

Asymmetric Categorization in the Sequential Auditory Domain

Charles Delb   (delbe@leadserv.u-bourgogne.fr),
Emmanuel Bigand (emmanuel.bigand@leadserv.u-bourgogne.fr)
Robert M. French (robert.french@u-bourgogne.fr)
LEAD-CNRS, UMR 5022, Dijon, France

Abstract

An unusual category learning asymmetry in infants was observed by Quinn *et al.* (1993). Infants who were initially exposed to a series of pictures of cats and then were shown a dog and a novel cat, showed significantly more interest in the dog than in the cat. However, when the order of presentation was reversed — i.e., dogs were seen first, then a cat and a novel dog — the cat attracted no more attention than the novel dog. A series of experiments and simulations seemed to show that this asymmetry was due to the perceptual inclusion of the cat category within the dog category because of the greater perceptual variability of dogs compared to cats (Mareschal & French, 1997; Mareschal *et al.*, 2000; French *et al.*, 2001, 2004). In the present paper, we explore whether this asymmetric categorization phenomenon generalizes to the auditory domain. We developed a series of sequential auditory stimuli analogous to the visual stimuli in Quinn *et al.* Two experiments on adult listeners using these stimuli seem to demonstrate the presence of an identical asymmetric categorization effect in the sequential auditory domain. Furthermore, we simulated these results with a connectionist model of sequential learning. Together with the behavioral data, we can conclude from this simulation that, as in the infant visual categorization experiments, purely bottom-up processes were largely responsible for our results.

Introduction

A number of years ago, Quinn, Eimas & Rosencrantz (1993) and Eimas, Quinn, & Cowan demonstrated an unexpected asymmetry in category acquisition in young infants. When 3- to 4-month-old infants were shown different photographs of either cats or dogs they were able to form perceptual categories of either groups of pictures. Infants who were first shown a number of different photographs of cats and are then a picture of a dog along with a picture of a novel cat will be more attentive to the dog than to the novel cat. This was interpreted as showing that the infants had formed a Cat category that excluded dogs. However, infants who were first shown different photographs of dogs and then a picture of a cat along with a novel dog were not preferentially attentive to either picture. This surprising finding was interpreted as showing that infants had formed a Dog category that *included* cats. In other words, infants show an exclusivity asymmetry in their development of some perceptual categories. Thus, the Dog category does not exclude cats whereas the Cat category excludes dogs.

Between 1997 and 2004, a number of papers were published that attempted to explain and expand on these findings. (Mareschal & French, 1997; Mareschal *et al.*, 2000; French *et al.*, 2001, 2004; etc.) These experiments seemed to demonstrate that the key relationship that produced these results was that the smaller variability of the Cat category

compared to the Dog category, meant that, perceptually, the latter category largely subsumed the former. This meant that bottom-up, purely perceptual mechanisms were enough to explain the categorization asymmetry observed by Quinn and colleagues.

The authors manipulated the variability of the Dog and Cat categories by selecting, in one experiment, a set different races of dogs with little variability and a set of cats with a much greater variability. In this way, even though the high-level categories (i.e., Dog and Cat) remained unchanged, their low-level perceptual variability had been reversed. The connectionist model that had been developed and which relied only on the statistical distributions of the features of the two categories, predicted a reversal of the original asymmetric categorization. This is, indeed, what the authors found. As a result, the authors were able to conclude that the infants were relying exclusively on statistical (i.e., bottom-up) properties of the stimuli in their category discrimination.

The acoustic domain

This ability of young infants to discriminate between two categories of complex visual stimuli in a purely bottom-up manner — i.e., in the absence of any conceptual knowledge of the stimuli — led us to wonder if there might be an analogous phenomenon in the acoustic domain.

Although the perception of music relies on many different perceptual dimensions, such as timbre, loudness, rhythm, and pitch, one of the most salient features of music is that of pitch. Pitch perception is, indeed, fundamental to melody in music. When memorizing a tune, people do not represent the melody as a series of independent pitches, but process each pitch relative to the others. This leads to the fundamental notion of *musical interval*. The “sequential distance” between two notes can be measured by the chromatic interval (Krumhansl, 1990) seems to be the basic unit in the memorization of melodies. Plantinga & Trainor (2005) showed that infants store melodies in terms of relative pitches and not absolute pitches.

Asymmetric effects in music perception

In perception of musical style, Bigand & Barrouillet (1996) claimed that (non-musician) participants who were familiarized with selections of Baroque music (narrow category, Bach) and then tested on a novel Baroque piece versus a selection of early 20th century tonal musical (e.g., Debussy), showed an significantly increased interest in the early 20th century selection. On the other hand, when participants were familiarized with selections of early 20th century music and then tested on a novel selection of early 20th century music versus a Baroque piece, there was no increased interest in the Baroque piece.

In light of these results and the asymmetric visual categorization asymmetry observed in infants, we decided to attempt to explore this phenomenon in a controlled acoustic environment using artificially produced musical stimuli presented to non-musician participants.

Experiment 1

The aim of this experiment was to assess an effect of interval distribution in the formation of melodic categories. To this end, we compared the exclusivity of auditory sequential categories formed during exposure to exemplars of melodies, statistically controlled in terms of their interval distributions. In this experiment we attempted to reproduce the category inclusion and distribution relationships that produced the asymmetric categorization results in Quinn *et al.*, 1993 (see Figure 1).

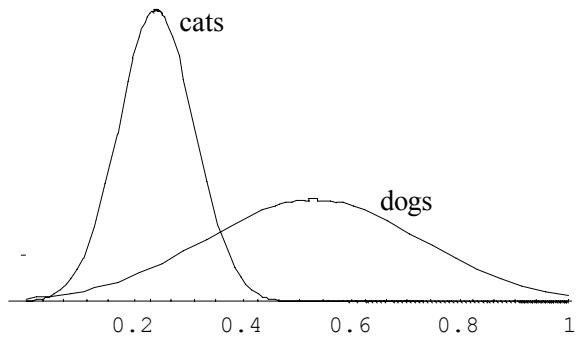


Figure 1: General inclusion and variability of the feature distributions for Dogs and Cats in Quinn *et al.*, 1993.

Participants

51 undergraduate psychology students from the University of Burgundy, all non-musicians, took part in the experiment.

Material

For each participant, two sets of melodies were created, one following a *narrow* distribution, the other a *broad* probability distribution of the occurrence of 11 different musical intervals (see Figure 2). Each melody was built with 5 consecutive intervals (e.g. 6 pitches), randomly chosen according to one of the distributions. Each set consisted of 72 sequences. The start note of the melodies was counterbalanced between the 12 possible pitches (yielding 6 different melodies per start-note). Each participant was randomly assigned one of the two distributions.

Even though specific intervals associated with each probability used in both the training and test phases are varied over participants, the probability of occurrence of these intervals followed one of the two distributions shown in Figure 2 (i.e., either narrow or broad). This was done to counterbalance the effect of prominence of some particular intervals in the process of melodic categorization.

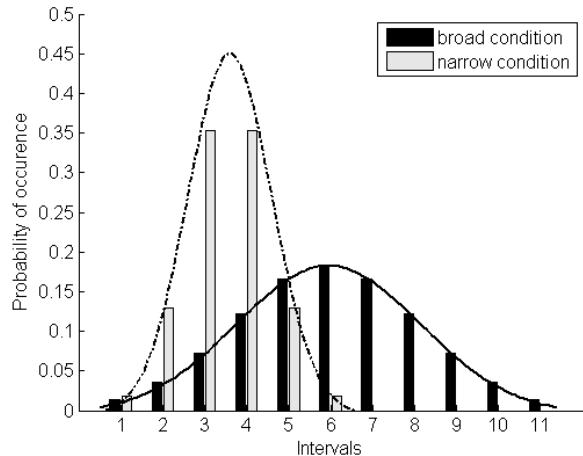


Figure 2: Probabilities of occurrence of 11 musical intervals, for the *narrow* (gray) and *broad* (black) conditions.

Melodies were synthesized with a MIDI synthesizer software (using the piano bank). Each tone was randomly played for either 250 or 500 ms, giving more “musicality” to the pitch sequences (in order to alleviate listener fatigue). A further feature that is of particular importance in melodic perception is the contour, or the pattern of ups (+) and downs (−) of pitches from one note to the other (see Dowling & Harwood, 1986). The contour of each melody was random, and there was no repetition of contour/interval patterns within the same set of melodies.

Procedure

Each participant was exposed to 84 pitch sequences. The first 60 sequences — which constituted the training phase — were exemplars drawn from one of the two distributions. The 24 remaining items — which constituted the test phase — were composed of 12 new items from the training distribution, and of 12 items from the other distribution. These last 24 items were randomly ordered.

A presentation program was written within MATLAB programming language. Melodies were played through headphones. At the end of the pitch sequence, participants were then asked if they thought they had previously heard the melody during the training session. The inter-trail interval between the subject's answer and the sounding of the next melody was 2 seconds.

Results

For each participant, the correct-response rate (percentage of correct recognition of the participant's training distribution) during the test phase was computed. Single-group t-tests were used to compare the performance of each group to chance levels (50%).

The *broad* group performance (50%) was not significantly different from chance, $t(24)=0$, $p=1$, whereas the *narrow* group's performance (55.6%) was significantly above chance, $t(25)=2.42$, $p=0.023$. This suggests that the *narrow* group learned aspects of the interval statistics of the *narrow* items, allowing them to recognize new melodies from this distribution appropriately.

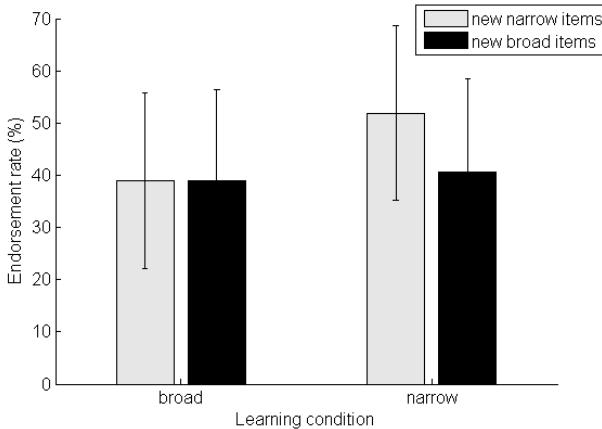


Figure 3: Mean endorsement rate during test phase for the *broad* group (left) and the *narrow* group (right). Error bars represent standard errors.

We then analyzed the endorsement rate (i.e., percentage of accepted test items) for both groups, for new broad and new narrow items (Figure 3). A 2x2 mixed-ANOVA (broad/narrow training condition x broad/narrow test items) revealed a significant main effect of test items ($F(1,49)=4.51$, $p<.05$). In other words, as in the Quinn *et al.*'s (1993) Dog-Cat studies, participants familiarized with the narrow (i.e., low-variability) stimuli excluded more contrasting items (in this case, new broad items), whereas participants familiarized with the broad (i.e., high-variability) category rejected narrow items at no better than chance. This mirrors the inclusion/exclusion relationships observed in Quinn *et al.* (1993)'s stimuli.

Experiment 2

In light of the results of the first experiment, we decided to tighten the constraints on the statistical properties of the sequences of notes comprising the melodies for the broad and narrow categories. This was done in order to determine if it was possible to enhance the effects found in the first experiment. In addition, during training, we decided to use a more incidental memory task.

Participants

24 students from the University of Burgundy, all non-musicians and having not participated to the experiment 1, took part in the experiment.

Material

Melodies were generated by a Markov process, yielding highly constrained Markov chains, where the probability of a specific event i , depends on the occurrence of a prior event. We used a 1st-order Markov model, which can be represented using a 2-dimensional transition matrix. The probability of a given event depends only on the event immediately preceding it.

Table 1: Transitional probabilities between pitches, for the *broad* distribution (in black in figure 4).

		To pitch						
		end	1	2	3	4	5	6
Fr o m pit ch	start	.1	.25	.15	.15	.2	.2	.1
	1	.2	.15	.15	.1	.25	.1	.2
	2	.1	.1	.15	.15	.2	.2	.25
	3	.1	.2	.15	.15	.2	.1	.25
	4	.1	.2	.2	.15	.15	.25	.1
	5	.1	.25	.15	.1	.15	.2	.2
	6	.1	.1	.25	.2	.15	.2	.15

A sequence is built by selecting a start-note according to the probabilities of the “start” row, selecting the second note according to its probability of occurrence after the first note, cycling through the table until the “end” column is reached (see table 1). The transitional probabilities between a note and the 6 possible following ones (plus the “end” code) followed either a *broad* or a *narrow* distribution (Figure 4).

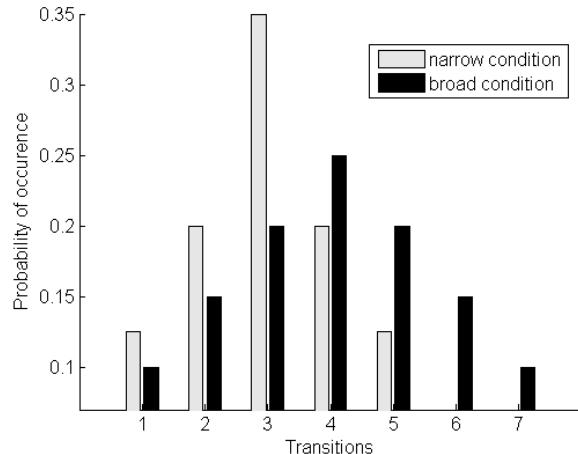


Figure 4: Probabilities of occurrence of the 7 transitions, in *narrow* (gray) and *broad* (black) conditions.

The transition matrices for the two categories were associated with four different pitch-sets (notes 1-6 in Table 1) : C4/D4/D#4/F4/G#4/A#4, D#4/F4/F#4/G#4/B4/C#4, A3/B3/C4/D4/F4/G4 or F#4/G#4/A4/B4/D4/E4, to avoid specific effects associated with one or the other pitch-set. Note that the musical intervals between each adjacent notes is identical across the 4 sets.

70 melodies were generated per condition, 60 of them being used as exemplars in a training phase. In the test phase, 20 pairs of melodies were used, composed of the 10 remaining sequences of the training condition, and of 10 sequences of the contrasting distribution. Each pair consisted of one melody from each set. The order within pairs was

counterbalanced. The number of tones for each melody in both group varied from 4 to 7 ($\mu_{\text{narrow}} = 5$, $\mu_{\text{broad}} = 4.7$). Tones' duration and contour were controlled as in Experiment 1.

Procedure

Participants were seated in front of a computer. Melodies were played through headphones. In a first phase, participants were asked to listen to the melodies, and to report the total number of pitches of each sequence. Feedback was given after each answer, indicating whether or not their reply was correct, and if not, reporting the right number of pitches. The inter-trail interval between the subject's answer and the next melody was 2 seconds. In a second phase, participants heard pairs of melodies. They were then asked to select, for each pair, the melody most similar to the ones they had heard in the first phase. Both groups had 12 participants.

Results

First, the data obtained from the training phase was analysed. The mean number of correct responses was calculated for each participant. This score was high for both groups (*broad* group : 88% (SD: 7.5), *narrow* group: 83.3% (SD: 11.1)). The pitch-counting task was relatively easy for the participants. The participants' strategy seemed to generally be to rehearse the melody, pitch by pitch, after hearing it. A 2-tailed t-test conducted with groups as the independent variable, and percentage of correct responses during the training phase as dependent variable yielded no significant effect, $t(22)=1.21$, $p>.2$, showing that melodies from both sets were not processed significantly differently according to the task (counting the tones).

The second set of analyses concerned the test phase. Here again, our results closely resembled those of the infant categorization experiments of Quinn *et al.*, (1993). The *broad* group performance (49.2%) was not significantly different from chance, $t(11)=-0.3$, $p>0.7$, whereas the *narrow* group's performance (57.5%) was significantly above chance, $t(11)=3.95$, $p=0.023$.

A 2x2 mixed-ANOVA (broad/narrow training condition x broad/narrow test items) revealed an effect of test items ($F(1,22)=6.25$, $p=.02$), but no significant main effect of training nor a significant interaction (Figure 5).

These results suggest that participants in the *narrow* condition learned the statistical distribution of the sequences, in terms of musical intervals, allowing them to recognize new melodies drawn from this distribution, whereas participants within the *broad* condition performed at chance level.

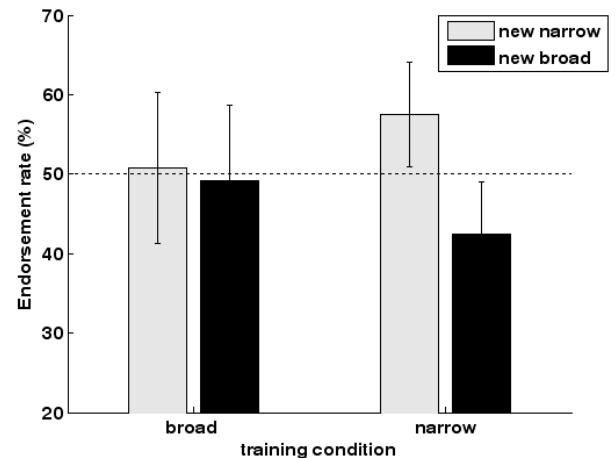


Figure 5 : Mean endorsement rate during test phase for the *broad* group (left) and the *narrow* group (right). Error bars represent standard errors.

Discussion

Experiment 2 confirmed and enhanced the effects observed in the first experiment.

The material used in these experiments was "music-like," rather than "musical," because although the stimuli were constructed with musical pitches, they could not have been encountered in a natural musical environment. So, it seems unlikely that top-down knowledge could have been responsible for the asymmetry we observed. However, to confirm that these results do not come from any influence of top-down musical knowledge, we tried to produce the results with a bottom-up connectionist model of sequence processing.

Simulation

The aim of the following simulation is to show that a purely bottom-up cognitive model, without any prior knowledge, can explain the results of our experiments. The main difference with the Quinn et al's visual stimuli resides in the sequential aspect of our auditory material. In fact, any attempt to simulate the asymmetric effect shown in experiments 1 and 2 must accommodate the sequential properties of the material. In order to model the process underlying the results described above, we used Simple Recurrent Network (SRN; Elman, 1990). These artificial neural networks are frequently used to encode sequential dependencies between elements of a sequence (see Cleeremans, 1993; Dienes 1993).

Procedure

The material used to train the networks is identical to that of Experiment 2, except that only one pitch-set was used, instead of 4. We used a localist coding scheme, i.e. each note of a sequence is coded by a unique bit in a 7-elements vector (1 bit per note + a "start/end" bit). Two groups of 20 SRNs were used, each being exposed to exemplars from either the narrow distribution, or the narrow one.

Each network was randomly initialized. Networks were composed of 7 input nodes, 3 hidden nodes and 7 output

nodes. Hidden nodes used a *sigmoid* transfert function, whereas output nodes used a *linear* activation function.

During the training phase, each item from a training set was presented twice consecutively. Stimuli were presented twice in order to simulate the strategy employed by participants: rehearsing the melody after hearing it (see Experiment 1). For each element of a sequence, the task of the networks was to predict the next element of that sequence. A mean square error (MSE) is computed for each element, giving a measure of prediction accuracy (i.e. the distance between the output computed by the network in response to an element and the actual desired output, that is, the next element in the sequence). Weights were updated at each time step, using a gradient descent with momentum training algorithm. The learning rate was set to .1, the momentum term to .4.

During the test phase, no weight change was allowed. Test items were presented in pairs, the order of items within each pair was random. The context units activation were reset to 0 between pairs, but not between items of a pair. This was meant to reflect the 2-alternative forced choice procedure of experiment 2. Within each test pair, the MSE was computed for each item. The sequence associated with the lowest MSE reflected the "choice" of the network. We then computed an endorsement rate for each test item type, across all test pairs.

Results

Figure 6 shows the mean endorsement rate for the two network groups, in response to new narrow and new broad items.

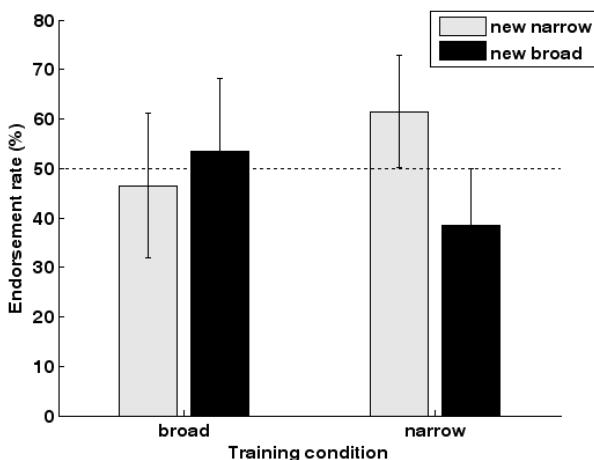


Figure 6: Mean endorsement rate during test phase for the the *broad* and *narrow* trained-networks. Error bars represent standard errors.

The simulation results closely resemble those of the experiment 2; networks trained with exemplars drawn from the broad category cannot distinguish new items from both categories, whereas networks trained with items from the narrow category produce more accurate predictions about the element of the narrow sequences, yielding a higher endorsement rate for new items from their own category than

from the other. Noting that these networks do not have any prior knowledge of the material we used, we can conclude that a simple connectionist model, which processes stimuli in a purely bottom-up fashion, is sufficient to simulate our behavioral results.

Conclusions

These preliminary results suggest that the categorization asymmetry in young infants observed by Quinn *et al.* (1993) is not limited to the visual domain. Rather, it is probable that this phenomenon also applies to auditory perception. Our results point to the importance of bottom-up (statistical) processing in the perception and categorization by non-musicians of sequential auditory stimuli.

Meulemans and Van der Linden (1997) have shown that, in an artificial grammar learning task, participants exposed to a small subset of a grammar were sensitive to the similarity of the test items with the training items, whereas with longer exposure, this similarity effect disappeared. The similarity was measured as the mean probability of occurrence in the training set of the bigrams (2 consecutive elements in a sequence) and trigrams (3 consecutive elements) composing a test item. This implies that the statistical distribution of the different bigrams and trigrams in the training set had an effect on the ability of the participant to discriminate sequences from their own category (grammar) from distractors. This suggests that the asymmetrical effects we described in our experiments could be eliminated by the acquisition of syntactic rules governing the elements of musical pitches.

Finally, these results are limited to a single auditory dimension (pitch intervals). It will also be necessary to investigate the influence of the myriad other dimensions of musical perception (e.g., duration, timbre, rhythm, etc.) and the interaction among these various dimensions.

Acknowledgments

This work was funded in part by a Framework 6 European Commission grant NEST 516542 to the third author.

References

- Bigand, E., & Barrouillet, P. (1996). Processi di classificazione degli stili nei bambini e negli adulti. *Quaderni della SIEM, semestrale di ricerca e didattica musicale*, 10(1), 81–93.
- Cleeremans, A. (1993) Mechanisms of implicit learning: Connectionist models of sequence processing. Cambridge, MA: MIT Press.
- Dienes, Z. (1993) Computational models of implicit learning. In D.C Berry & Z. Dienes (Eds.) *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum.
- Dowling, J & Harwood, D. (1986). *Music Cognition*. NY: Academic Press.
- Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants, *Journal of Experimental Child Psychology*, 58, 418-431.
- Elman, J.L. (1990) Finding Structure in Time, *Cognitive Science*, 14, 179-211.

French, R. M., Mareschal, D., Mermilliod, M., & Quinn, P. C. (2004). The Role of Bottom-up Processing in Perceptual Categorization by 3- to 4-month-old Infants: Simulations and Data. *Journal of Experimental Psychology-General*, 133(3), 382-397.

French, R. M., Mermilliod, M., Quinn, P., & Mareschal, D. (2001). Reversing Category Exclusivities in Infant Perceptual Categorization: Simulations and Data. In *Proceedings of the 23nd Annual Conference of the Cognitive Science Society*, NJ:LEA, 307-312.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Mareschal, D. and French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In *Proceedings of the 19th Annual Cognitive Science Society*, New Jersey: LEA, 484-489.

Mareschal, D., French, R. M. & Quinn, P. (2000). A Connectionist Account of Asymmetric Category Learning in Early Infancy. *Developmental Psychology*, 36, 635-645.

Meulemans, T. & Van der Linden, M. (1997) Associative Chunk Strength in Artificial Grammar Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23.

Plantinga, J. & Trainor, L. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, 98, 1-11.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants, *Perception*, 22, 463-475.