Reconsidering human cross-situational learning capacities: a revision to Yu & Smith's (2007) experimental paradigm

Kenny Smith (kenny.smith@northumbria.ac.uk)
Cognition and Communication Research Centre, Division of Psychology, Northumbria University,
Northumberland Building, Northumberland Road, Newcastle upon Tyne, NE1 8ST, UK

Andrew D. M. Smith (andrew@ling.ed.ac.uk)
Language Evolution and Computation Research Unit, PPLS, University of Edinburgh,
Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK

Richard A. Blythe (R.A.Blythe@ed.ac.uk)
SUPA, School of Physics and Astronomy, University of Edinburgh,
James Clerk Maxwell Building, Mayfield Road, Edinburgh, EH9 3JZ, UK

Abstract

Cross-situational learning allows word learning despite exposure-by-exposure uncertainty about a word's meaning, by combining information across exposures to a word. A number of experimental studies demonstrate that humans are capable of cross-situational learning. The strongest claims here are made by Yu and Smith (2007), who provide experimental data suggesting that adult humans are capable of using cross-situational learning to rapidly learn the meanings of multiple words simultaneously and despite considerable uncertainty at each exposure. We identify a flaw in their testing regime which throws their conclusions into doubt, and conduct a new experiment which remedies this methodological flaw. Our data supports a more limited view of the ability of adults to rapidly and simultaneously apply cross-situational learning in conditions of (relatively) high referential uncertainty.

Keywords: word learning; cross-situational learning

Introduction

Learning the meaning of a new word is a challenge: as famously noted by Quine (1960), there are in principle infinitely many meanings which could be consistent with a single utterance of an unfamiliar word, or any sequence of such utterances. However, despite this theoretical difficulty, humans manifestly *do* learn the meanings of new words, and rapidly too (Bloom, 2000).

A productive area of investigation has been to explore how children eliminate some of the theoretically-possible but in practice spurious candidate meanings for a new word, in order to make a better guess as to the word's true meaning. A number of such heuristics have been identified, including (to name but two examples) the mutual exclusivity heuristic (each object should have only one name, and therefore already-labelled objects can be excluded as candidate referents for new words: Markman & Wachtel, 1988) and the shape bias (words generalise by shape, so a new word probably refers to a category-typical shape rather than colour, texture, etc: Landau, Smith, & Jones, 1988).

However, heuristics of this sort are unlikely to uniquely identify the true meaning of a word on every word learning encounter: some residue of uncertainty will remain. Various authors (e.g. Pinker, 1989; Siskind, 1996; Gillette, Gleitman, Gleitman, & Lederer, 1999; Yu & Smith, 2007; L. B. Smith & Yu, 2008; Frank, Goodman, & Tenenbaum, in press) have argued that words can still be learned despite this sort of uncertainty, by combining information across exposures via *cross-situational learning*. While a number of flavours of cross-situational learning exist (see e.g. Siskind, 1996; Frank et al., in press), the basic premise is as follows. Each situation in which a word is used provides a number of candidates for that word's meaning. Multiple uses therefore produce multiple sets of candidate meanings, and the learner can make use of this cross-situational information, for example by assuming that the true meaning of the word lies at the intersection of these sets of candidate meanings.

Several experimental studies have shown that adults and children are capable of cross-situational learning, both from naturalistic stimuli (e.g. Gillette et al., 1999; Piccin & Waxman, 2007) and more stylised materials (Yu & Smith, 2007; L. B. Smith & Yu, 2008). Perhaps the most striking demonstration of human cross-situational learning prowess is provided by Yu and Smith (2007), henceforth YS, who show that adult learners are able to learn multiple words simultaneously despite a small number of exposures to each word and referential uncertainty at each exposure. We show here that a flaw in YS's testing regime means that they risk substantially overestimating human cross-situational learning proficiency.

In the next section we provide an analysis of their testing regime and demonstrate that their human participants are in fact outperformed by a simple non-cross-situational learning procedure. This severely undermines the conclusion that their experimental results indicate powerful cross-situational learning abilities. We then describe a replication and extension of YS's results using a more appropriate testing regime. Our results provide only partial support for YS's original conclusions. At intermediate levels of referential uncertainty, human participants can indeed use cross-situational learning to rapidly learn multiple words in parallel. However, our results also show that at the highest level of referential uncertainty tested, this ability breaks down — human cross-situational learning abilities may be somewhat weaker than suggested by YS. Furthermore, our results suggest problems at lower levels of referential uncertainty, possibly arising from the difficulty of integrating information across widely separated exposures.

Analysing Yu & Smith (2007)

In order to demonstrate conclusively that a group of word learners are capable of cross-situational word learning, the following three steps are necessary: (1) Present those learners with a sequence of training exposures to a target word or target words; (2) Test those learners on their ability to correctly identify the meaning of each target word; (3) Demonstrate that the resulting performance is significantly better than that which could be obtained by any learner using a single one of those training exposures (i.e. the learning performance truly indicates cross-situational, rather than one-shot, learning).

YS present a series of word-learning studies which they argue meet these three criteria. We believe a deficiency in their testing regime (item 2 above) means that they are not in fact able to satisfy the third requirement: a learner capable of only remembering a single training exposure for each word outperforms their human participants under their testing regime.

YS use the following experimental set-up (their Experiment 1) to demonstrate cross-situational learning. Adult participants are presented with a series of exposures to words (aurally presented) paired with referent objects (presented visually). Example exposures (using our materials, not those used by YS) are illustrated in Figure 1. At each exposure 2, 3 or 4 words are presented simultaneously, depending on condition (these are referred to as the 2x2, 3x3 and 4x4 conditions respectively), with all participants experiencing all conditions. Participants are trained on 18 words in each condition, with the training set being constructed such that each word is paired with its referent object six times.

During testing, participants are presented with each word in turn, and asked to pick out the correct referent object for that word from an array consisting of the correct object plus three foils, which are themselves referents of other words and will therefore have been encountered six times during training (see Figure 1d). YS show that human participants perform significantly better than a memoryless baseline strategy which selects randomly among the four test objects, which would get 25% of test items correct on average: human learners score on average between 53% and 88%, depending on condition, with greater levels of referential uncertainty (i.e. the 4x4 condition) leading to reduced performance.

The baseline performance that YS evaluate their participants against is that which would be achieved by a learner with no memory of any of the training exposures they received. Humans perform better than this. However, there are other learners who might perform better than this memoryless learner without making use of cross-situational information. In order to demonstrate that humans are performing cross-situational learning, it must also be demonstrated that they are outperforming these non-cross-situational learners.

Consider the following learner, who we will term the *one-exposure learner*. The one-exposure learner remembers the

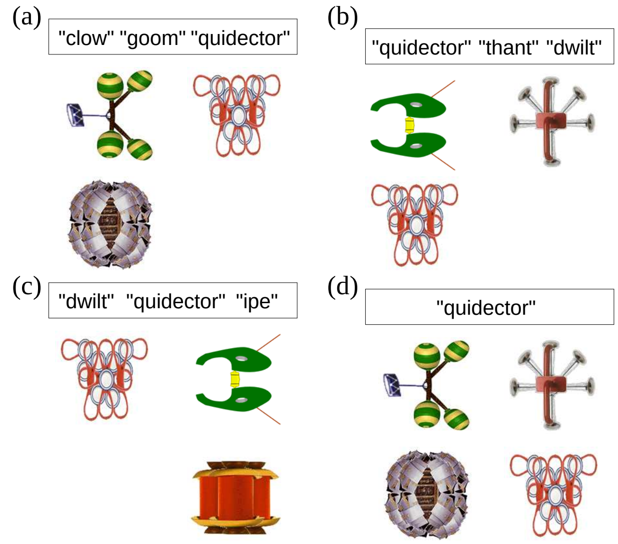

Figure 1: Training (a–c) and test (d) exposures in the YS 3x3 condition.

details of only one of each of the six exposures they receive for each word — e.g. the first exposure, or the last. On testing on a particular target word, this learner selects at random from all the test objects which it saw paired with the target word on the single exposure it remembers. For example, looking at the training and testing episodes depicted in Figure 1, if the one-exposure learner remembers the exposure depicted in Figure 1(a), upon testing on the array shown in Figure 1(d) the one-exposure learner would guess at random among three possible referent objects (those objects present in the single remembered exposure). Alternatively, if the learner only remembered the exposure depicted in Fig. 1(b) or (c), the learner would guess at random among two possibilities in the test array (if it remembered exposure b) or correctly identify the target referent as it is the only object from the single remembered exposure present in the test array (if c was the sole remembered exposure). This learner is clearly not integrating information across exposures, since it only remembers a single training exposure for each word. Nor is it learning from the test exposure — given a different test array, this one-exposure learner might perform differently.[1]

How well would the one-exposure learner perform on YS's test regime? We have previously provided a mathematical formalism which can be straightforwardly adopted to calcu-

[1]Given that the test array constrains which referent objects can be selected, the one-exposure learner could be said to be exploiting cross-situational information between the single remembered exposure and the test array. While this is a potentially interesting interpretation, we believe it is desirable to separate this weak form of cross-situational learning (which essentially exploits the reduction in uncertainty afforded by the test regime) with cross-situational learning across multiple exposures (which is more relevant to cross-situational word learning in the real world). As we will argue in the rest of the paper, the YS testing regime obscures the extent of this more interesting form of cross-situational learning by using a test which rewards one-exposure learning.

late this (equation 1 in K. Smith, Smith, Blythe, & Vogt, 2006). We will repeat the relevant expression here, modified to the question at hand for ease of exposition.

YS's experimental deign can be expressed as follows. A learner attempts to learn W words, each paired uniquely with one of W associated referent objects. At each exposure to a particular target word, a learner sees the target word (and a number of other words) paired with the target referent plus C non-target referent objects (the *training foils*: $C = 1$ in YS's 2x2 condition, $C = 2$ in the 3x3 condition, $C = 3$ in the 4x4 condition). Those training foils are drawn from a set of $M = W - 1$ objects. During testing, the learner is presented with the target word and asked to identify the target referent from a set consisting of the target referent plus T ($= 3$ in YS's paradigm) non-target referents (*test foils*) drawn from the set of $M = W - 1$ non-target referents.

For a one-exposure learner, the relevant question is: for a given word, how many of the T test foils were also present in the set of C training foils during the single remembered exposure to the target word? If there is no overlap between training and test foils the one-exposure learner will correctly identify the target referent as the only object it remembers co-present with the target word. If there is some overlap between these sets of foils, the one-exposure learner will guess at random between the target object and the members of this overlapping set.

The probability of O items being present in the overlap between T test foils and C training foils is given by

$$Q(O|T,C,M) = \binom{T}{O} \cdot \binom{M-T}{C-O} \cdot \binom{M}{C}^{-1}. \qquad (1)$$

The first term is the number of ways of correctly selecting overlapping items: there are $\binom{T}{O}$ ways in which the desired number of overlapping foils (O) can be chosen from the test foils T. The second term is likewise the number of ways of correctly selecting non-overlapping items: $M - T$ gives the number of referents which are *not* test foils, and we must select $C - O$ training foils from this set. The number of valid combinations of training foils and test foils which satisfy the desired condition is the product of these two expressions, and the probability of obtaining O objects which were present in both the test and training foil sets is obtained by multiplying this quantity by the probability of a given set of training foils.

Once we have calculated this quantity it is relatively easy to calculate the probability of a one-exposure learner guessing the target meaning correctly on test. This is:

$$P_{\text{one}}(T,C,M) = \sum_{O=0}^{O=T} \frac{1}{O+1} \cdot Q(O|T,C,M) \qquad (2)$$

where the sum is over the possible sizes of overlapping sets, and the fraction gives the probability of correctly selecting the target meaning by chance from the union of the overlapping set and the target meaning. It can be shown that this

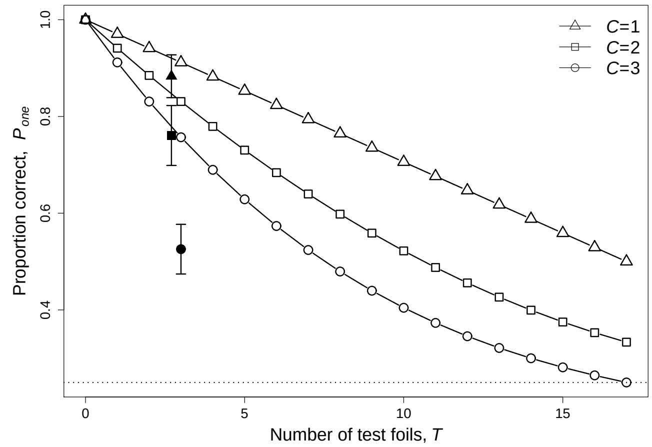

Figure 2: Probability of correctly guessing the target for a one-exposure learner, for various numbers of test foils (T). Mean and 95% confidence interval for the mean ($1.96 \times SD/\sqrt{N}$) of YS's human learners are given as solid points plus error bars, with shapes coded according to C, offset as necessary to avoid obscuring other points. Horizontal dotted line gives performance of the YS memoryless learner.

expression has the comparatively convenient closed form

$$P_{\text{one}}(T,C,M) = \frac{1}{T+1}\left[\frac{M+1}{C+1} - \binom{M-T}{C+1} \cdot \binom{M}{C}^{-1}\right]. \qquad (3)$$

Figure 2 shows, for $M = 17$ and $C = 1$, 2 or 3 (the parameters used by YS in their Experiment 1) the probability of this one-exposure learner correctly identifying the target referent under various testing regimes, including $T = 3$ (the number of test foils used by YS) and $T = M$, the hardest possible test where the learner is confronted with all possible word referents at every test. We also plot the human performance from Yu and Smith (2007), Experiment 1.

There are several things to note. Firstly, the one-exposure learner outperforms YS's 25% baseline in all cases where $T = 3$, and in nearly all other test regimes. Secondly, and most importantly, under the $C = 1$ condition human performance is not significantly different from that of the one-exposure learner ($t(37) = 1.27$, $p = 0.21$)[2], whereas under the $C = 2$ and $C = 3$ conditions humans perform significantly *worse* than the one-exposure learner($C = 2$: $t(37) = 2.22$, $p = 0.03$; $C = 3$: $t(37) = 8.83$, $p < 0.001$).[3] Finally, the one-exposure learner performs worst on the test when $T = M$, at which point it will achieve a proportion $1/C + 1$ correct (guessing among the target and the C training foils on the sin-

[2] All p values reported in this paper are for two-tailed tests.

[3] YS report a second set of experimental results (their Experiment 2), which involves comparing a replication of the 4x4 condition of Experiment 1 with two conditions where participants learn smaller lexicons (9 words) in the 4x4 conditions, with more repetition of each word. Performance in the 9-word lexicon conditions is indistinguishable from one-exposure learner performance: greatest $t(27) = 1.57$, $p = 0.128$.

gle remembered training exposure). If we wish to eliminate a non-cross-situational learning strategy like the one-exposure learner as a candidate explanation for human behaviour, the best approach is therefore to test on the full array of referent objects, rather than a subset.

The fact that the one-exposure learner performs better than humans is slightly puzzling — can we reject non-cross-situational learning on the basis that human performance is inconsistent with this learning behaviour? Unfortunately not. One straightforward way to account for this mismatch is to introduce the notion of a *forgetful one-exposure learner*. As before, this learner only remembers a single exposure to each target word, but within that exposure forgets each referent object (including the target) with probability f. We can provide the following expression for the probability that a forgetful one-exposure learner will guess correctly on a particular target word:

$$P_{\mathrm{f}}(f,T,C,M) = \left(\tfrac{1}{T+1}\right) f^{C+1} + (1-f)\left[\sum_{C'=0}^{C'=C} P(C'|C,f)\cdot P_{\mathrm{one}}(T,C',M)\right] \tag{4}$$

where $P(C'|C,f)$ gives the probability of remembering C' of the C training foils, given by the expression

$$P(C'|C,f) = \binom{C}{C'}\cdot(1-f)^{C'}\cdot f^{C-C'}\ . \tag{5}$$

The second term in the expression for P_{f} gives the probability of correctly identifying the target during testing, weighted by the probability of remembering the target $(1-f)$ and between 0 and C of the training foils. The first term covers the case where all details of the training exposure are forgotten, including the target, in which case the forgetful one-exposure learner picks at random among the $T+1$ possibilities. The other possibilities (forgetting the target and remembering one or more of the training foils) will lead to incorrect guesses and can therefore be omitted.

Close matches between the forgetful one-exposure learner and the mean performance of YS's human participants can be achieved by assuming that f increases with C. For example, human behaviour is not significantly different from that of the one-exposure learner if we assume the following f values: $C=1,\ f=0; C=2,\ f=0.1; C=3,\ f=0.4; t(37)\leq 0.37$.

Based on these results, we therefore cannot reject the null hypothesis that humans are incapable of cross-situational learning and are achieving the observed levels of performance by simply (partially) remembering a single exposure from the sequence of exposures. It is important to emphasise that we know that humans *are* capable of cross-situational learning — as discussed above there are a number of other empirical demonstrations of cross-situational learning (e.g. Gillette et al., 1999; Piccin & Waxman, 2007). However, YS's study makes the strongest claims about human cross-situational learning (both in terms of its rapidity and simultaneity), and a replication of the YS experimental paradigm with a more demanding test regime (such that participants must identify the target referent from an array of all possible referent objects) is required to support this conclusion — at present, we cannot rule out non-cross-situational learning as a potential explanation for the observed behaviour, or at least as a confounding factor masking the true learning abilities of their participants.

An Experimental Test

We therefore ran an experimental study to remedy the flaw in the Yu and Smith (2007) method. As described below, we ran two groups of participants: a Control group, who underwent a direct replication of YS's Experiment 1, and an Experimental group, where we replaced the flawed YS test with a more robust test of cross-situational learning ability (each test comprises the target plus all 17 possible test foils).

Method

Participants 48 undergraduate psychology students at Northumbria University participated in the study as part of a participation co-op scheme.

Materials Participants were asked to learn pairings of a referent object and a spoken (nonsense) word form. 54 novel referent objects were created by cutting and pasting together components parts of pictures of technological artifacts to produce novel objects — see Figure 1 for examples. We created 54 nonsense words (using the English Lexicon Project Website: Balota et al., 2007) which followed English phonotactics and were stratified according to number of syllables (1–3), stress (first or second syllable) and initial sound (vowel, single consonant, consonant cluster). These words were grouped into three sets, such that each set had a similar sample of the various word types and the subjectively more confusable words were in different sets. Spoken forms of these words were produced using the Victoria voice on the Apple Mac OS X built-in speech synthesiser.

Design and Procedure Following YS, participants were explicitly briefed on the task: they would have to work out which object went with which word, multiple objects would appear on the screen and their associated words would be spoken, there was no relationship between where the object appeared on the screen and the order in which the words were spoken, and their task was therefore to work out across trials which word went with which object. Participants were tested in groups of between 1 and 5, seated at a PC in a room with the PCs distributed around the periphery facing the walls. Participants observed objects being displayed on the monitor and listened to words being presented over headphones.

As in the YS study, participants were tested on three sets of 18 word-object pairings. Each set of 18 word-object pairings was created by pairing each word from one of the word sets with a referent object selected randomly without replacement from the set of 54 referent objects. The degree of referential uncertainty varied between word sets, with either two, three or four words and their referents being presented at each exposure trial (the 2x2, 3x3 and 4x4 conditions, corresponding to C of 1, 2 or 3). All participants experienced all three conditions. Following YS, exposure times were designed so that total training time was the same in each condition (see Table 1). Given that the training sequence is independent of the test regime, we designed the training sequences such that participants were paired across the Control and Experimental group — for every member of the Control group, there was a participant in the Experimental group who received an identical series of training exposures but underwent a more rigorous test. Order of presentation of the three conditions and three sets of word forms was counterbalanced across participants.

Table 1: The training regimes

Condition	# trials	Time per trial (secs)	Total time
2x2	54	6	324
3x3	36	9	324
4x4	27	12	324

After training on a word set was completed, participants were tested: each word from the current word set was presented aurally and the participants were instructed to select the associated object from the test array by clicking on it using the mouse. Participants were randomly assigned to one of two groups. The Control group were tested using the YS test regime — on each test trial, they were required to identify the target referent from an array of four objects, the actual target object and three foils selected at random from the set of 18 referent objects in use for this word set. Replicating the YS regime allows us to check for any differences with their basic result arising from differences in participants or materials. The Experimental group were tested using what we identify above as the correct test regime — on each test trial, they were required to identify the target referent from an array of all 18 referent objects associated with this word set.

Results

Figure 3 shows the results from our two groups, alongside the results from YS's Experiment 1. Levels of performance in the Control condition correspond fairly well with those of YS, the greatest difference being slightly lower performance of our participants in the 3x3 condition (YS: $M = 13.69$ words learned; Control: $M = 12.92$), but this difference is not significant ($t(60) = 0.83$, $p = 0.412$). This gives us some confidence that our materials and participant pool are roughly comparable to those of YS.

Focusing on the contrast between the Control and Experimental groups: as suggested by Figure 3, an ANOVA with referential uncertainty as a within-subjects factor and three between-subjects factors (test regime, order of presentation of the three levels of referential uncertainty, order of presentation of the three sets of word forms) reveals a main effect of referential uncertainty during training ($F(2,72) = 99.84$, $p < 0.001$) and of test configuration ($F(1,36) = 20.62$, $p < 0.001$). There is also a significant interaction between referential uncertainty and order of presentation of the three levels

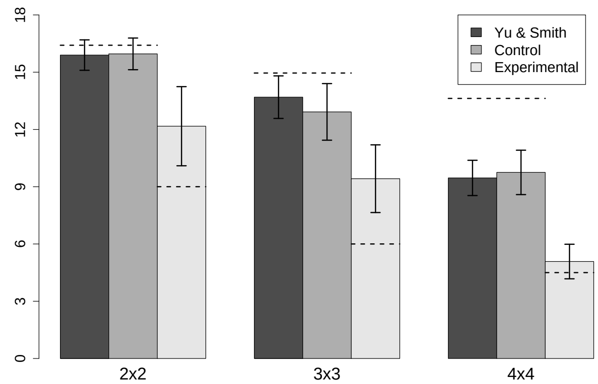

Figure 3: Mean performance (out of 18 words) of YS's participants and our participants, organised by condition. Error bars give the 95% confidence interval of the mean. Dashed horizontal lines give one-exposure performance — note that one-exposure performance for YS and Control groups is sometimes far greater than human performance, as discussed with reference to Figure 2.

of referential uncertainty ($F(10,72) = 3.97$, $p < 0.001$), indicative of a practice effect: participants perform relatively poorly on their first word set (averaging across test regimes and levels of referential uncertainty, $M = 9.54$ words correct) relative to their second and third sets ($M = 11.42$ and $M = 11.69$ words correct respectively). Counterbalancing of presentation orders means this practice effect does not alter the overall pattern of results we report in the remainder of the paper. All other main effects and interactions are n. s.

Looking at the performance of matched pairs of participants across the two testing conditions, participants in the Experimental group perform worse across the board: smallest $t(23) = 3.253$, $p = 0.004$, occurring in the 2x2 condition. Post-hoc tests on both the Control and Experimental groups show that performance differed significantly between each level of referential uncertainty (all Bonferroni-corrected p values < 0.01).

We can also ask whether performance in the Experimental group gives a clear signal that our participants are doing cross-situational learning: are they significantly better than the best non-cross-situational performance level, afforded by one-exposure learning? One-sample tests show that participants in the Experimental group perform above the one-exposure chance levels in the 2x2 condition (mean one-exposure learner performance of 9 words, $W = 204$, $p = 0.012$)[4] and in the 3x3 conditions (mean one-exposure learner performance of 6 words, $t(23) = 3.77$, $p = 0.001$), but not in the 4x4 condition (mean one-exposure performance of 4.5 words, $t(23) = 1.26$, $p = 0.219$).

[4]The non-normal nature of the distribution of scores on the 2x2 condition ($A^2 = 0.782$, $p = 0.032$) necessitates use of a one-sample Wilcoxon Signed Rank Test rather than a one-sample t-test.

Discussion

The finding that, in the Experimental 4x4 condition, our participants do not perform significantly above the one-exposure level of performance is a major departure from the conclusions reached by YS, and suggests that, given a more careful construction of the testing regime, our participants may not be capable of doing cross-situational learning in the 4x4 condition, contrary to the conclusions drawn by YS. In other words, due to a flawed testing regime, YS may have overestimated the human capacity for rapid, simultaneous cross-situational learning under higher levels of referential uncertainty.

We had anticipated the strongest signal of cross-situational learning in the 2x2 condition. In fact, the signal of cross-situational learning here (relative to the one-exposure baseline) is equivalent or slightly weaker to that in the 3x3 condition (mean difference from the one-exposure level = 3.17 words in the 2x2 condition, 3.42 in the 3x3 condition). While this might merely reflect the high level of the one-exposure baseline in the 2x2 condition, another interpretation is possible and perhaps worthy of exploration. In the 2x2 condition participants receive 54 randomly-ordered exposures, as opposed to 36 in the 3x3 condition and 27 in the 4x4 condition. Consequently, consecutive exposures to a given word in the 2x2 condition are disproportionately likely to be broken up by intervening exposures to other words. All other things being equal, information across trials is therefore more likely to be forgotten in the 2x2 condition than in the other conditions. This might explain why performance in this condition is not even better differentiated from one-exposure performance — it may be that the relatively fragmented nature of the training stimuli in the 2x2 condition fosters a behaviour more in line with the one-exposure baseline. Note that the test configuration used by YS would not reveal this problem, given the high baseline level of one-exposure performance in their 2x2 condition: essentially, it doesn't matter if you can't remember more than one exposure, since you'll probably guess right on test anyway.

Under this interpretation, learners are therefore faced with two difficulties in applying cross-situational learning in this experiment: (1) degree of referential uncertainty, the 4x4 condition being the most challenging; (2) forgetting between exposures, with the 2x2 condition being hardest. Performance is greatest (relative to the one-exposure baseline) when both these pressures are minimised — in the 2x2 condition, it may be that forgetting effects drag performance back towards the baseline despite low referential uncertainty. We are currently running a series of follow-up experiments to explore the impact of the interleaving of exposures on performance (point 2 above), as well as exploring whether cross-situational learning is possible under higher levels of referential uncertainty (e.g. 4x4) for smaller ($<$ 18 word) lexicons.

Conclusions

Our analysis of the flaw in the testing regime used by Yu and Smith (2007) shows that their results do not support the conclusion that adult humans are capable of rapidly learning multiple words in the face of relatively high referential uncertainty. Our replication and extension of their work shows that more limited conclusions are justified: while our participants were clearly doing cross-situational learning under the low (2x2) and intermediate levels (3x3) of referential uncertainty, there was no clear signal of cross-situational learning under the highest level of referential uncertainty tested (4x4). These results suggest that human capacities for rapid and simultaneous cross-situational may be more limited than suggested by Yu and Smith (2007).

Acknowledgements

A. D. M. S. is supported by ESRC grant RES-062-23-1537. R. A. B. is an RCUK Academic Fellow. Thanks to Ronan O'Ceallaigh for helpful early discussions.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. (in press). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*.

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

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299–321.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, *20*, 121–157.

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: new evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*, 295–323.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

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

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 1–38.

Smith, K., Smith, A. D. M., Blythe, R. A., & Vogt, P. (2006). Cross-situational learning: a mathematical approach. In P. Vogt, Y. Sugita, E. Tuci, & C. Nehaniv (Eds.), *Symbol grounding and beyond* (pp. 31–44). Berlin: Springer.

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

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