

Children's referent selection and word learning: insights from a developmental  
robotic system

Katherine E. Twomey<sup>1</sup>, Anthony F. Morse<sup>2</sup>, Angelo Cangelosi<sup>2</sup> & Jessica S. Horst<sup>3</sup>

Author note

<sup>1</sup>Department of Psychology, Lancaster University, UK.

<sup>2</sup>Centre for Robotics and Neural Systems, University of Plymouth, UK

<sup>3</sup>School of Psychology, University of Sussex, UK

Katherine E Twomey is a Senior Research Associate in the International Centre for Language and Communicative Development (LuCiD) at Lancaster University. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged. Empirical and simulated data are available on request from the authors.

Correspondence concerning this article should be addressed to Katherine E. Twomey, Department of Psychology, Lancaster University, Bailrigg, Lancaster, UK, LA1 4YF. Contact: [k.twomey@lancaster.ac.uk](mailto:k.twomey@lancaster.ac.uk).

### Abstract

It is well-established that toddlers can correctly select a novel referent from an ambiguous array in response to a novel label. There is also a growing consensus that robust word learning requires repeated label-object encounters. However, the effect of the context in which a novel object is encountered is less well-understood. We present two embodied neural network replications of recent empirical tasks which demonstrated that the context in which a target object is encountered is fundamental to referent selection and word learning. Our model offers an explicit account of the bottom-up associative and embodied mechanisms which could support children's early word learning and emphasises the importance of viewing behaviour as the interaction of learning at multiple timescales.

Keywords: word learning; referent selection; fast mapping; self-organizing maps; developmental robotics; iCub; Epigenetic Robotics Architecture; competition; novelty

**Children’s referent selection and word learning: insights from a developmental robotic system**

Toddlers perform impressively when confronted with the seemingly difficult task of choosing the referent of a new word in an ambiguous environment (e.g., Clark, 1995). A rich empirical literature shows that children from as young as 18 months reliably map novel labels to novel referents without being explicitly taught the “correct” mapping (Axelsson, Churchley, & Horst, 2012; Carey & Bartlett, 1978; Houston-Price, Plunkett, & Harris, 2005). Theories as to the mechanisms underlying *referent selection* range from innate, top-down knowledge to bottom-up associative mechanisms. Early accounts proposed pre-existing word learning biases, or *assumptions* (Markman, 1990; Markman, 1994). For example, children’s tendency to map novel labels to objects of the same type, rather than the same thematic category could be guided by a *taxonomic assumption* (Markman & Hutchinson, 1984) by which the same “kind of things” share a label; hence, a new furry, barking animal is called *dog*, but a new bone-shaped toy is not. Similarly, a *whole object* assumption could explain why children preferentially map new labels to an entire object rather than one of its parts (Markman & Wachtel, 1988). Further, when confronted with an array of objects, all-but-one of which are familiar, children will reliably map a novel label to the novel object. This could be the result of a *mutual exclusivity (ME)* assumption, by which objects have one and only one label (Markman & Wachtel, 1988; Merriman, Bowman & MacWhinney, 1989). For example, if a child sees a furry, barking animal, a scaly, swimming animal, and a pink, feathered animal, an ME assumption would prompt him/her to map the new label *flamingo* onto the feathered animal, because the other two objects are called *dog* and *fish*.

Although these constraints neatly describe what children do when hearing new words, precisely what children need to know in order correctly identify the referent of a new label is the subject of some debate. For example, Golinkoff and colleagues (1992) argued that children learn flexible lexical principles rather than possessing innate, hard-and-fast biases. Indeed, studies showing that word learning biases can be overridden by – or even depend on – sociopragmatic cues suggest that referent selection is a flexible behaviour (Baldwin, 1993; Tomasello & Akhtar, 1995). The source of ME-type behaviour has been particularly controversial. Clark (1990) argued for a *principle of contrast* by which children assume that adults use words consistently, so new labels must contrast in meaning to already-known labels and therefore refer to previously unlabelled objects. In contrast, Mervis & Bertrand (1994) posited a *novel-name-nameless-category (N3C)* strategy, whereby children make a simple novelty-to-novelty mapping, linking new words directly to unlabelled objects. Others argue that they use a process of elimination, explicitly ruling out known competitor objects before mapping the novel label to the novel object (Halberda, 2006). Importantly, these accounts differ as to the importance they place on children's well-documented attentional bias towards novel over familiar items, or *novelty preference* (Fantz, 1964; Houston-Price & Nakai, 2004). While attention to novelty is fundamental to N3C, its role in ME is less clear. Further, whichever mechanism underlies ME-type behaviour, it is disputed whether successful referent selection is the result of explicit, metacognitive reasoning about a label's potential referents, or whether it is due to low-level associative processes (e.g., Smith, Jones, Yoshida, & Colunga, 2003).

Importantly, there is substantial evidence that simply disambiguating the referent of a novel word once is not on its own sufficient for word learning (Bion,

Borovsky, & Fernald, 2013; Horst & Samuelson, 2008; Kucker, McMurray, & Samuelson, 2015; Mather & Plunkett, 2009; McMurray, Horst, & Samuelson, 2012; Munro, Baker, McGregor, Docking, & Arculi, 2012; Twomey, Ranson & Horst, 2014). Rather, word learning is the result of incremental *cross-situational learning*, in which label-referent mappings are gradually strengthened via repeated encounters with the mapping (e.g., Fazly, Alishahi, & Stevenson, 2010; Horst, McMurray & Samuelson, 2006; Smith & Yu, 2008; Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014). For example, a child might learn that a furry, meowing animal with four legs is called *cat* after s/he encounters the label alongside a toy cat at nursery, a photograph of a cat in a storybook, and her pet cat at home. However, while there is a broad consensus that cross-situational learning is a domain-general learning mechanism that can drive word learning, developmental psychologists disagree about the mechanisms that allow children to solve the in-the-moment referent selection puzzle.

### **Computational and robotic insights into development.**

Recent interdisciplinary research has begun to address this issue by integrating insights from developmental psychology with computational and robotic techniques to explore the perceptual and cognitive processes underlying empirically observed behaviour (Cangelosi, Schlesinger, & Smith, 2015; Gliozzi, Mayor, Hu, & Plunkett, 2009; McMurray et al., 2012; Morse & Cangelosi, in press; Samuelson, Smith, Perry, & Spencer, 2011; Westermann & Mareschal, 2014). Computational models of word learning simulate how children behave (e.g., pointing to the flamingo and not the dog or the fish) based on what they see and hear (e.g., one novel object, two known competitor objects and the novel word *flamingo*). Just like children, models have internal representations that change with learning which in turn give rise to certain

behaviours. In sharp contrast to children, however, we can inspect these representations as they develop over time to examine the relationship between internal representation and external behaviour. Critically, because a model's cognitive mechanisms are explicitly defined, by examining the computations which drive representational change in the model, we can build an explicit account of the mechanisms that explain cognitive development in the child (Westermann & Mareschal, 2012). In all models, simplification is needed to render the work feasible. Thus, there are differences between learning in robotic simulations and learning in humans. However, these differences are essential to theory development: simplification can tell us which components of a system are necessary for capturing a given behaviour (McClelland, 2009; Morse & Cangelosi, in press). Thus, simulations offer a unique opportunity for developing explicit, mechanistic theories of cognitive processes, which make clear predictions for subsequent empirical testing.

The goal of the current studies was to examine the potential mechanisms driving children's behaviour in two word learning tasks. Recent work in computational developmental psychology suggests that the body's location in space plays an important role in young children's early word learning (Morse, Benitez, Belpaeme, Cangelosi, & Smith, 2015; Samuelson et al., 2011). These studies suggest that using an embodied system is important to understand the mechanisms driving word learning. Thus, we chose to explore these mechanisms using iCub, a developmental robot designed specifically as an embodied platform for developmental research (Cangelosi et al., 2015; Metta et al., 2010). The following sections describe two robotic replications of recent empirical studies designed to explore the mechanisms thought to be at play in referent selection and word learning; specifically competition from non-target objects and referent novelty. Finally, we

discuss the implications of the modelling work for our understanding of referent selection and word learning in children.

### **Experiment 1:**

#### **The effect of competition on referent selection and word learning**

##### **Target empirical data**

With the goal of narrowing down the possible mechanisms underlying ME-type behaviour, a recent empirical study directly tested the predictions of the N3C and ME accounts. Horst, Scott and Pollard (Horst, Scott, & Pollard, 2010; henceforth HSP) explored which of these accounts best explained children's word learning by manipulating the number of known competitor objects children saw during referent selection. N3C predicts that the number of known competitors should not affect word learning, because referent selection involves simply mapping novelty to novelty: on this account, competitors are irrelevant. In contrast, ME predicts that increasing the number of competitors present during referent selection will make the task more difficult, because children must rule out all known competitors before making the novel label-novel object mapping (see also Halberda, 2006).

**Design and procedure.** To test these predictions, HSP presented 36 30-month-old children with referent selection trials consisting of an array of 3D age-appropriate toys, one of which was novel and the rest of which were known to children. The number of competitor objects seen during referent selection varied between conditions: trials consisted of a novel object and two, three, or four known competitors (see Fig. 1 for an example referent selection trial). In the two-competitor condition, for example, a trial might consist of a plastic cone with multicoloured strings attached to it, a small plastic horse, and a small plastic block. In the four-

competitor condition, a trial array might consist of the cone, the horse, the block, a spoon and a toy car. All other aspects of the design were held constant.

[FIGURE 1 ABOUT HERE]

During the referent selection phase, children were presented with four sets of objects, across eight referent selection trials. Each novel object was presented twice and served as a target once. On each trial, children were allowed to look at the objects for three seconds before being asked to select either a known or the novel object (e.g., known trial: *Can you show me the car?*; novel trial: *Can you show me the fode?*; target objects were named five times; thus, two trials per set.) Children therefore had an equal amount of experience with each novel target during referent selection, which is critical to enable a robust test of word learning (Axelsson & Horst, 2013). After referent selection children were presented with four test trials (see Fig. 1). Each of the four novel targets appeared on every test trial, and children were asked for each object in turn. If children had retained the novel label-object associations formed during the referent selection phase, then they should pick the target object at levels greater than expected by chance.

**Results.** In line with existing studies, children were very good at referent selection, performing significantly above chance on both known and novel trials regardless of the number of competitors present. However, only children in the two-competitor condition retained novel labels at levels greater than expected by chance, and did so significantly more reliably than children in the three- and four-competitor conditions (see black bars, Fig. 3). An analysis of reaction times during referent selection revealed that children in the two-competitor condition selected novel objects marginally faster than children in the four-competitor condition. Thus, longer reaction times during referent selection were related to poorer word learning. The authors

reasoned that the disambiguation task in the two-competitor condition was less onerous than in the other conditions. Put differently, these children only had to identify two known objects before mapping the novel label to the novel referent – a quicker and simpler undertaking than identifying three or four known objects.

**Paying attention to what an object is *not*.** The authors concluded that referent selection – and eventual word learning – involves paying attention to competitor objects in order to establish what the referent is not, as well as paying attention to the novel object to establish what it is (see also Fitneva & Christiansen, 2011; Zosh, Brinster, & Halberda, 2013). On this account, referent selection is influenced not only by novelty, but also by knowing the names of the competitor objects; subsequent word learning is therefore the product of learning which associations are correct (e.g., novel object-*fode*), but also of learning which associations are wrong (e.g., cow-*fode*; see also McMurray et al., 2012). The implication, therefore, is that word learning emerges from the interaction of multiple timescales of development: in-the-moment referent selection, medium-term cross-situational learning, and long-term vocabulary learning (McMurray et al., 2012). However, for a full understanding of word learning it is critical that we understand not just *what* children do, but also *how*. In the following simulations we make these mechanisms explicit by using a developmental robotic system (*iCub*; Metta et al., 2010) to implement a connectionist architecture (*Epigenetic Robotics Architecture*; Morse, de Greeff, Belpeame, & Cangelosi, 2010).

### **The iCub and the Epigenetic Robotics Architecture**

iCub's design reflects the approximate physical proportions of a 3-year-old child. iCub has 53 bodily degrees of freedom (neck: 3; eyes: 3; arms: 14; hand: 18; leg: 12, torso: 3), and sensors (e.g., cameras, microphones), which encode a range of naturalistic perceptual input, approximating young children's perceptual

environments (see Fig. 2). Like children, then, iCub integrates visual, auditory, tactile and proprioceptive information to generate behaviour, for example auditory and visual information in a word learning task (although which modalities contribute to a given simulation are decided *a priori* by the modeller). Thus far, iCub has captured a range of developmental phenomena, for example motor development (Tikhanoff, Cangelosi, & Metta, 2011), visuomotor development (Shaw, Law, & Lee, 2014), intrinsically motivated exploration (Maestre, Cully, Gonzales, & Doncieux, 2015), affordance-based verb learning (Marocco, Cangelosi, Fischer, & Belpaeme, 2010), and spatially-grounded noun learning (Morse et al., 2015; for a review see Cangelosi & Schlesinger, 2015).

[FIGURE 2 ABOUT HERE]

A version of the Epigenetic Robotics Architecture (*ERA*) served as the architecture in both the current simulations, as depicted in Figure 3. The ERA consists of a network of Self-Organising Maps (*SOMs*; Kohonen, 1998): connectionist networks that reorganise their internal structure based on a winner-takes-all response to input stimuli. At the end of learning, SOMs reflect the structure of the input in their own topological structure; that is, neurons that are close together in the network fire in response to perceptually similar stimuli (e.g., the colours red and pink). SOMs naturally lend themselves to categorisation of complex naturalistic stimuli such as inputs generated by iCub's sensors.

The model comprises two visual SOMs that receive processed video information from iCub's cameras. One map receives an HSV (hue, saturation, value; Alvy Ray, 1978) spectrogram of each object in view and so represents colour, and the other receives shape information about each object (e.g., circleness, squareness, convexity, elongation; Montesano, Lopes, Bernardino, & Santos-Victor, 2008).

Speech recognition, via the commercial software Dragon Dictate™, is used to provide speech-to-text input for the words, where each word dynamically activates a single unit in a label field. The visual SOMs are bidirectionally coupled to the field of label inputs via Hebbian-like links to form a dynamic spreading activation network.

Objects that are primed cause iCub to look at them or to reach and point to them. A detailed discussion of the ERA is available in Morse et al. (2010). The model is robust to parameter variation. Higher (or lower) learning rates, for example, strengthen and weaken the response percentage globally, preserving the between group effects. However, a summary of the parameters relevant to this particular implementation is provided in the Appendix.

[FIGURE 3 ABOUT HERE]

For an object in a particular region, colour information is extracted by determining the location in HSV colour space of each pixel in that region. Ignoring the white background of the table, pixels with a saturation value greater than a threshold of 0.2 are allocated to one of 36 bins each representing 10 degrees of the 360 degree HSV colour continuum, which generates a histogram-like colour profile for each object. Each object profile is unique and based on the entire range of the colour SOM. Thus, the model takes into account differences between uniformly and multicoloured objects.

### **Simulating Referent Selection and Word Learning in a Robotic System**

**Design and Procedure.** The procedure in the robot experiment was kept as close as possible to the procedure in the empirical task. As in the empirical study, the experiment was run 12 times per condition and trial order and counterbalancing were the same. The robot was initially provided with background training to simulate

infants' everyday experience with objects prior to the onset of word learning (see Appendix).

***Simulating children's known vocabulary.*** To simulate children's existing vocabularies, we taught the robot a "familiar" vocabulary in an initial training session (not including the novel words used in the subsequent experiment). The SOMs were provided with object and label input for the 18 competitor objects it would encounter during the referent selection phase. Based on recent empirical work demonstrating that individual objects tend to dominate infants' visual fields during word learning (Smith, Yu, & Pereira, 2011), the experimenter placed each object centrally in the robot's field of vision on a white surface and allowed the SOMs to settle – equivalent to allowing children to look at objects before providing the target label. Once the SOMs had settled (that is, once the system had formed a representation of that object; approximately 3s), the experimenter provided the label SOM with the appropriate input, keeping the object in view (reflecting the ostensive labelling shown to facilitate word learning in children; Axelsson et al., 2012). Each object received 20 unambiguous labelling events. Thus, the robot began the experiment with a robust known vocabulary. Clearly, this vocabulary is substantially smaller than children in the empirical study, who had a mean productive vocabulary of 468.92 words according to a UK adaptation of the widely-used MacArthur-Bates Communicative Development Inventory (Fenson et al., 1993; Klee, Marr, Robertson & Harrison, 2001). We made this assumption to render the task tractable, however the relationship between vocabulary and performance in word learning tasks is the focus of existing research (e.g., Borovsky, Ellis, Evans, & Elman, 2015; Perry & Samuelson, 2011; Samuelson, 2002)

***Cross-situational learning and referent selection.*** The robot was presented with the same referent selection and test trials as children in the empirical task, again across two-, three- and four-competitor conditions. During referent selection the robot was presented with four sets of objects on a white tabletop via eight referent selection trials. Each set consisted of a novel object and two, three or four known competitors selected from the pre-trained set (see Fig. 1 for an example referent selection trial). As in HSP, object locations and trial order were pseudorandomised across trials. Thus, the same set of objects was never presented on successive trials, known/novel trials occurred no more than twice in succession and each novel label/object pair was encountered in first, second, third or last position equally often. All objects were placed in the robot's field of vision and the SOMs were allowed to settle (intended to reflect the three-second pause before labelling in the empirical study). Then, the experimenter labelled the object five times with either a known (pretrained) or novel label. Following labelling, the robot moved its head to centre its field of vision on each object in turn, activating a node in the label SOM. If the SOM activated the appropriate label node for the target object, the robot's response was scored as correct, and if not, the robot's response was scored as incorrect. For example, on a known trial, activation of the *horse* node in response to the horse object would be scored correct, and activation of the *yok* label would be scored as incorrect. Each novel object was presented twice and served as a target once.

***Testing word learning.*** After referent selection the robot was presented with four test trials that proceeded in an identical manner to the referent selection trials. As in the empirical study, each of the four novel targets appeared on each test trial, and each served as the target on one trial. If the model had learned and retained novel label-object associations during the referent selection phase, then it should activate

the appropriate label node in response to each novel object at levels greater than expected by chance. Again, retention was scored by monitoring node activation.

**Results.** Results from the referent selection trials are depicted in Figure 4. In line with children, the model successfully mapped known labels to known objects (100% correct on all known trials). Critically, the robot also mapped novel labels to novel objects, and did so at levels greater than expected by chance (two-competitor:  $t(11) = 19.53, p < .0001, d = 7.84$ ; three-competitor:  $t(11) = 9.93, p < .0001, d = 3.33$ ; four-competitor:  $t(11) = 8.94, p < .0001, d = 2.83$ . Note that chance = 0.33, 0.25 and 0.20 in the two-, three- and four-competitor conditions, respectively, and all  $t$ -tests reported are two-tailed). Thus, the model captured HSP's referent selection results, mapping novel labels to novel objects without explicit instruction or the ability to reason explicitly about its choices.

[FIGURE 4 ABOUT HERE]

Results from the test trials are depicted in Figure 5. Here, the model retained novel label-object mappings at levels greater than expected by chance (0.25) in the two-competitor condition only (two-competitor:  $t(11) = 8.86, p < .0001, d = 5.34$ ; three-competitor,  $t(11) = 1.39, ns., d = 0.84$ ; four-competitor,  $t(11) = 0.56, ns., d = 0.34$ ). This is the same pattern of results observed in the empirical study. We submitted the model's proportion of correct choices on test trials to an ANOVA with condition (two-competitor, three-competitor, four-competitor) as a between-subjects factor. The effect of condition was significant,  $F(1,34) = 34.19, p < .0001, \eta^2_G = 0.50$ ; planned comparisons revealed that the model made significantly more correct choices in the two-competitor condition than in the three- or four- competitor conditions (both  $ps < .0001$ ). The model therefore also captured HSP's retention results.

[FIGURE 5 ABOUT HERE]

To compare the robot and empirical data, we submitted proportion correct choices for both datasets to an omnibus ANOVA with data (empirical, robot) and condition (2-competitor, 3-competitor, 4-competitor) as between-subjects factors and trial (referent selection, retention) as a within-subjects factor, and their associated interactions. Results are reported in Table 1; critically, as highlighted in bold, whether data were empirical or robotic had no effect on responses. Thus, word learning in this embodied simulation can emerge from the interaction between long-term vocabulary learning, and in-the-moment referent selection, without the need for top-down reasoning.

[TABLE 1 ABOUT HERE]

### **Discussion**

In Experiment 1 we used an embodied neural network model (Metta et al., 2010; Morse et al., 2010) to replicate 30-month-old children's behaviour in a word learning task (Horst et al., 2010). Children in the empirical task were presented with a referent selection phase in which they were asked to select novel objects in response to novel labels, from an array in which all other objects were known. When tested on retention of novel label-object mappings, only children who had initially encountered novel objects alongside two competitors successfully retained those mappings; children who saw more competitors did not retain novel labels. We taught the model a known vocabulary and then presented it with a maximally similar task. The model correctly mapped known and novel labels to target objects during referent selection – that is, it exhibited the same in-the-moment referent selection as the children. At test, only when novel objects had initially been encountered alongside two (but not three or

four) competitors did the model successfully retain label-object mappings, again replicating children's behaviour in HSP.

### **“Mutual exclusivity” can emerge from simple associations**

By implementing the ERA in iCub we demonstrate that a behaviour some have argued to depend on complex, top-down inferential reasoning (Markman, 1990; Markman, 1994) can emerge from low-level associative processes. Word learning in the two-competitor condition demonstrates that simply reinforcing newly-formed label/object associations via cross-situational learning allows these associations to be reactivated without supporting context, that is, known competitor objects (cf. Smith & Yu, 2008). Finally, we capture the effect of competition seen in the empirical task, demonstrating that the complex world learning phenomena seen in HSP can emerge from the simple associative mechanisms governing the model's behaviour. We return in detail to these mechanisms in the General Discussion.

Overall, then, Experiment 1 demonstrated that ME-type behaviour can arise from simple associative learning across situations without the need for a top-down, metacognitive reasoning system. Importantly, in both Experiment 1 and HSP novelty was controlled: children and the model encountered every novel object and every novel label an equal number of times during the referent selection phase. Indeed, if novelty had been driving behaviour during this phase, the novel object would have been chosen even in response to known labels. Clearly, novel objects were not so salient to children or iCub that ME-type behaviour during referent selection was overridden. Nonetheless, children's novelty preference is well-documented in a range of paradigms (e.g., Fantz, 1964; Golinkoff, Ma, Song, & Hirsh-Pasek, 2013; Houston-Price & Nakai, 2004), and it remains possible that increased attention to novelty may affect children's referent selection when relative novelty between referents is

manipulated. Horst, Samuelson, Kucker & McMurray (2011; henceforth HSKM) therefore explored the extent to which relative referent novelty affects children's choices during referent selection.

### **Experiment 2:**

#### **The effect of novelty on referent selection**

##### **Target empirical data**

HSKM presented 12 24-month-old children with referent selection trials consisting of three known or three novel objects, presented in the same manner as in HSP with similar stimuli. Critically, children had been allowed to play with a subset of the novel objects before the experiment began. Thus, on novel trials, one “supernovel” object had never been seen before, and none of the three objects had been labelled. On each trial, the experimenter asked the child to retrieve an object with a known label on known trials (e.g., *Which one is the cow?*) or a novel label on novel trials (e.g., *Which one is the fode?*). If the small amount of engagement with the objects before referent selection was sufficient to render them “familiar,” on novel trials children should systematically reject these just-seen objects as candidate referents and map the label to the supernovel object. However, if object novelty plays no part in referent selection, children should respond at chance levels.

Again, as expected, children selected the known object in response to the known label at rates greater than expected by chance. Critically, on novel object trials, children systematically chose the supernovel object. Thus, just two minutes' experience with objects before referent selection was sufficient to trigger children's novelty preference. Critically, HSKM demonstrated that referent selection, hitherto assumed to be a linguistic mechanism, is affected by nonlinguistic factors, and specifically, object salience as mediated by novelty. Nonetheless, the question of *how*

children solved the novel referent selection trials remains unanswered – is this behaviour the result of children explicitly reasoning about what they had or had not previously seen, or could it emerge from low-level associative learning? We explored this by using the same robotic system in a similar task to replicate HSKM's empirical results.

**Design and procedure.**

Model architecture and model parameter were the same as in Experiment 1.

*Simulating children's known vocabulary.* We pre-trained the robot with 12 known objects using the same procedure as in Experiment 1.

*Cross-situational learning and referent selection.* Procedure on referent selection trials was identical to the procedure used in Experiment 1. The experimental design followed HSKM, and we ran the simulation 12 times to reflect the 12 participants in the empirical study. Specifically, over the course of the experiment, the robot saw 12 known and 18 novel toy objects. However, before referent selection the experimenter prefamiliarised the robot with eight of the objects by presenting each object centrally in the robot's field of vision on a white surface and allowing the SOMs to settle, in line with the two-minute prefamiliarisation phase in HSKM. Critically, no object was labelled during this phase. During referent selection the robot was presented with four known and eight novel trials. Each known trial set consisted of three different objects from the set of pre-trained known objects. Each novel trial consisted of two prefamiliarised novel objects and one supernovel object labelled with a different novel word. Each supernovel object appeared once, and pre-familiarised objects were counterbalanced such that each trial consisted of a different combination of objects.

**Results.** Results are depicted in Figure 6. As anticipated, the model successfully mapped known labels to known objects ( $t(11) = 16.43, p < .0001, d = 6.51$ ), replicating the results of the known referent selection trials in the current Experiment 1. Critically, on novel trials, the model mapped novel labels to supernovel objects at rates greater than expected by chance ( $t(11) = 15.27, p < .0001, d = 6.36$ ). To compare the robot and empirical data we submitted proportion correct choices for both datasets to an omnibus ANOVA with data (empirical, robot) as a between-subjects factor, trial (known, novel) as a within-subjects factor, and a data-by-trial interaction term. Overall, there were more correct responses on known than novel referent selection trials. Critically, however, whether data were empirical or robotic had no effect on responses. Thus, Experiment 2 captured the empirical data presented in HSKM, and specifically, the remarkable effect of novelty on referent selection by which even brief familiarisation with novel objects prior to the referent selection task leads children – and iCub – to systematically map novel labels to never-before-seen, supernovel objects.

[FIGURE 6 ABOUT HERE]

[TABLE 2 ABOUT HERE]

### Discussion

Experiment 2 extended the findings of Experiment 1 by exploring the effect of novelty on children's referent selection reported by Horst and colleagues (HSKM; 2011). HSKM presented children with multiple three-object referent selection trials. On known trials, children systematically choose the correct target referent from an array of three known objects. On novel trials, children were asked to choose a referent

from an array of three novel objects; however, two of the novel objects were more familiar than the third, having been briefly presented to children before the task (without labelling). Children systematically mapped novel labels to the “supernovel” object. We presented the same embodied robotic system described in Experiment 1 with HSKM’s referent selection task and replicated the novelty effect seen in children’s responding – and again, without building in a mechanism for reasoning about whether it had encountered objects previously. Convergent with McMurray et al. (2012), the current studies demonstrate that apparently complex word learning behaviour can emerge over time from the dynamic interaction of multiple timescales. Specifically, long-term vocabulary acquisition facilitates medium-term cross-situational learning by supporting in-the-moment referent selection (McMurray et al., 2012). In the following section we discuss these timescales and their interactions with the mechanisms driving word learning in our system.

### **General Discussion**

The current simulations capture children’s ability to learn words in the widely-used referent selection tasks described by HSP and HSKM (Faubel & Schoner, 2008; although for a pilot study, see Twomey, Horst, & Morse, 2013). Like referent selection and word learning in children, these phenomena in the current simulations are affected by competition and novelty during initial disambiguation. As such, this work supports associative accounts of word learning: referent selection and word learning in the current simulations can emerge bottom-up from simple associations, without recourse to a complex, top-down reasoning system (Smith, 2000). Critically, because computational models in general make mechanisms explicit (Westermann & Mareschal, 2012), our embodied neural network simulations not only capture what

children in the empirical tasks did, but also suggest how they did it. As with all computational and robotic models, however, this account needs to be tested: if the assumptions of this model reflect the mechanisms underlying learning in human children, it should be possible to use an experimental task to capture the predictions our model makes about children's behaviour. These predictions and related empirical work are discussed in more detail below.

### **Referent selection via cross-situational associative learning.**

The referent selection task in Experiment 1 supports the ME account of the processes underlying children's referent selection and word learning, demonstrating that referent selection involves attention to known competitor objects as well mapping the novel label to the novel object (see also Fitneva & Christiansen, 2011; Zosh et al., 2013). Specifically, because inhibition from the strong known label-known object connections prevents the formation of new known label-novel object connections, meaning that the only mapping not subject to inhibition on novel label trials is that between the novel label and the novel object. Thus, in line with McMurray et al. (2012), Experiment 1 points to an account of referent selection not as a bias towards novel, unlabelled referents, but as a bias away from known, labelled referents.

Experiment 2 elaborates this account. Here, we see that referent selection is not a purely linguistic phenomenon, but rather an interplay between linguistically- and visually-mediated information – when labelling is controlled for, the novelty of the objects themselves affects whether children (or the model) learn the names for those objects (Kucker & Samuelson, 2011). Further, these data make a strong prediction for future empirical work. On a given novel trial, the model's mapping of novel labels to the supernovel objects hinged critically on the characteristics of the three potential targets – only when all three objects shared some visual characteristic

(e.g., colour) were label/just-seen object mappings inhibited sufficiently for the model to map the novel name to the novel referents. Thus, Experiment 2 makes the novel, testable prediction that referent selection should be affected by the degree of perceptual similarity between objects. An empirical test of this prediction is underway.

### **Multiple timescales and the effect of embodiment**

The current work shows that word learning can emerge from the interaction of knowledge across multiple timescales (cf. Horst et al., 2006; Smith, Colunga, & Yoshida, 2010; Thelen & Smith, 1994). The longest timescale relates to vocabulary. Participants began these experiments with a pre-existing vocabulary; in the model's case, a pre-trained set of robust label-object mappings, and in children's case, a set of known words built up over the preceding 24 months. The intermediate timescale relates to cross-situational learning during the course of the study. The robot learns words by forming and strengthening associations between the label node and the representations in the colour and shape SOMs on every novel label trial, while children formed and strengthened associations between novel labels and novel objects by completing multiple referent selection trials. The in-the-moment timescale relates to the disambiguation task itself. Here, long-term vocabulary knowledge interacts with online information in the visual scene: labels for the known objects are activated when their referents are recognised. Using a robotic system allows us to watch this mechanism unfold, and shows that label activation (in the absence of hearing that label) is critical for this ME-type behaviour. While 18-month-old infants have been shown in a visual priming task to activate labels when their referents are seen in silence (Mani & Plunkett, 2010), testing this prediction in a word learning task remains an important challenge for future empirical work.

Importantly, the current work illustrates an additional timescale: the micro-level temporal dynamics of activation decay during the time taken to look from one object to another. During referent selection the robot was provided with a label, which caused a spike of activation in the relevant label node. The robot then “looked” at each object in turn to establish whether the current label was associated with each object. Note that because “looking” involved a physical turn of the robot’s head, looking at all objects took less time in the two-competitor condition than in the four-competitor condition (this was also the case for children in the empirical study). In addition, activation in the label node decayed over time. Thus, since learning was applied at the end of the trial, label-object mappings were weaker in the three- and four-competitor conditions than in the two-competitor condition; so weak, in fact, that the model was unable to reactivate them sufficiently to form correct mappings at test. In line with existing embodied explorations of word learning, which show that the spatial orientation of the body can affect the formation of label-object mappings (Morse et al., 2015; Samuelson et al., 2011), the current work demonstrates that the layout of the task environment entrains the spatial location of the robot’s head, which in turn affects the amount of time required to encode all objects in the array. Our model therefore predicts that children’s performance in similar tasks can be boosted or impaired by reducing or increasing the amount of time it takes to scan all possible referents of the novel words by increasing the spatial distance between them. Critically, this prediction would not have emerged from a non-embodied system without building in a timing mechanism *a priori* (e.g., look at the first object at 2s, the second at 4s, and the third at 5s). Thus, this prediction emerged “for free” from the interaction between the SOMs’ learning mechanism and the iCub’s body. Capturing this prediction empirically would strongly support the current hypothesis that the

body affects word learning via micro-level temporal interactions. An experiment with two-year-old children is planned to provide the critical test of this account.

Both the empirical and robotic studies discussed here involved a simplified learning environment and a highly structured task relative to the real-world learning environment children experience outside the lab (Horst & Simmering, 2014; Oudeyer & Smith, in press). It is therefore possible that while low-level associative learning can account for the results of HSP and HSKM, additional information such as sociopragmatic cues (e.g., Brooks & Meltzoff, 2005; Moore, Mueller, Kaminski, & Tomasello, 2015; Schulze & Tomasello, 2015), distributional information (e.g., Gillette, Gleitman, Gleitman, & Lederer, 1999; Medina, Snedeker, Trueswell, & Gleitman, 2011; Twomey, Chang, & Ambridge, 2014; Yuan, Fisher, & Snedeker, 2012) or existing semantic category representations (e.g., Borovsky et al., 2015; Borovsky & Elman, 2006) may play a part in iCub's – and children's – word learning in more complex environments, pointing to further fruitful work in the rapidly-expanding field of developmental robotics. Overall, however, the current studies represent the first full experimental replication of the results of a widely-used word learning paradigm using an embodied robotic system. As such, they contribute in a broader sense to the emerging interdisciplinary literature in the cognitive sciences that in recent years has begun to apply mathematical, computational and robotic innovations to some of the decades-old enigmas of developmental psychology. In parallel, the current work helps us build an explicit account of the complex and subtle temporal, environmental and physiological interactions that drive word learning and cognitive development.

### References

- Alvy Ray, S. (1978). Color gamut transform pairs. *SIGGRAPH Computer Graphics*, 12(3), 12–19. DOI: 10.1145/965139.807361
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3. DOI: 10.3389/fpsyg.2012.00088
- Axelsson, E. L., & Horst, J. S. (2013). Testing a word is not a test of word learning. *Acta Psychologica*, 144(2), 264-268. DOI: 10.1016/j.actpsy.2013.07.002
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, 20(02), 395–418. DOI:10.1017/S0305000900008345
- Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, 126(1), 39–53. DOI: 10.1016/j.cognition.2012.08.008
- Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2015). Lexical leverage: category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*. DOI: 10.1111/desc.12343
- Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33(04), 759–790. DOI: 10.1017/S0305000906007574
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543. DOI: 10.1111/j.1467-7687.2005.00445.x

- Cangelosi, A., Schlesinger, M., & Smith, L. B. (2015). *Developmental robotics: From babies to robots*. MIT Press.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17–29.
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17(2), 417–431. DOI: 10.1017/S0305000900013842
- Clark, E. V. (1995). *The Lexicon in Acquisition*. Cambridge University Press.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science*, 146(668-670). DOI: 10.1126/science.146.3644.668
- Faubel, C., & Schoner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks*, 21(4), 562–576. DOI: 10.1016/j.neunet.2008.03.007
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6), 1017–1063. DOI: 10.1111/j.1551-6709.2010.01104.x
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular Publishing Group.
- Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, 35(2), 367–380. DOI: 10.1111/j.1551-6709.2010.01156.x
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2), 135–176. DOI: 10.1016/S0010-0277(99)00036-0

- Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, 33(4), 709–738. DOI: 10.1111/j.1551-6709.2009.01026.x
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28(1), 99–108. DOI: 10.1037/0012-1649.28.1.99
- Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the Intermodal Preferential Looking Paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, 8(3), 316–339. DOI: 10.1177/1745691613484936
- Gurney, K. (1997). *An introduction to neural networks*. CRC Press.
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), 310–44. DOI: 10.1016/j.cogpsych.2006.04.003
- Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*.
- Horst, J. S., McMurray, B. & Samuelson L. K. (2006). Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture. *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum & Associates.
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. DOI: 10.1080/15250000701795598
- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, 118(2), 234–244. DOI: 10.1016/j.cognition.2010.10.015

- Horst, J. S., Scott, E. J., & Pollard, J. P. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706–713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Horst, J. S., & Simmering, V. R. (2015). Category learning in a dynamic world. *Frontiers in Psychology*, *6*. DOI: 10.3389/fpsyg.2015.00046
- Horst, J. S., McMurray, B. & Samuelson L. K. (2006). Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture. *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum & Associates
- Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341–348. DOI: 10.1002/icd.364
- Houston-Price, C., Plunkett, K., & Harris, P. (2005). “Word-learning wizardry” at 1;6. *Journal of Child Language*, *32*(1), 175–189. DOI: 10.1017/S0305000904006610
- Kohonen, T. (1998). The Self-Organizing Map, a possible model of brain maps. *Brain and Values*, 207–236 568.
- Klee, T., Marr, C., Robertson, E., & Harrison, C. (1999). The MacArthur Communicative Development Inventory: Toddler (British English adaptation). Newcastle, England: Newcastle University Press.
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. DOI: 10.1111/cdep.12110

- Kucker, S. C., & Samuelson, L. K. (2011). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy, 17*(3), 295–323. DOI: 10.1111/j.1532-7078.2011.00081.x
- Maestre, C., Cully, A., Gonzales, C., & Doncieux, S. (2015). Bootstrapping interactions with objects from raw sensorimotor data: a Novelty Search based approach. In *Proceedings of the 5th International Conference on Development and Learning and on Epigenetic Robotics*. Providence, RI.
- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science, 21*(7), 908–913. DOI: 10.1177/0956797610373371
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*(1), 57–77. DOI: 10.1207/s15516709cog1401\_4
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua, 92*(1-4), 199–227. DOI: 10.1016/0024-3841(94)90342-5
- Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*(1), 1–27. DOI: 10.1016/0010-0285(84)90002-1
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*(2), 121–157. DOI: 10.1016/0010-0285(88)90017-5
- Marocco, D., Cangelosi, A., Fischer, K., & Belpaeme, T. (2010). Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Frontiers in Neurorobotics, 4*. DOI: 10.3389/fnbot.2010.00007

- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in Mutual Exclusivity. *Infancy, 14*(1), 60–76. DOI: 10.1080/15250000802569702
- McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1*(1), 11–38. DOI: 10.1111/j.1756-8765.2008.01003.x
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375. DOI: 10.1037/0033-295X.88.5.375
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 83877. DOI: 10.1037/a0029872
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences, 108*(22), 9014–9019. DOI: 10.1073/pnas.1105040108
- Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, i*–129. DO: 10.2307/1166130
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name Nameless Category (N3c) principle. *Child Development, 65*(6), 1646–1662. DOI: 10.2307/1131285
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C., Rosander, K., Lopes, M., Santos-victor, J., Bernardino, A., & Montesano, L. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks Neural Networks, 23*(8-9), 1125–1134. DOI: 10.1016/j.neunet.2

010.08.010

- Montesano, L., Lopes, M., Bernardino, A., & Santos-Victor, J. (2008). Learning object affordances: from sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, *24*(1), 15–26. DOI: 10.1109/TRO.2007.914848
- Moore, R., Mueller, B., Kaminski, J., & Tomasello, M. (2015). Two-year-old children but not domestic dogs understand communicative intentions without language, gestures, or gaze. *Developmental Science*, *18*(2), 232–242. DOI: 10.1111/desc.12206
- Morse, A. F., Benitez, V. L., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2015). Posture affects how robots and infants map words to objects. *PLoS ONE*, *10*(3), e0116012. DOI: 10.1371/journal.pone.0116012
- Morse, A. F. & Cangelosi, A. (in press). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cognitive Science*.
- Morse, A. F., de Greeff, J., Belpaeme, T., & Cangelosi, A. (2010). Epigenetic Robotics Architecture (ERA). *IEEE Transactions on Autonomous Mental Development*, *2*(4), 325–339. DOI: 10.1109/TAMD.2010.2087020
- Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science*, *7*(2), 141–8. DOI: 10.1111/j.1467-7687.2004.00331.
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*. DOI: 10.3389/fpsyg.2012.00041
- Oudeyer, P.-Y., & Smith, L. (in press). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*. Retrieved from <http://www.pyoudeyer.com/OudeyerSmithTopicsCogSci14.pdf>

- Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, 2, 345. DOI: 10.3389/fpsyg.2011.00345
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016–1037. DOI: 10.1037//0012-1649.38.6.1016
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS One*, 6(12), e28095. DOI: 10.1371/journal.pone.0028095
- Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, 136, 91–98. DOI: 10.1016/j.cognition.2014.11.036
- Shaw, P., Law, J., & Lee, M. (2014). A comparison of learning strategies for biologically constrained development of gaze control on an icub robot. *Autonomous Robots*, 37(1), 97–110. DOI: 10.1007/s10514-013-9378-4
- Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, ... G. Hollich (Eds.), *Becoming a Word Learner: A Debate on Lexical Acquisition* (pp. 51–80). New York: Oxford University Press.
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, 34(7), 1287–1314. DOI: 10.1111/j.1551-6709.2010.01130.x
- Smith, L. B., Jones, S. S., Yoshida, H., & Colunga, E. (2003). Whose DAM account? Attentional learning explains Booth and Waxman. *Cognition*, 87(3), 209–213.

- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. DOI: 10.1016/j.cognition.2007.06.010
- Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. DOI: 10.1111/j.1467-7687.2009.00947.x
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass.: MIT Press.
- Tikhanoff, V., Cangelosi, A., & Metta, G. (2011). Integration of speech and action in humanoid robots: iCub simulation experiments. *Autonomous Mental Development, IEEE Transactions on*, *3*(1), 17–29. DOI: 10.1109/TAMD.2010.2100390
- Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*(2), 201–224. DOI: 10.1016/0885-2014(95)90009-8
- Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, *73*, 41–71. DOI: 10.1016/j.cogpsych.2014.05.001
- Twomey, K. E., Horst, J. S., & Morse, A. F. (2013). An Embodied model of young children's categorization and word learning. In L. J. Gogate & G. Hollich (Eds.), *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence* (pp. 172–196). Hershey, PA: Information Science Reference. DOI: 10.4018/978-1-4666-2973-8.ch008

- Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, 23(2), 105–122. DOI: 10.1002/icd.1824
- Westermann, G., & Mareschal, D. (2012). Mechanisms of developmental change in infant categorization. *Cognitive Development*, 27(4), 367–382. DOI: 10.1016/j.cogdev.2012.08.004
- Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120391. DOI: 10.1098/rstb.2012.0391
- Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, 83(4), 1382–1399. DOI: 10.1111/j.1467-8624.2012.01783.x
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. DOI: 10.1111/j.1467-9280.2007.01915.x
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, 21(1), 1–22. DOI: 10.3758/s13423-013-0443-y
- Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, 17(1), 20–28. DOI: .1080/10888691.2013.748420

### Appendix

Input SOMs were initialised with random connection weights. To provide an approximation of the developmental history children accumulate prior to learning their first words (that is, via encounters with multiple objects simultaneously), the 20 objects used in this study were simultaneously placed in view and the SOMs trained using standard equations 1 (SOM activity) & 2 (SOM learning rule; Kohonen, 1998):

$$(1) \quad \text{BMU} = \underset{i}{\operatorname{argmax}} \left( 1 - \sqrt{\Sigma} (a_j - w_{ij}) \right)$$

Where the Best Matching Unit (BMU;  $i$ ) is the unit whose weight vector  $w$  is closest to the current input vector  $a$ .

$$(2) \quad \Delta w_{ij} = \alpha \exp \left( -\frac{\text{dist}^2}{2\text{size}} \right) (a_j - w_{ij})$$

The weights of each unit  $j$  in the neighbourhood of the BMU are then modified to move closer to the current input vector, with changes scaled according to the distance of that unit from the BMU ( $\text{dist}$ ) in the SOM (i.e., not in terms of the input space) and neighbourhood size ( $\text{size}$ ).

As is typical of SOMs (Gurney, 1997), the neighbourhood size and learning rate ( $\alpha$ ) decrease monotonically until the neighbourhood size is 1 to allow the network to settle into a stable state, at which point both the neighbourhood size and the learning rate of the SOM are fixed to allow learning to continue at a low rate.

The two SOMs and the label field are fully connected via Hebbian-type links (Hebb, 1949; Munakata & Pfaffly, 2004), which propagate activation as in equation 3 (IAC spreading activation; cf. McClelland & Rumelhart, 1981) and learn as in equation 4 (Hebb-like learning rule):

$$(3) \quad \text{net}_i = \Sigma w_{ij} a_j + \beta \text{BMU}$$

$$\text{If } \text{net}_i > 0 \quad \Delta a_i = (\text{max} - a_i) \text{net}_i - \text{decay}(a_i - \text{rest})$$

$$\text{Else} \quad \Delta a_i = (a_i - \text{min}) \text{net}_i - \text{decay}(a_i - \text{rest})$$

The net input ( $net_i$ ) to each unit in the whole network is either

- the sum of spreading activation
- or, the sum of spreading activation plus external activation if this node happens to be the BMU of a SOM or a currently active word.

Parameter values for both robotic tasks were as follows: External Input Bias ( $\beta$ ) = 0.5; Max = 1; Min = -0.2; Decay = 0.5; Rest = -0.01.

(4) If  $a_i > 0$  OR  $a_j > 0$ :

$$\text{if } a_i a_j > 0 \quad \Delta v_{ij} = \lambda a_i a_j (1 - v_{ij})$$

$$\text{else} \quad \Delta v_{ij} = \lambda a_i a_j (1 + v_{ij})$$

$$\text{else} \quad \Delta v_{ij} = 0$$

The Hebb-like learning rule increases the strength of a weight ( $v$ ) between SOMs if both units connected by this weight are positively active, or reduces its strength if only one is positively active. This change is scaled according to the product of the units' activity and how close the weights are to 1 or -1, respectively, for positive and negative weight changes. Finally each field is fully connected by fixed inhibitory connections. The experiment reported here used the learning parameter value  $\lambda = 0.005$ .

Note that adaptive connections exist only between the SOMs and label field, while constant-valued (-0.8) inhibitory spreading activation connections exist within each SOM and within the label field.

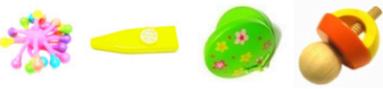
All other model parameters are the same as reported by Morse et al. (2015). The architecture itself differs in that it includes separate shape and color SOMs, but does not include a SOM receiving postural input.

*Table 1.* Results from omnibus ANOVA, Experiment 1.

Effect	<i>df</i>	<i>F</i>	<i>p</i>	$\eta^2_G$
<b>Data</b>	<b>(1,66)</b>	<b>0.021</b>	<b>0.89</b>	<b>.00010</b>
Condition	(2,66)	25.99	< .001***	.20
Trial	(2,132)	228.70	< .001***	.70
<b>Data x condition</b>	<b>(2,66)</b>	<b>0.44</b>	<b>0.65</b>	<b>00042</b>
<b>Data x trial</b>	<b>(2,132)</b>	<b>0.71</b>	<b>0.49</b>	<b>.0073</b>
Condition x trial	(4,132)	8.72	< .001***	.015
<b>Data x condition x trial</b>	<b>(4,132)</b>	<b>0.50</b>	<b>0.74</b>	<b>.010</b>

*Table 2.* Results from omnibus ANOVA, Experiment 2.

Effect	<i>df</i>	<i>F</i>	<i>p</i>	$\eta^2_G$
<b>Data</b>	<b>(1,22)</b>	<b>1.38</b>	<b>.25</b>	<b>.022</b>
Trial	(1,22)	6.05	.022	.15
<b>Data x trial</b>	<b>(1,22)</b>	<b>0.25</b>	<b>.62</b>	<b>.0071</b>

Condition	Example referent selection trial	Example test trial
2-competitor	 (e.g.) <i>Which one's the fish?</i>	 (e.g.) <i>Which one's the fode?</i>
3-competitor	 (e.g.) <i>Which one's the fode?</i>	
4-competitor	 (e.g.) <i>Which one's the car?</i>	

*Figure 1.* Example referent selection and test trials for the empirical and robot tasks.

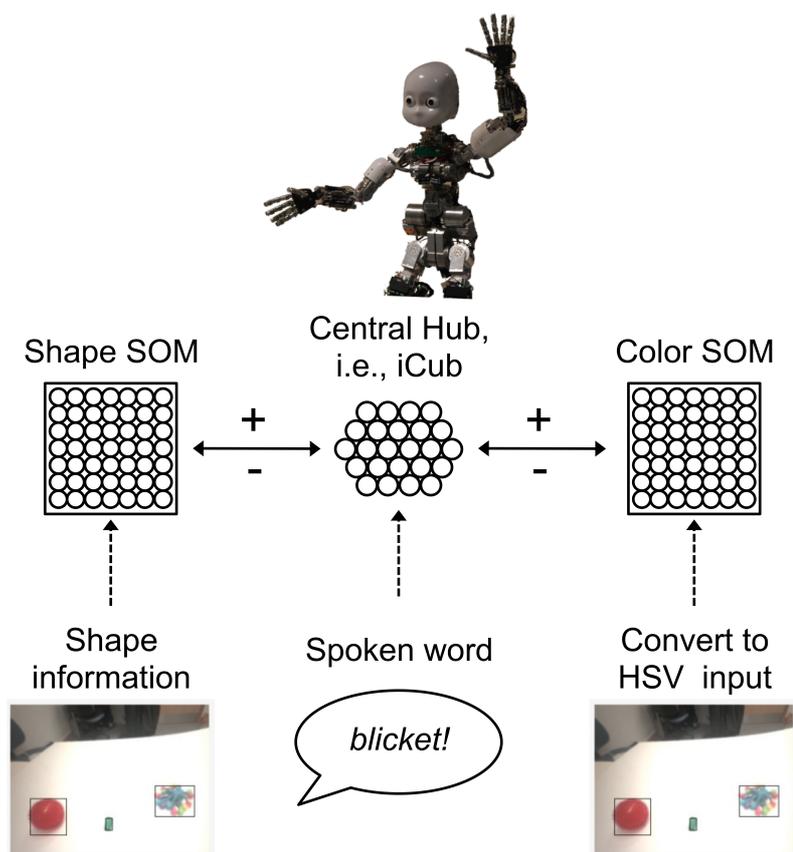
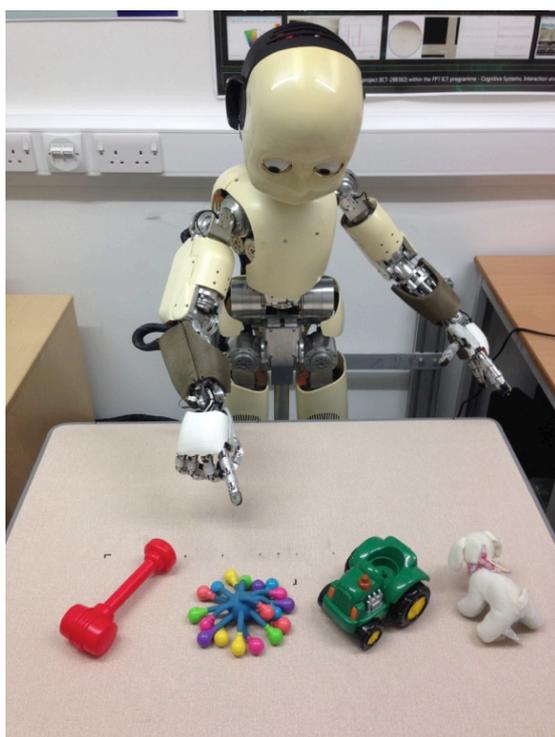


Figure 2. Model architecture



*Figure 3.* iCub during referent selection.

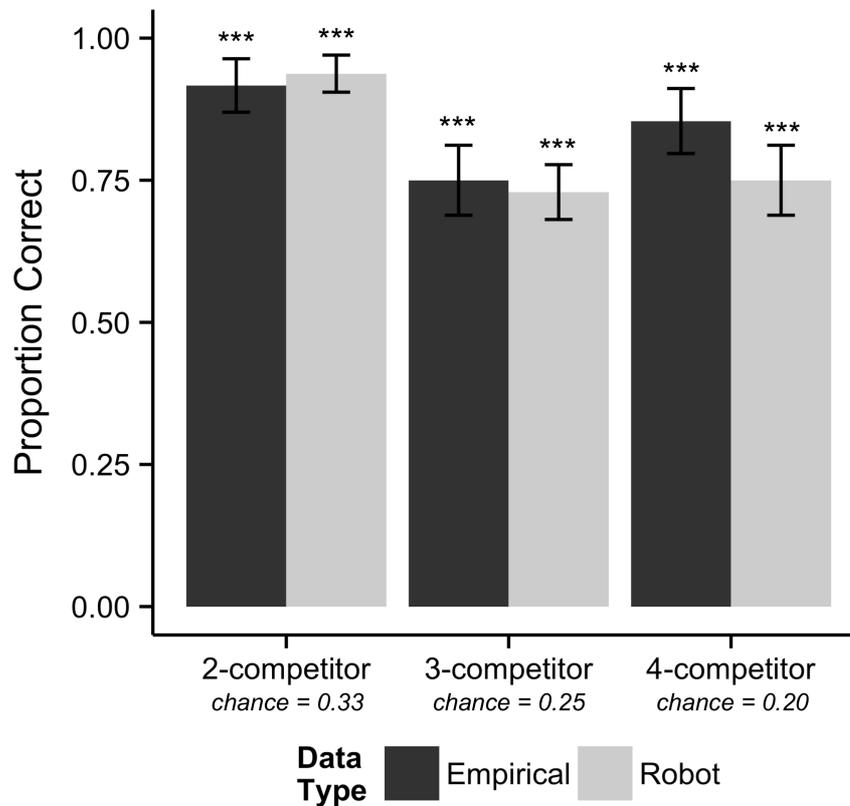
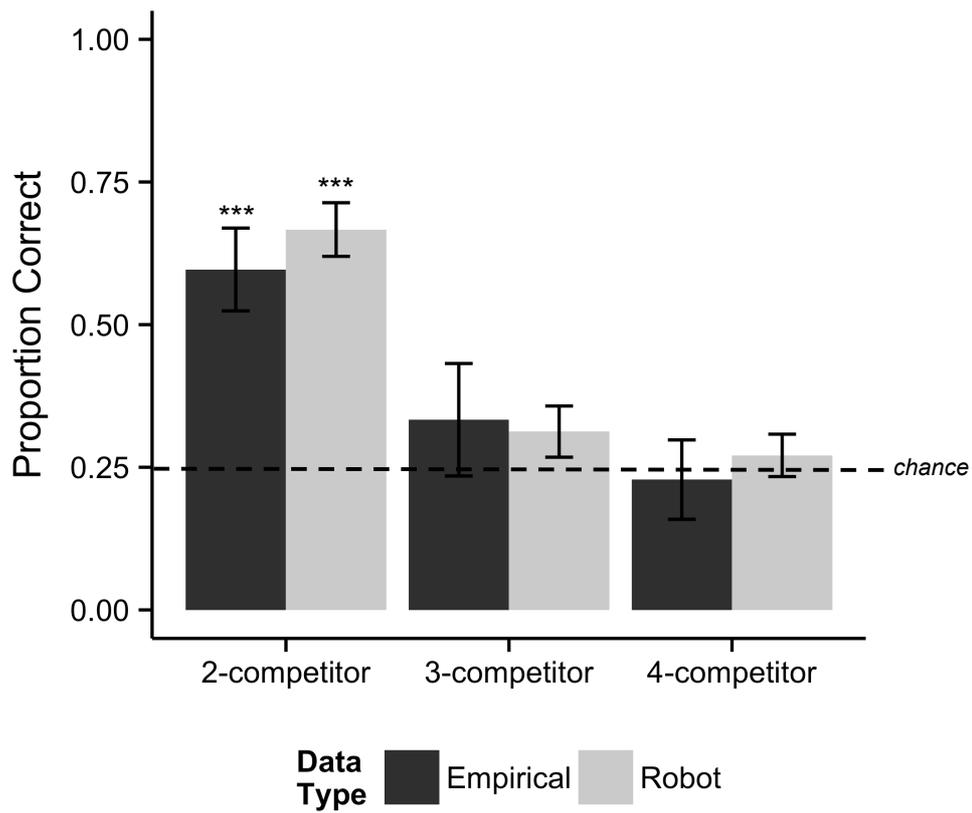
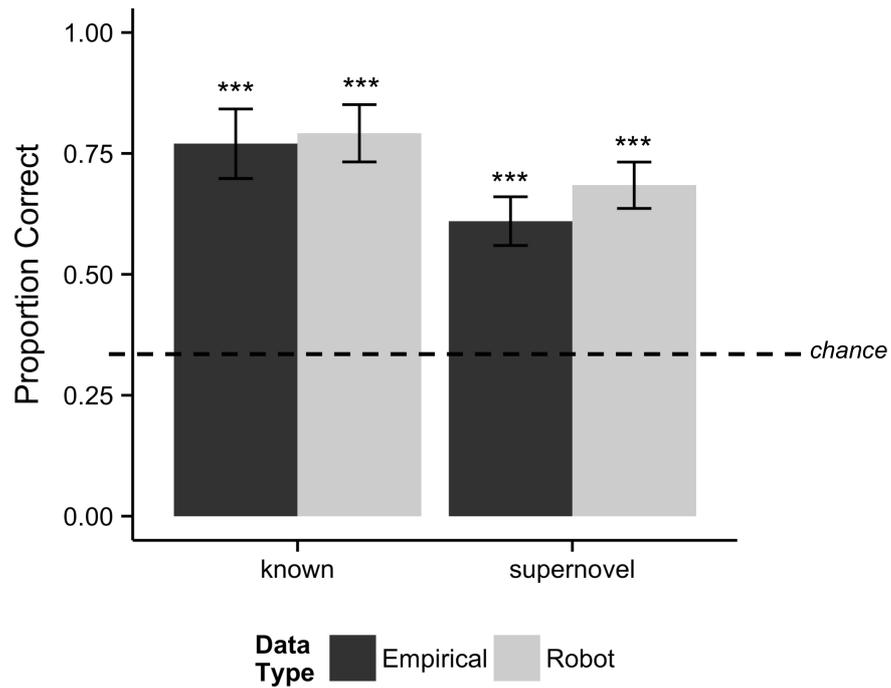


Figure 4. Proportion correct of children's (grey bars) and the model's (black bars) correct choices on novel trials during referent selection, Experiment 1. \*\*\* $p < .001$  (compared to chance, as stated in the figure).



*Figure 5.* Proportion correct of children's (grey bars) and the model's (black bars) correct choices (dark blue bars) on test trials, Experiment 1. \*\*\* $p < .001$  (compared to chance; 0.25).



*Figure 6.* Proportion correct of children's (grey bars) and the model's (black bars) correct choices on known and novel referent selection trials, Experiment 2. \*\*\* $p < .001$  (compared to chance; 0.33).