A neurocomputational model of the effect of learned labels on infants' object representations

Arthur Capelier-Mourguy (a.capelier-mourguy@lancaster.ac.uk)
Department of Psychology, Lancaster University
Lancaster, LA1 4YF United Kingdom

Katherine E. Twomey (k.twomey@lancaster.ac.uk)
Department of Psychology, Lancaster University
Lancaster, LA1 4YF United Kingdom

Gert Westermann (g.westermann@lancaster.ac.uk)
Department of Psychology, Lancaster University
Lancaster, LA1 4YF United Kingdom

Abstract

The effect of labels on nonlinguistic representations is the focus of substantial debate in the developmental literature. A recent empirical study (Twomey & Westermann, 2016) suggested that labels are incorporated into object representations, such that infants respond differently to objects for which they know a label relative to unlabeled objects. However, these empirical data cannot differentiate between two recent theories of integrated label-object representations, one of which assumes labels are features of object representations, and one which assumes labels are represented separately, but become closely associated with learning. We address this issue using a neurocomputational (auto-encoder) model to instantiate both theoretical approaches. Simulation data support an account in which labels are features of objects, with the same representational status as the objects' visual and haptic characteristics.

Keywords: connectionist model, label status, word learning

The nature of the relationship between labels and nonlinguistic representations has been the focus of recent debate. On one account, label representations are qualitatively different to object representations (Waxman & Markow, 1995). This *labels as invitations to form categories* (henceforth *label as invitations*) approach views labels as conceptual markers acting as abstract, top-down indicators of category membership, and assumes that labels are represented separately from their referents. In contrast, the *labels as features* view assumes that label representations are integrated into object representations (Gliozzi, Mayor, Hu & Plunkett, 2009; Sloutsky & Fisher, 2004). On this account, labels have no special status; rather, they contribute to object representations in the same way as other features such as shape and color. More recently, Westermann & Mareschal (2014) suggested a *compound representations* account in which labels are encoded in the same representational space as objects, and drive learning over time, but are not integrated within visual object representations. Rather, they become closely associated with object representations over learning. Importantly, although this view is superficially similar to the *labels as invitations* approach, it involves substantially different mechanisms. In the former, labels are qualitatively different from other features, and act in a top-down way to guide categorization by directing infants' attention to category-relevant exemplar features. In contrast, the compound representations view assumes that labels have the same status as other features with respect to how they are perceived. Specifically, they are not abstract guides to categories: like visual features, labels are low-level perceptual features. However, they are represented separately from visual features, and thus have equivalent representational status. In this sense, they are diagnostic – rather than deterministic – of categorization. This common status with other features allows labels to be embedded in object representations, rather than associated via an abstract link. However, despite substantial empirical work (e.g., Gelman & Coley, 1990; Gliga, Volein, & Csibra, 2010; Sloutsky & Fisher, 2004, 2012; Westermann & Mareschal, 2014) and a handful of computational investigations (Gliozzi et al., 2009; Mirolli & Parisi, 2005; Westermann & Mareschal, 2014), there is no current consensus as to the status of labels in object representations, and the debate goes on.

Nonetheless, the existence of a broader relationship between language and representation is not in dispute: multiple studies have demonstrated that language encodes perceptual distinctions (Boroditsky, 2001) and can influence representations on-line (Lupyan, 2016). While the effect of language on representation has been established in adults, however, when and how in development this relationship emerges is less clear. Studies demonstrate that labels can guide infants' online category formation in infants (Althaus & Westermann, 2016; Plunkett, Hu, & Cohen, 2008), and that learned, but unlabeled, category representations affect their in-the-moment behavior in the lab (Bornstein & Mash, 2010), but until recently the link between learned labels and representations had not been directly tested. Twomey & Westermann (2016; henceforth T&W) sought to trace the roots of this relationship to the earliest stages of language development, in 10-month-old infants. Infants were trained by their parents over the course of a week with two objects via 3-minute play sessions. Critically, infants were taught a label for one of the objects, but not for the other. After the training phase, infants participated in a looking time task in which they were shown images of each object in silence. On the hypothesis that (previously learned) labels would affect infants' object representations, the authors predicted that infants should exhibit different looking times to the labeled and unlabeled objects. Their predictions were upheld: infants maintained interest for longer in the previously labeled than the unlabeled object. Infants of this age have been repeatedly shown to engage preferentially with novel stimuli when familiarized for sufficient time (for a review, see Houston-Price & Makai, 2004). Thus, infants' longer looking to the labeled object across familiarization was interpreted as a novelty response to the previously labeled object. The authors concluded that labels shape object representations from the very beginnings of language acquisition.

These data shed light on the status of labels debate. Specifically, they support both the labels as features and the compound representations theories: if a label is an integral part of an object's representation, there will be a mismatch between that representation and the object in the real world when the label is missing. In contrast, the labels as invitations account predicts that removing the label should have no effect on infants' responses to objects when those objects are presented in silence, because labels are not integrated into object representations. Thus, the behavioral data do not support the labels as invitations view. However, these empirical data cannot differentiate between the labels as features and compound representations views. Computational models, on the other hand, allow researchers to explicitly test the mechanisms specified by these theories against empirical data. Thus, in the current study we explored which of the labels as features and compound representations explains T&W's results by implementing both accounts in neural network models.

Model Architecture

We used a simple three-layer auto-encoder model to implement both the labels as features and the compound representations accounts. Auto-encoders are feed-forward connectionist neural networks consisting of an input layer, a smaller hidden layer and an output layer. These models have successfully captured data from infant categorization tasks (Cottrell & Fleming, 1990; Mareschal & French, 2000; Twomey & Westermann, 2015; Westermann & Mareschal, 2012, 2014). These models reproduce input patterns on their output layer by comparing input and output activation after presentation of training stimuli, then using this error metric to adjust the weights between units using back-propagation of error (Rumelhart, Hinton, & Williams, 1986). The sum of the square of these error values (SSE) is typically used as a proxy for looking time (Mareschal & French, 2000; Westermann & Mareschal, 2012, 2014), and we use this index in the current simulations. The network consisted of 19 input units, 15 hidden units, and 19 output units. Hidden units used a sigmoidal activation function while output units used a linear activation function, and weights were initialized randomly between -0.25 and 0.25. We used a learning rate of 0.1, a momentum of 0.1, and a Fahlman offset of 0.1.

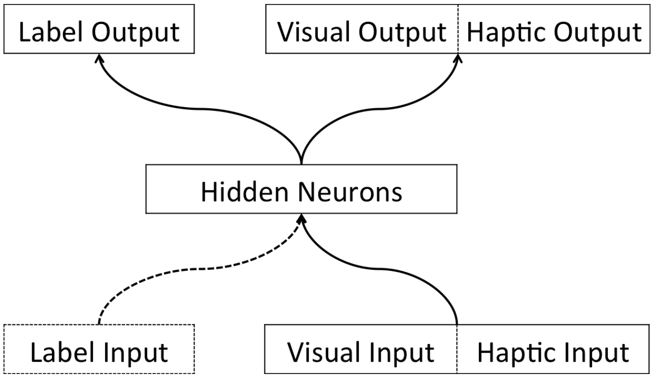


Figure 1. Network architecture

Labels as features Model (LaF)

Figure 1, including the dashed Label input, depicts the LaF model. To represent the label as a feature equivalent to all other features, we included it both at the input and the output level. Thus, the label had the same status as all other features in the model's representation.

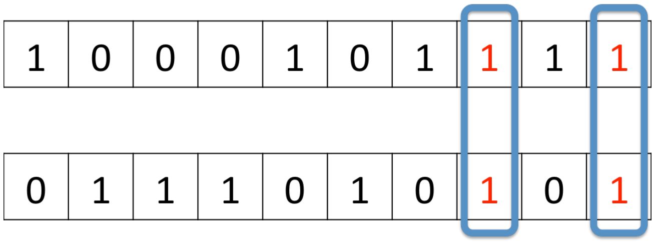
Compound Representation Model (CR)

Figure 1, excluding the dashed label input, depicts the CR model. We based this model on the Westermann and Mareschal (2012) auto-encoder-type dual-memory model in which labels are encoded as separate outputs. Thus, when an object is presented as an input the label is retrieved as part of that representation, but the label does not act as an object feature at a perceptual level. Note that in this model, we used only 18 of the input units, making it a partial auto-encoder.

Stimuli

Simuli reflected the visual, haptic and label characteristics of T&W's 3D object stimuli.

Visual input. In T&W's empirical study stimuli were two small toys, one painted blue and one painted red. One toy was a castanet, and the other was two wooden balls joined with string. Thus, the stimuli were visually dissimilar, but both consisted of two wooden components connected with string/elastic. To reflect the partial overlap in visual appearance of these objects, we encoded the visual component of our stimuli as pseudorandom patterns of activation over 10 units. We kept the total number of active units constant for each object, but allowed distribution to vary; however, two out of the ten units overlapped to represent commonalities between stimuli (see Figure 2).



1	0	0	0	1	0	1	1	1	1
0	1	1	1	0	1	0	1	0	1

Figure 2. Visual stimuli presented to the network. Blue boxes represent overlapping features.

Haptic input. As well as visual experience, infants in T&W received haptic input when handling or mouthing the stimuli. We reasoned that the degree of overlap in this input would vary between infants. Because both objects were wooden and experienced simultaneously, infants would have experienced some minimum overlap in haptic experience of the objects. On the other hand, because the objects had different affordances, this overlap would never have been exact. Thus, we encoded haptic input over 8 units, with overlap varying randomly between 2 and 6 units between simulations. Haptic stimuli were presented to the model simultaneously with the visual stimuli and encoded in an identical fashion.

Label input. Label input consisted of a single unit, activated for the labeled object only.

Procedure

In line with T&W's experiment, our procedure consisted of two phases. First, to simulate the 3D object play sessions, we trained the model with both objects, one with a label and one without a label. Then, we tested the model in a familiarization task in which the label was absent, as in T&W. Specifically, we ran each architecture in a test condition in which the label unit was inactive for both stimuli.

Play sessions

To reflect the likely differences in playing time across children, and playing time with individual objects for each child, the total number of iterations for which the model received each stimulus during background training was selected randomly from a normal distribution of mean 500 and standard deviation 200, for each stimulus. Duration of presentation of both stimuli was randomised based on the same distribution. Presentation began with random selection of one of the stimuli. Then, each stimulus was presented 30 times and for a total number of presentations as determined before.

Familiarization Training.

Before familiarization training, we added noise to hidden-to-output weights (from a uniform distribution ranging from 0.001 to 1) to simulate the likely memory decay from infants' final play session, which had taken place the previous day. Then, we removed the label from the inputs and outputs for the LaF condition, and from the outputs only for the CR condition.

Familiarization then proceeded as follows: in line with T&W, stimuli were interleaved (with learning) for 100 iterations, or until the sum squared error (SSE) between the input and output fell below a threshold of 0.01. The threshold reflects infants' looks away from the screen as the trial proceeds (Westermann & Mareschal, 2012, 2014). Each stimulus was presented for eight trials; the familiarization phase therefore consisted of 16 trials in total. The initial stimulus was counterbalanced across simulations.

Infants' looking time on a given trial was indexed as number of presentations, either until error fell below threshold or until the maximum number of presentations was reached.

Results

Results from the CR and LaF models are depicted in Figures 3 and 4.

We submitted looking time (SSE) to a 2 (model; CR, LaF) x 2 (condition; label vs. no-label) x (trial; 1 - 8) mixed ANOVA. Overall, the CR and LaF models' looking time differed ($F(1, 67168) = 1166.73$, $p <. 0001$, $\eta_p^2 = .017$), and decreased rapidly across trials ($F(7, 67168) = 28697.85$, $p < .0001$, $\eta_p^2 = .751$). There was also a small but robust difference in looking times to the labeled versus the non-labeled object, $F(1, 67168) = 55.07$, $p < .0001$, $\eta_p^2 = .001$). All two- and three-way interactions also contributed to the model (all Fs > 2.82, all ps < .004). To understand these interactions we conducted individual ANOVAs for each simulation. The CR model's looking time decreased rapidly across trial ($F(7, 33584 = 15307.88$, $p < .0001$, $\eta_p^2 = .761$), and there was a much smaller effect of condition ($F(1, 33584) = 4.18$, $p = .041$, $\eta_p^2 < .001$). However, unlike in T&W, there was no trial-by-condition interaction. Thus, although the CR model did show an effect of previously-learned labels on in-task looking times, it did not capture the nuanced pattern of results in the empirical study.

The LaF model's looking times also decreased across trial ($F(7, 33584) = 13940.44$, $p < .0001$, $\eta_p^2 = .74$), and this model showed a somewhat stronger effect of condition ($F(1, 33584) = 64.55$, $p < .0001$, $\eta_p^2 = .002$). Critically, the effect of trial interacted with the effect of condition $(7,33584) = 8.67$, $p < .001$, $\eta_p^2 = .002$). As depicted in Figure 4, post-hoc pairwise comparisons (Bonferroni-corrected) demonstrated that looking times initially decreased more rapidly in the no-label condition (all ps < .0001) but that this difference disappeared by the end of familiarization. Thus, while both models replicate T&W's overall finding that labels affect object representations, only the LaF model reflects the exact pattern of results reported by T&W: when all else is held equal, teaching the LaF model a label for one object but not another leads to a more rapid decrease in looking time to the unlabeled object in a subsequent, silent familiarization phase.

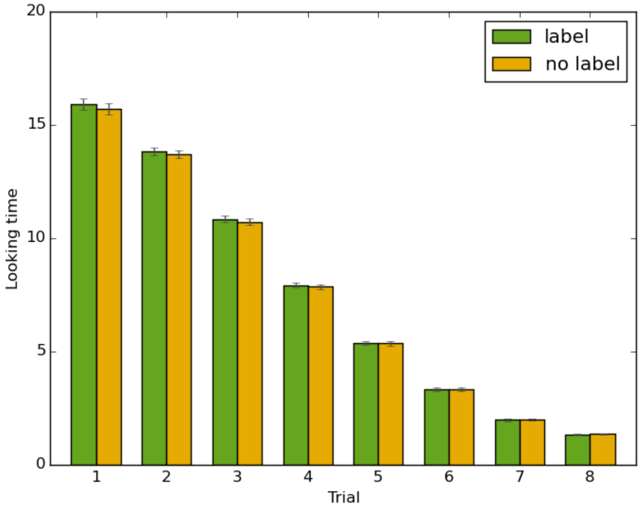


Figure 4. Results from the compound representations model.

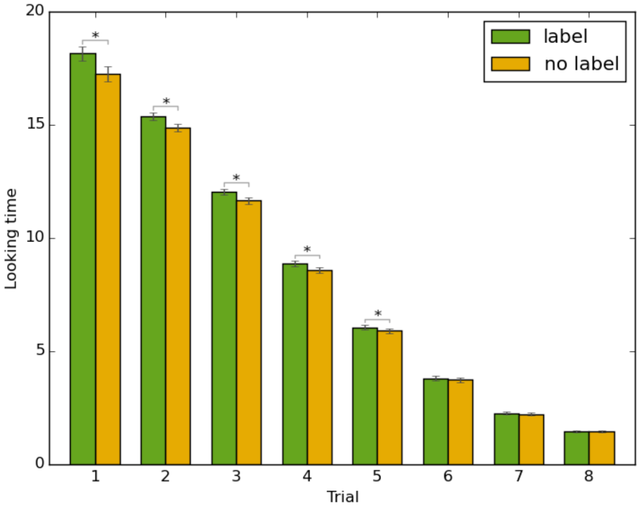


Figure 3. Results from the labels as features model. *p < .001

Discussion

In the current study we tested two possibilities for the relationship between labels and object representations using a neurocomputational model to capture recent empirical data (Twomey & Westermann, 2016). The target data showed that learned labels affect 10-month-old infants' looking times in a silent familiarization phase, suggesting that knowing a label for an object directly affects its representation, even when that object is presented in silence. As noted by T&W, both the compound representations and labels as features accounts predict some effect of labels on object representations, however the empirical data could not shed light on which of these two accounts best explained the pattern of results they observed. To untangle these two possibilities, we implemented both accounts in simple auto-encoder models (cf., Mareschal & French, 2000; Twomey & Westermann, 2015). In the compound representations model we instantiated labels on the output layer. This model learned to associate labels with inputs over time such that the presence of visual/haptic input for an object would consistently activate the label, but nonetheless, label representations were separate from visual and haptic object representations (Westermann & Mareschal, 2014). In the labes as features model, labels were represented on the input as well as on the output layer with the same status as the visual and haptic components of object representations (Gliozzi et al., 2009; Sloutsky & Fisher, 2004). Only the labels as features model captured the more rapid decrease in looking to the no-label stimulus exhibited by the infants in T&W's empirical study.

This work offers converging evidence that labels may have a low-level, featural status in infants' representations. In line with recent computational work (Gliozzi et al., 2009; Westermann & Mareschal, 2014) we chose to explore such low-level accounts to establish whether a simple associative model could account for the nuances of T&W's data. We did so for two reasons. First, the labels as invitations account assumes that in their communicative context, labels prompt infants to attend to the similarities between category exemplars (Ferguson & Waxman, 2016; Fulkerson & Waxman, 2007; Waxman & Markow, 1995). It is not clear how this view relates to T&W's data: although it predicts that representations of category exemplars will be more similar to one another in the presence of labels, it does not state whether this effect will be evident for a single object in the absence of a label or other communicative information, as in the familiarization task presented to infants in T&W's study (e.g., Ferguson & Waxman, 2016; Futó, Téglás, Csibra, & Gergely, 2010). In contrast, our labels as features model offers a parsimonious account of T&W's results, in which looking time differences emerge from a low-level novelty effect. Specifically, as argued by T&W, over background training the label is learned as part of the object representation. Thus, when the object appears without the label there is a mismatch between representation and external input. This mismatch leads to an increase in network error, our proxy for looking time, capturing the empirical data without a need for high-level communicative cues.

The current simulations also relate closely to the dual memory model presented by Westermann & Mareschal (2014). Our labels as features architecture suggests that if infants were to be taught a label for a *category* of objects – rather than a single object as described here – the absence of a label during familiarization should nonetheless provoke a similar novelty response. Interestingly, however, Westermann and Mareschal (2014) make the opposite prediction: familiarization time to new exemplars of a previously-learned category should be faster when those exemplars have previously been labeled. A key difference between the current model and that of Westermann & Mareschal is the latter's separation of memory into long- and short-term components. This allows the presence of a label during background training to actively restructure category representations, pulling exemplars closer to the prototypical category member. From this perspective, new category members should be perceived as more similar to learned representations when that category has a label, irrespective of the presence of the label in-the-moment. In contrast, labels in the present model are simply shared features, the absence of which gives rise to a mismatch. It is of course possible that both predictions are correct, but come into play at different developmental stages. For example, early in language development infants could initially form simple, featural associations between labels and objects, as in the current model. Over time, however, label representations could become more deeply entrenched, leading to the "magnet"-type effect on representations predicted by Westermann & Mareschal (cf. Deng & Sloustky, 2015; Kuhl, 1991). An empirical test of this possibility is important for a detailed understanding of labels' status in object representations.

It is important to note that other computational work has explored the effect of labeling and representation in this age group. Gliozzi et al. (2009) used a self-organizing map (SOM; Kohonen, 1998) architecture to capture empirical data from a categorization task with 10-month-old infants. In this network, labes are represented as units in SOMs in the same way as visual features. This model could capture T&W's results for similar reasons to the success of our labels as features model, although this remains an empirical question. However, the two networks make very different assumptions about learning mechanisms, highlighting an important issue for both infancy research and computational work. Gliozzi and colleagues' model learns in an unsupervised way, strengthening associations between units in its SOMs using "fire together, wire together" Hebbian learning. In contrast, our model learns by comparing what it "sees" to what it "knows" and updating its representations in proportion to any discrepancy. Thus, the current results are compatible with an error-based learning account to development, in which infants learn by tracking mismatches between representation and environment (Heyes, 2015). Whether unsupervised learning, error based learning, or some combination of both drives early development is a profound theoretical issue outside the scope of the current paper; for now we highlight the importance of bearing in mind the link between the technical assumptions of a computational model and the implications for (developmental) theory. Taken together with T&W, however, the current work demonstrates how language can shape representation and even change behavior from the bottom up.

Acknowledgments

This work was supported by a Doctoral Scholarship Grant from the Leverhulme Trust and by the International Centre for Language and Communicative Development (LuCiD). The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

References

Althaus, N., & Westermann, G. (2016). Labels constructively shape categories in 10-month-old infants. *In Press, Journal of Experimental Child Psychology.* http://dx.doi.org/10.1016/j.jecp.2015.11.013

Bornstein, M. H., & Mash, C. (2010). Experience-based and on-line categorization of objects in early infancy. *Child Development, 81*(3), 884–897.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology, 43*(1), 1–22. http://doi.org/10.1006/cogp.2001.0748

Cottrell, G. W., & Fleming, M. (1990). Face recognition using unsupervised feature extraction. In *Proceedings of the International Neural Network Conference* (pp. 322–325).

Deng, W. (Sophia), & Sloutsky, V. M. (2015). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, *51*(3).

Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition, 146*, 185–189. http://doi.org/10.1016/j.cognition.2015.09.020

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 105*(1), 218–228.

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative Function Demonstration induces kind-based artifact representation in preverbal infants. *Cognition, 117*(1), 1–8. http://doi.org/10.1016/j.cognition.2010.06.003

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*(5), 796.

Gliga, T., Volein, A., & Csibra, G. (2010). Verbal labels modulate perceptual object processing in 1-year-old children. *Journal of Cognitive Neuroscience, 22*(12), 2781–2789. http://doi.org/10.1162/jocn.2010.21427

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.

Heyes, C. (2015). When does social learning become cultural learning? *Developmental Science.*

Kohonen, T. (1998). The Self-Organizing Map, a possible model of brain maps. *Brain and Values*, 207–236 568.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. http://doi.org/10.3758/BF03212211

Lupyan, G. (2015). The paradox of the universal triangle: concepts, language, and prototypes. *The Quarterly Journal of Experimental Psychology*, (just-accepted), 1-69.

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy, 1*(1), 59–76. http://doi.org/10.1207/S15327078IN0101_06

Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. *Frontiers in Psychology, 4*. http://doi.org/10.3389/fpsyg.2013.00491

Mirolli, M., & Parisi, D. (2005). Language as an aid to categorization: A neural network model of early language acquisition. *Progress in Neural Processing, 16*, 97.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*(2), 665–681. http://doi.org/10.1016/j.cognition.2007.04.003

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology-General, 133*(2), 166–188.

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology, 111*(1), 65–86. http://doi.org/10.1016/j.jecp.2011.07.007

Twomey, K. E., & Westermann, G. (2015). A neurocomputational model of curiosity-driven infant categorisation. *Proceedings of the 5th Joint IEEE International Conference on Development, Leaning and Epigenetic Robotics.* Providence, RI.

Twomey, K. E. & Westermann, G. (2016, August). A learned label modulates object representations in 10-month-old infants. *Poster to be presented at the 38th Annual Meeting of the Cognitive Science Society, Philadelphia, PA.*

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology, 29*(3), 257–302.

Westermann, G., & Mareschal, D. (2012). Mechanisms of developmental change in infant categorization. *Cognitive Development, 27*(4), 367–382.

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120391.