bisyllabic words (referring and non-referring for the noun-only and verb-only conditions, and the order of object referring and motion referring for the noun-and-verb condition) within utterances was randomised and whether the left or right scene contained the target object and/or motion was also randomised. Participants were able to pause to rest after six blocks of training.

For the noun-and-verb condition it is possible to learn to solve the task by responding either only to the noun-object pairing or to the verb-object pairing. In order to determine whether participants had learned the nouns or the verbs or both in this condition, we conducted an additional test at the end of training. To test verb learning, participants were presented with two scenes containing a neutral object performing different motions, heard the single motion-referring word, and were

required to select the scene described by the word. This task could only be solved if participants had learned the verb-motion mapping. To test noun learning, participants viewed two stationary objects and heard the single object-referring word. The correct object could only be selected if participants had learned the nouns. One trial for each verb and each noun was presented, and no feedback was given on performance. After completing the test, it was immediately repeated to increase power for analysis.

3. Results

Our analysis modelled the probability (log odds) of response accuracy, considering variation across participants and materials, as well as the effect of block (to measure learning across the experiment), the effect of experimental condition, to determine whether certain conditions resulted in better learning, and the interaction between block and experimental condition. We took into account the fact that observations were clustered for each participant and for each stimulus action or object, by performing a series of Generalized Linear Mixed-effects Models (Baayen, 2008; Jaeger, 2008). The models were initially fitted specifying just random effects to account for variation by participants and stimuli in overall accuracy (random intercepts). We then added the experimental condition (noun only, verb only, noun-and-verb), block (1-12) and the interaction between condition and block as fixed effects. Likelihood ratio test comparisons showed that inclusion

of the interaction term significantly improved model fit compared to a model including just main effects of condition and block ($\chi(2)^2 = 44.7$, $p = 2 \times 10^{-10}$).

Following the recommendations of Barr, Levy, Scheepers, and Tily (2013; see also Baayen, 2008), we fitted further models adding both random intercepts and random slopes for the random effects, to measure individual variation in participants and stimuli changing accuracy over the blocks. A likelihood ratio test comparison of models showed that a model with both random intercepts and slopes (with correlations between participant random effects only – correlations for stimulus random effects were found to be invalid) fit the data better than a model with just random intercepts ($\chi(4)^2 = 399.0$, $p < 2.2 \times 10^{-16}$) while the pattern of fixed effects remained the same. We report a summary of the final model in Table 2.

---------- Insert Table 2 about here ……

The results demonstrate that participants learned during the experiment (the effect of block was significant and positive). The effect of experimental condition was not significant, with overall accuracies for all conditions similar and all greater than chance. However, the significant interaction between block and condition indicated that rate of learning was greater in the noun only condition than the verb only condition, and did not differ to the noun-and-verb condition. The model predictions are shown in Fig. 2 for each participant, demonstrating that for most participants, most of the time, performance was above chance (.5), and that learning progressed through the study.

---------- Insert Figure 2 about here ……

For the additional tests of whether nouns or verbs or both had been learned in the noun-and-verb condition, we tested a generalized linear mixed-effects model including fixed effects of stimulus type (noun or verb), test occasion (first or second test) and random effects on intercepts due to participants or action or object stimulus items. The model demonstrated that the average probability of producing a correct response was significantly greater than chance (estimated intercept log odds for the model = 1.56, SE = 0.28, z = 5.56, p = 2×10^{-8}, taking noun learning as the reference or baseline condition). Learning was not significantly affected by whether nouns or verbs were being learnt or by first or second test occasion, or by the interaction between those effects (all *p*s > 0.05). The accuracy for the offline tests are shown in Fig. 3.

---------- Insert Figure 3 about here ……

To assess the relationship between the training and offline tests, we added to the model of offline performance a fixed effect representing the percentage accuracy of each participant over the whole online training test. The effect of training accuracy was significant (estimate = .08, SE = .01, z = 6.22, $p = 5 \times 10^{-10}$), indicating that participants who were more accurate in online training were also more accurate in offline performance.

4. Discussion

We tested whether cross-situational learning was robust for acquiring word-referent mappings in situations including multiple grammatical categories of word. Previous studies had constrained the interpretation of the grammatical category, either through presenting only objects in the learner's environment (Smith & Yu, 2008; Yu & Smith, 2007), or by providing additional cues to constrain the interpretation of the word as a verb (Childers, 2011; Childers et al., 2012; Scott & Fisher, 2012). We showed that both nouns and verbs could be learned by adults with multiple possible categories amongst the speech references as well as in the environmental referents. This situation more closely resembles the task facing the language learner when complex sentences (Koehne & Crocker, in press) are uttered in the presence of multiple moving objects with different characteristics (Monaghan & Mattock, 2012; Yu & Ballard, 2007).

Furthermore, our study demonstrated that verb meanings could be acquired in the verb-only condition without the possibility of the learners first acquiring the labels for the objects that appeared, and without the requirement for additional known syntactic cues to grammatical category constraining the words' possible referents (Scott & Fisher, 2012; Waxman et al., 2009). Hence, for adults, verbs can be learned using similar sources of information to nouns, in line with Golinkoff et al.'s (1995, 1996) approach to discovering processing similarities in noun and verb learning (see also Golinkoff & Hirsh-Pasek, 2008).

The interaction between block and condition demonstrated that there was faster learning in the noun only condition compared to the verb only condition, which is compatible with observations of a noun learning advantage in children (Imai et al., 2008) and in adults (Gillette et al., 1999; Snedeker & Gleitman, 2004). The results are also consistent with previous studies comparing noun and verb learning that have also attempted to control for diverse information sources (Childers & Tomasello, 2002, 2003). However, when there was uncertainty about both nouns and verbs in the noun-and-verb condition, there was no significant difference in the accuracy with which nouns and verbs were learned as indicated in the offline tests.

The controlled environment of our study precludes explanations of the noun learning advantage based on ambiguity of referents for verbs (Gleitman, 1990; Gleitman & Gleitman, 1997), frequency and distributional differences between nouns and verbs (Mintz, 2006; Monaghan et al., 2007), greater seriality in verb-motion presentations (Gillette et al., 1999), distinctions in the number of potential referents for nouns and verbs, or requirements that nouns are learned prior to verbs (Fisher et al., 1994; Gillette et al., 1999; Gleitman, 1990). However, other properties of the stimuli may have resulted in a noun learning advantage. First, word-order flexibility may have adversely affected verbs, because nouns can occur in subject and object position. However, we feel this is unlikely, because verbs tend to occur in a wider range of contexts than do nouns (Monaghan et al., 2007), and verbs can occur phrase-initially as in imperatives, phrase-medially (in subject verb object constructions), as well as phrase-finally (as in subject intransitive-verb, or

subject finite-verb infinite-verb constructions). Second, participants may have exhibited a bias to interpret the nonwords as nouns either because of their phonological properties – first syllable stress is more likely for nouns than verbs (Arciuli, Monaghan, & Seva, 2010), or because they interpreted our instructions to view the scene as an instruction to attend to the objects. Yet, if this was the case then we would anticipate a distinction between noun learning and verb learning that occurs early in training. However, the interaction between language condition and block indicated a distinction between conditions only at later blocks of training, suggesting that initial biases did not substantially drive the effects.

We suggest that the noun learning advantage in our study is due to the salience or the complexity of the referents themselves (Gentner & Boroditsky, 2001). Motions require encoding of temporal sequences, whereas the objects are permanent and encapsulated on the screen at all times. Alternatively, distinctions between the motions, though reliably indicated in our pilot study, may have been less prominent than distinctions observed among the objects. Deciding between these alternatives is a matter of further research.

The studies we have presented were novel word learning studies by adults. So what relevance do the current studies have for the processes of language acquisition, rather than limited to language learning? There are clear distinctions between adults' and infants' word learning that provide caveats for the conclusions from our study. Adults have already acquired conceptual distinctions between nouns and verbs, and awareness of different grammatical categories within speech, which may assist in learning. Yet, even when such conceptual acquisition has

occurred, noun-learning advantages are generally observed (Gillette et al., 1999; Snedeker, Geren, & Shafto, 2007). Relatedly, one feature of the current experimental setup is that the adult participants already had an unambiguous label for the motions, whereas the objects were novel, which may affect comparisons in learning between nouns and verbs. An alternative approach would be to use motions that participants had never observed before, but then this would mean that it would not be possible to determine whether the motions are interpreted in the same way by every participant, thus introducing potential ambiguity into the verb-motion mapping (e.g., Gleitman, 1990).

The use of adult language learners also provides potential advantages for pinpointing the specific processes involved in verb acquisition. Our participants already had pre-individuated concepts for the objects and the motions presented in the experiment (Gentner & Boroditsky, 2001), as attested to by the unambiguous labelling of the motions by participants in the pilot study. Thus, conceptual distinctions between nouns and verbs that proffer a clear "package of features" (cf. Snedeker & Boroditsky, 2004) to individuate the motions and the objects were not at issue in the study. Yet, it remains an open question as to whether similar performance can be observed with infants as with adults (e.g., Gillette et al., 1999), as in previous extensions of cross-situational learning studies from adult to infant learners of single categories (Smith & Yu, 2008; Yu & Smith, 2007).

In conclusion, we have shown that when presented with a complex utterance and a complex scene involving nouns and verbs and objects and actions, a language

learner can acquire, through multiple experiences, whether “Gavagai” refers to objects or actions: whether “Gavagai” is or does.

References

Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioural, and computational investigations. *Journal of Memory and Language, 63*, 180-196.

Baldwin, D. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832-843.

Bernal, S., Lidz, J., Millote, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development, 3*, 325-341.

Bornstein, M., Cole, L., Maital, S.K., Park, S.Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean and American English. *Child Development, 75*, 1115-1140.

Childers, J. B. (2011). Attention to multiple events helps two-1/2-year-olds extend new verbs. *First Language*, *31*, 3–22.

Childers, J. B., Heard, M. E., Ring, K., Pai, A., & Sallquist, J. (2012). Children use different cues to guide noun and verb extensions. *Language Learning and Development*, *8*, 233-254.

Childers, J. B. & Paik, J. H. (2009). Korean-and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, *36*, 201-224.

Childers, J. B. & Tomasello, S. (2002). Two-year-olds learning novel nouns, verbs, and conventional actions from massed or spaced exposures. *Developmental Psychology, 38*, 967-978.

Childers, J. B. & Tomasello, S. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science, 6*, 185-190.

Fiser, J. & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458-467.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L.R. (1994). When is it better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua, 92*, 333-375.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development, Volume 2: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early verb learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp.215-256). Cambridge: Cambridge University Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 1-55.

Gleitman, L.R., Cassidy, K., Nappa, R., Papafragou, A. & Trueswell, J.C. (2005). Hard words. *Language Learning and Development, 1*, 23-64.

Gleitman, L. & Gleitman, H. (1997). What is a language made out of? *Lingua, 100*, 29-55.

Goldin-Meadow, S., Seligman, M.E.P., & Gelman, R. (1976). Language in the two-year-old: Receptive and productive stages. *Cognition, 4*, 189-202.

Golinkoff, R.M., Hirsh-Pasek, K. & Mervis, C. B. (1995). Lexical principles can be extended to the acquisition of verbs. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names of things Young children's acquisition of verbs* (pp. 185-221). Hillsdale, NK: Erlbaum.

Golinkoff, R.M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development, 67*, 3101-3119.

Horst, J. S., Scott, E. J. & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science, 13*, 706-713.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R., et al. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79*, 979–1000.

Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development, 73*, 93–109.

Koehne, J. & Crocker, M.W. (in press). The interplay of cross-situational word learning and sentence-level constraints. *Cognitive Science*, in press.

Markman, E.M. (1990). Constraints children place on word learning. *Cognitive Science, 14*, 57-77.

Mintz, T. H. (2006). Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs,* pp. 31-63. New York: Oxford University Press.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.

Monaghan, P. & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition, 123*, 133-143.

Oviatt, S. (1980). The emerging ability to comprehend language. *Child Development, 51*, 97-106.

Quine, W.V.O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Schwartz, R. G. & Leonard, L. B. (1984). The role of input frequency in lexical acquisition. *Journal of Child Language, 10*, 57-66.

Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163-180.

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-61.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cognitive Science, 35*, 480-498.

Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568.

Snedeker, J., Geren, J., & Shafto, C. (2007). Starting over: International adoption as a natural experiment in language development. Psychological Science, 18(1), 79-87.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (pp. 603–636). Cambridge MA: MIT Press.

Tomasello, M. & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development, 10*, 201-224.

Tomasello, M. & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language, 19*, 311-333.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). 24-month-old infants' interpretation of novel verbs and nouns in dynamic scenes. *Cognitive Psychology, 59,* 67-95.

Yu, C. & Ballard, D.H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149-2165.

Yu, C. & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414-420.

Acknowledgements

Thanks to Jozsef Fiser and Dick Aslin for generously providing their shape stimuli, to Emma Threadgold for recording the speech stimuli, and to Muriel Zhu for assisting with data collection. The research was supported by an EPS student bursary awarded to Alastair Smith.

Appendix 1: Speech stimuli

Bisyllabic words:

barget, bimdah, chelad, dingep, fisslin, goorshell, haagle, jeelow, kerrwoll, limeber, makkot, nellby, pakrid, rakken, shooglow, sumbark

Monosyllabic (function) words:

tha, noo

Table 1. Conditional probabilities of co-occurrence of auditory (words) and visual (objects and motion) stimuli in the three experimental conditions.

Condition	Auditory	Visual	P(V|A)	P(A|V)
Noun only	Object referring word	Object identity of target	1	.500
	Object referring word	Object identity of non-target	.143	.143
	Object referring word	Motion of target	.125	.125
	Object referring word	Motion of non-target	.125	.125
	Non-referring word	Object identity of target	.125	.125
	Non-referring word	Object identity of non-target	.125	.125
	Non-referring word	Motion of target	.125	.125
	Non-referring word	Motion of non-target	.125	.125
Verb only	Motion referring word	Object identity of target	.125	.125
	Motion referring word	Object identity of non-target	.125	.125
	Motion referring word	Motion of target	1	.500
	Motion referring word	Motion of non-target	.143	.143
	Non-referring word	Object identity of target	.125	.125
	Non-referring word	Object identity of non-target	.125	.125
	Non-referring word	Motion of target	.125	.125
	Non-referring word	Motion of non-target	.125	.125
Noun and verb	Object referring word	Object identity of target	1	.500
	Object referring word	Object identity of non-target	.143	.143
	Object referring word	Motion of target	.125	.125
	Object referring word	Motion of non-target	.125	.125
	Motion referring word	Object identity of target	.125	.125
	Motion referring word	Object identity of non-target	.125	.125
	Motion referring word	Motion of target	1	.500
	Motion referring word	Motion of non-target	.143	.143

Table 2. Summary of the Generalized Linear Mixed-effects Model of (log odds) accuracy of response over blocks and experimental conditions; note that the noun only condition is taken as the reference level for estimation of the effect of variation in levels of the factor experimental condition. R syntax for final model is:

```
glmer(accuracy ~ condition*block + (block + 1|Subject) +
(1|targetobject) + (1|targetaction) + (block +
0|targetobject) + (block + 0|targetaction), family =
binomial)
```

	Estimated		Wald confidence intervals			
Fixed effects	coefficient	SE	2.50%	97.50%	z	Pr(>|z|)
(Intercept)	-0.2669	0.1544	-0.5482	0.0143	-1.7300	0.0840
Experimental condition (noun vs. noun-verb)	-0.0193	0.1797	-0.3539	0.3149	-0.1100	0.9150
Experimental condition (noun vs. verb)	0.2661	0.2028	-0.0948	0.6266	1.3100	0.1900
Block effect	0.1871	0.0421	0.1049	0.2694	4.4500	< 0.0001
Experimental condition (noun-verb):block interaction	0.0117	0.0594	-0.1047	0.1282	0.2000	0.8440
Experimental condition (verb):block interaction	-0.1287	0.0596	-0.2447	-0.0126	-2.1600	0.0310

Random effects	Name	Variance	Std.Dev	Corr
Subject effect on intercepts	(Intercept)	0.0447	0.2115	
Random effect of subjects on slopes of block effects		0.0258	0.1606	-0.6800
Item effect (actin) on intercepts	block	< 0.0001	0.0023	
Random effect of items (actions) on slopes of block	block	0.0001	0.0096	
effects	(Intercept)	0.0153	0.1238	
Item effect (objects) on intercepts	(Intercept)	0.0084	0.0919	
Random effect of items (objects) on slopes of block effects				

AIC	BIC	logLik	deviance
16496.736	16594.681	-8235.368	16470.736

13824 observations, 48 participants, 8 target action stimuli plus null action, 8 target object stimuli plus null object

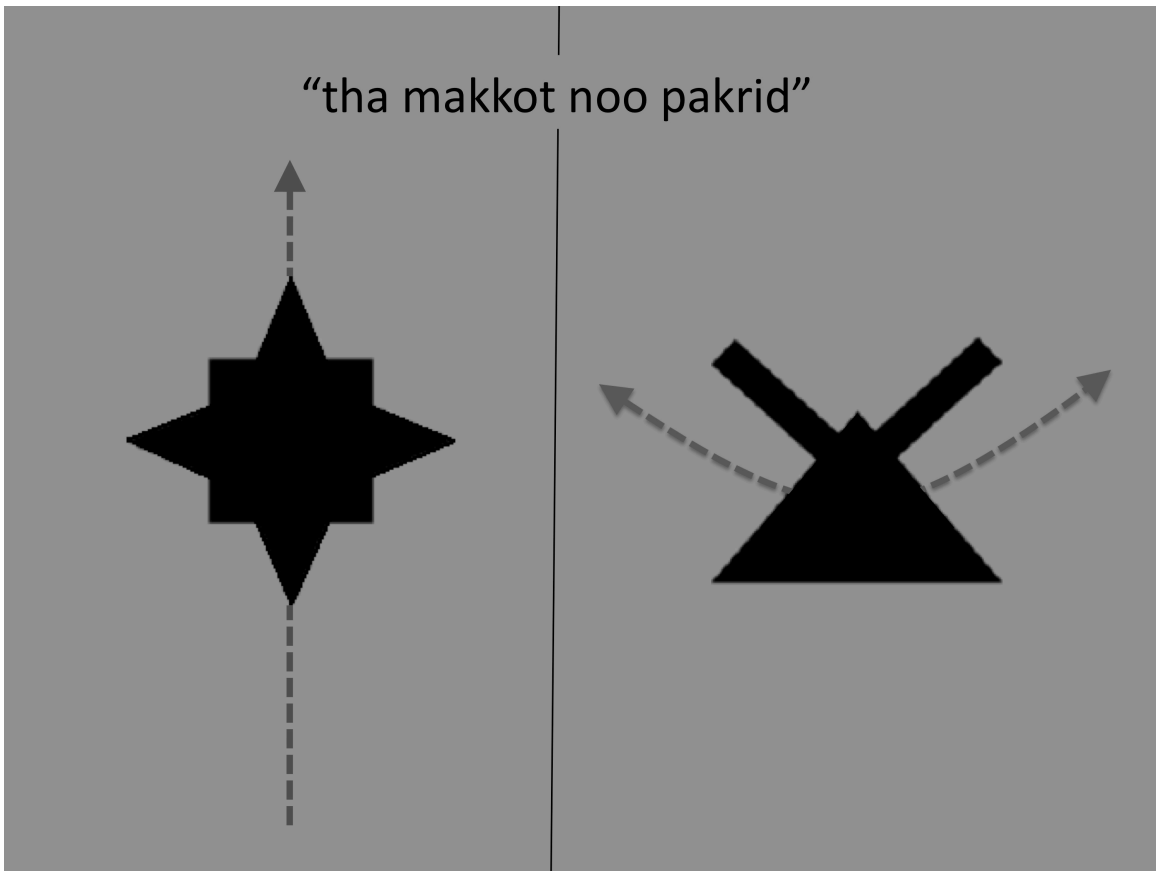
Figure Captions

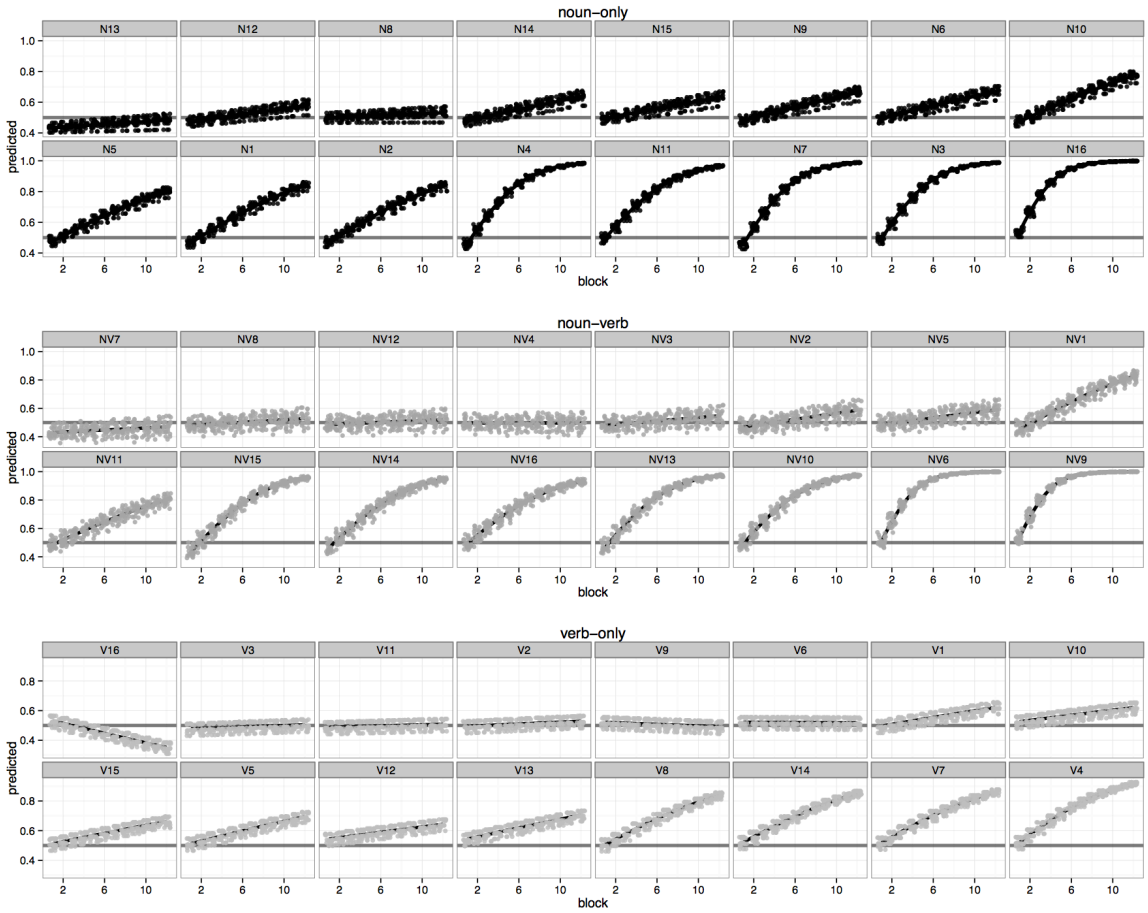
Fig. 1. Example of a learning trial. Two moving objects are observed. Arrows indicate the movement path of the object. The four word phrase is simultaneously heard, with “tha” and “noo” function words and “makkot” and/or “pakrid” referring to the motion and/or object, according to condition, in one of the scenes.

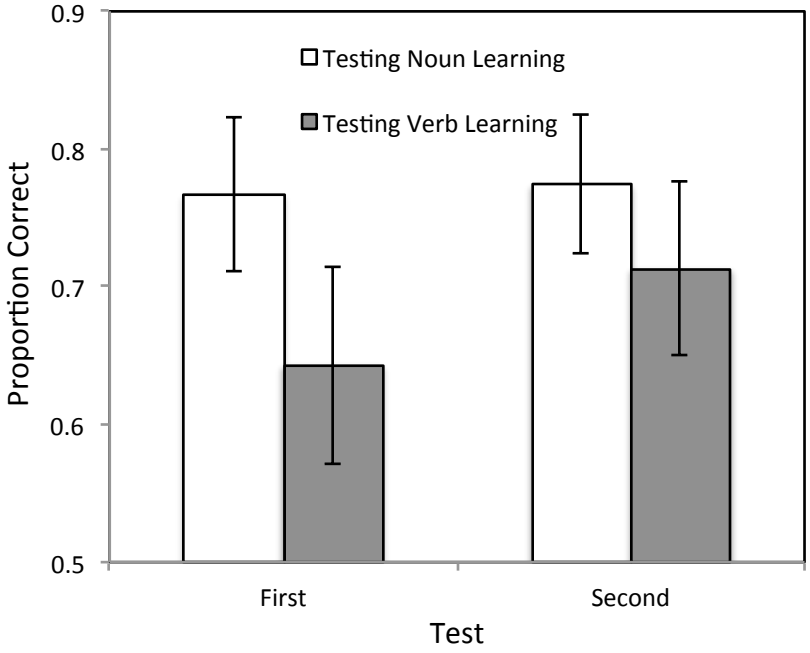
Fig. 2. Graph showing the Generalized Linear Mixed-effects Model predicted values of the probability that a response is correct, for each participant, in each trial.

Note to Fig. 2. Predictions are shown for each participant, individually. Predicted performance is shown separately for each condition: noun only; verb only; and noun and verb learning. For each set, plots are ordered from top left to bottom right by the overall percentage of correct responses actually *observed* for each participant. Participant codes are distinguished as: N# for the noun only condition (points in black); NV# for the noun-verb condition (points in dark grey); and V# for the verb only condition (points in light grey). In each plot, a curve has been added to indicate the trend, a horizontal line added at predicted probability = .5 shows chance level performance. Point location has been jittered to alleviate over-plotting.

Fig. 3. Means and SEM for off-line tests of independent noun and verb learning in the noun-and-verb condition.


“tha makkot noo pakrid”


noun–only
N13
N12
N8
N14
N15
N9
N6
N10
N5
N1
N2
N4
N11
N7
N3
N16
predicted
block
noun–verb
NV7
NV8
NV12
NV4
NV3
NV2
NV5
NV1
NV11
NV15
NV14
NV16
NV13
NV10
NV6
NV9
predicted
block
verb–only
V16
V3
V11
V2
V9
V6
V1
V10
V15
V5
V12
V13
V8
V14
V7
V4
predicted
block


0.9
0.8
0.7
0.6
0.5
Proportion Correct
Testing Noun Learning
Testing Verb Learning
First
Second
Test