Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (Saguinus oedipus)

Marc D Hauser, Harvard University
Stanislas Dehaene
Ghislaine Dehaene-Lambertz
Andrea L Patalano
Brief article

Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (Saguinus oedipus)

Marc D. Hauser a,*, Stanislas Dehaene b, Ghislaine Dehaene-Lambertz c, Andrea L. Patalano b

a Department of Psychology & Program in Neurosciences, Harvard University, Cambridge, MA 02138, USA
b Inserm Unit 334, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, Orsay, France
 c LSCP, CNRS UMR 8554 et EHESS, Paris, France

Received 21 January 2002; accepted 9 May 2002

Abstract

Studies using operant training have demonstrated that laboratory animals can discriminate the number of objects or events based on either auditory or visual stimuli, as well as the integration of both auditory and visual modalities. To date, studies of spontaneous number discrimination in untrained animals have been restricted to the visual modality, leaving open the question of whether such capacities generalize to other modalities such as audition. To explore the capacity to spontaneously discriminate number based on auditory stimuli, and to assess the abstractness of the representation underlying this capacity, a habituation-discrimination procedure involving speech and pure tones was used with a colony of cotton-top tamarins. In the habituation phase, we presented subjects with either two- or three-syllable sequences that varied with respect to overall duration, inter-syllable duration, and pitch. In the test phase, we presented subjects with a counterbalanced order of either two- or three-tone sequences that also varied with respect to overall duration, inter-syllable duration, and pitch. The proportion of looking responses to test stimuli differing in number was significantly greater than to test stimuli consisting of the same number. Combined with earlier work, these results show that at least one non-human primate species can spontaneously discriminate number in both the visual and auditory domain, indicating that this capacity is not tied to a particular modality, and within a modality, can accommodate differences in format. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Number discrimination; Abstract concepts; Monkeys; Comparative data

* Corresponding author. Fax: +1-617-496-7077.
E-mail address: mdhauser@wjh.harvard.edu (M.D. Hauser).

0010-0277/02 - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0010-0277(02)00158-0
1. Introduction

The capacity for numerical representation is phylogenetically ancient, dating back at least as far as birds (Dehaene, 1997; Gallistel, 1990; Hauser, 2000). Currently, there are at least two significant controversies concerning the nature of non-linguistic numerical representations in animals and in infants (Gallistel & Gelman, 2000; Hauser & Carey, 1998). First, what are the mechanisms underlying numerical representations in organisms lacking language? Second, to what extent is the capacity for numerical representation spontaneously available, abstract, and capable of computing over input from multiple modalities? In this paper, we focus on the second problem, asking whether a non-human primate is capable of spontaneously discriminating auditory stimuli based on number as opposed to some other continuous variable; we explore the abstractness of this ability by assessing whether these animals recognize the similarity between stimuli that converge with respect to number while diverging with respect to format (i.e. speech vs. tones).

Studies of number discrimination in the wild and in captivity motivate the logic of our experiment. Research on foraging behavior indicates that animals spontaneously use various statistical computations to determine rates of return, including Bayesian statistics (Giraldeau & Caraco, 2000; Stephens & Krebs, 1986). Although none of these studies show that animals use a counting system to quantify food intake, all show that they can approximate the amount of food obtained, the time spent obtaining it, or some combination of amount and time (rate). Complementing these foraging observations are studies that ask whether animals attend to the number of individuals in a group when assessing the level of competition. In a wide range of species, coalitions of two to three individuals outcompete lone individuals for access to food or mates (Aureli & de Waal, 2000; Harcourt & de Waal, 1992). Further, playback experiments with lions and chimpanzees show that the level of aggressive response to auditory simulations of foreign intruders depends on the number of both intruders and defenders (McComb, Packer, & Pusey, 1994; Wilson, Hauser, & Wrangham, 2001). Together, these studies suggest that numerical quantification is spontaneously available, ecologically important, and possibly generalizable across visual as well as auditory inputs. They do not, however, provide a clear specification of the limits or abstractness of number discrimination.

Operant training studies in the laboratory have shown that animals represent number as an analog magnitude with scalar variability (Gallistel, 1990; Meck & Church, 1983). Specifically, animals can use number as a discriminative cue to obtain food rewards, and can discriminate large differences (2 vs. 9) more readily than small differences (2 vs. 3) and small numbers more readily than large even when the absolute difference between pairs is held constant (e.g. 2 vs. 3 as opposed to 8 vs. 9; Brannon & Terrace, 2000). These latter findings show that number discrimination adheres to Weber’s law. These studies also show that number is represented in an abstract sense in that animals discriminate objects or events across different modalities. For example, Fernandes and Church (1982) trained rats to press the right lever if they heard two sounds, but the left lever if they heard four sounds. By systematically varying the duration of the sounds, as well as the interval between sounds, results showed that the discrimination was based on the number of stimuli rather than other factors; when the experimenters substituted light flashes for sounds, the rats immediately transferred. Further, in a study by Church and
Meck (1984), results showed that rats were capable of integrating the number of visual and auditory stimuli, even when they were presented simultaneously.

In contrast to operant experiments, studies conducted in the wild and in the laboratory using spontaneous techniques have yielded different patterns of results and have yet to provide any evidence that number is represented in a format other than visual. Specifically, studies of wild rhesus monkeys and captive cotton-top tamarins (Hauser & Carey, 1998; Hauser, Carey, & Hauser, 2000; Hauser, MacNeilage, & Ware, 1996; Sulkowski & Hauser, 2000; Uller, Hauser, & Carey, 2001) indicate that the number of visually presented objects can be discriminated, but with a limit of approximately 4, and with no evidence of the analog magnitude signature (i.e. Weber’s law). For example, in a two box choice paradigm developed by Hauser et al. (2000), rhesus monkeys selectively picked up to 4 pieces of apple over 3, but failed to show a preference for 5 over 4 and even 8 over 4. These results have been interpreted as being consistent with an object file-like mechanism (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn & Storm, 1998), with constraints operating in terms of the number of objects that can be tracked and held in short-term working memory. Although it is currently unclear why the analog magnitude system is tapped in some situations, and an object file system in others, a further limitation on current studies of spontaneous number discrimination is the lack of research using non-visual input. The following study addresses this gap, describing habituation-discrimination experiments on cotton-top tamarin monkeys. We ask not only whether tamarins can discriminate auditory stimuli based on number as opposed to some continuous measure, but also whether this discrimination is based on an abstract representation that operates over auditory stimuli that differ in format (speech vs. tones).

2. Methods

2.1. Subjects

We conducted playback experiments on a colony of 13 adult cotton-top tamarins, five males and eight females. Subjects are housed in social groups, consisting of at least one breeding pair, and often, their offspring. All subjects have been tested in auditory experiments involving both conspecific and heterospecific auditory signals (Ghazanfar, Flombaum, Miller, & Hauser, 2001; Hauser, Newport, & Aslin, 2001; Miller, Dibble, & Hauser, 2001; Ramus, Hauser, Miller, Morris, & Mehler, 2000). Importantly, the current colony of animals has been run with a habituation-discrimination procedure, similar to the one implemented here, yielding robust and replicable results in the absence of training.

2.2. Stimuli

We habituated subjects to two or three synthetic consonant–vowel syllables whose elements were drawn without replacement from a set of three consonants (p, t, or k) and three vowels (a, i, or u) (e.g. pika, pikatu). Fundamental frequency (195, 225 or 255 Hz) and intensity (78, 76 and 73 dBC) were systematically varied. Syllable duration varied between 225 and 450 ms for two-syllable stimuli, and between 112 and 225 ms for
three-syllable stimuli. This ensured that the interval between the first and the last syllable ranged between 225 and 450 ms and was equalized on average in those two sets.

The test stimuli were always a series of two or three tones of 100 ms duration. The fixed onset asynchrony of the tones (225 ms) insured that the difference between two and three occurred at a fixed point in time (450 ms = onset of the third tone). All two or three tones within a given stimulus had the same pitch, timbre and intensity that were randomly selected on each trial amongst three possibilities (262, 330, or 394 Hz; brass, wood, or guitar; 78, 76, or 73 dBC).

Controls were introduced to insure that discrimination was based on numerosity rather than on any other confounded parameter (see Fig. 1).

2.2.1. Total duration

This parameter was operationally defined as the delay between the onset of the first and the last event (tone or syllable). Total duration could not be used for discrimination since it was equalized across the two categories (syllables, tones) of context stimuli (two or three syllables).

2.2.2. Tempo

As a consequence of the manipulation of total duration, the tempo (delay between the onsets of two consecutive events) was twice as short in the three-syllable stimuli as in the two-syllable stimuli. Discrimination could not be based on tempo, however, because for the test stimuli with two or three tones, this parameter was equalized at 225 ms.

2.2.3. Individual event duration

Because each syllable filled the entire interval between event onsets, individual syllable duration was identical to tempo and hence varied between the two-syllable and three-syllable context sets. However, individual tone duration was equalized across the test stimuli and thus could not be used for discrimination.

2.2.4. Total stimulation

The use of longer syllables in context two than in context three stimuli would appear to generate a confound in that total auditory stimulation – operationalized as the sum of the durations of all sound periods – is greater on average for the two-syllable stimuli than for the three-syllable stimuli. However, there was considerable variability and overlap between the two- and three-syllable stimuli on this dimension. Furthermore, because the individual tones were of a fixed duration, this temporal cue was reversed in the test stimuli, where total stimulation was greater for three-tone stimuli than for two-tone stimuli. Thus, discrimination based on total stimulation would predict that the test stimuli deviated more from the habituation stimuli on numerically identical trials than on numerically different trials. A contrario, a significant dishabituation response on numerically different trials compared to numerically identical trials cannot be attributed to a discrimination of total stimulation.

In some habituation studies in infants, any differential response during the test period, whether revealing a preference for the novel or for the old stimuli, is interpreted as evidence for discrimination (e.g. Starkey, Spelke, & Gelman, 1990). Such an ambiguity, however, cannot affect the present playback method where the test period starts only after
Fig. 1. (Top) Sample synthetic speech and tone stimuli used respectively as habituation and test items, illustrating the important variability in phonemic content, tempo, and duration. For simplicity, additional variations in pitch and intensity have not been depicted. (Bottom) Scheme used to control for non-numerical parameters. Each plot shows the percentage of stimuli presented at each given parameter value, separately for the two habituation numerosities (first two rows, marked H2 and H3) and the two test numerosities (last two rows, marked T2 and T3); white histograms correspond to two-item stimuli while hatched histograms correspond to three-item stimuli. Overall duration was equalized for H2 and H3, while tempo was equalized for T2 and T3. Thus, neither parameter could explain the monkeys’ differential reaction to H2 followed by T3 or to H3 followed by T2, relative to their respective controls H2-T2 and H3-T3. Finally, in terms of total stimulation duration, both H2 and H3 were more similar to T3 than to T2. Thus, discrimination based on total stimulation would predict systematically greater dishabituation to T2 than to T3, particularly in subjects habituated with H2, which was not observed.
the subject has stopped orienting to the habituation set. This creates an asymmetry
between the old and new stimuli during the test period, so that subjects may orient
more to the new stimuli, or equally to the new and old stimuli, but never more to the
old stimuli. Thus, greater orienting to the numerically different stimuli can only imply
numerical discrimination, not discrimination based on total duration.

2.3. Playback methods

In brief, each subject was presented during a habituation phase with multiple exemplars
of either two- or three-syllable speech stimuli until they failed to respond on three conse-
cutive trials and subsequently, in a test phase, were presented with eight alternating trials
of two- and three-tone stimuli. We predicted that subjects would show a higher proportion
of responses to tone stimuli differing in number from the habituation stimuli. Thus, for
example, subjects habituated to two-syllable stimuli would respond more to three-tone
than to two-tone stimuli.

The test cage (45 × 45 × 20 cm) was located inside an auditory chamber (Industrial
Auditory Company, Inc., Model 400-A). A thin black cotton sheet separated the cage from
an Alesis Monitor One speaker (frequency range: 45–18000 Hz ± 3 dB) mounted on a
shelf above, behind, and to the left of the subject.

A video camera attached to a monitor outside the chamber was used to view the session
and to capture the experiment onto a videotape. Stimuli were played using an Audiomedia
II sound card outputting to the speaker.

Each subject was removed from its home cage and transported to the testing room.
Inside the testing room, we transferred the subject into the test cage. We habituated half of
our colony to two-syllable exemplars and the other half to three-syllable exemplars. The
inter-stimulus interval ranged from a minimum of 10 s to a maximum of 60 s (habituation
trials: mean = 28.5, SD = 33.8; test trials: mean = 24.0, SD = 11.6). We played stimuli
only when the test subject was looking down and away from the speaker. A positive
response was defined as looking up and back toward the speaker either (a) during playback
and continuing after stimulus offset or (b) after stimulus offset, but within a 2 s window.
Responses occurring during playback, but ending before stimulus offset, were excluded
because these orienting responses terminated before the number of tones had been
presented; as in previous playback experiments, responses occurring after a 2 s window
are considered ambiguous with respect to the causal relationship between playback and
response. Other responses were coded as “no”, “ambiguous”, or “bad”. “No” responses
involved either no movement at all, or movement in a direction that was unambiguously
opposite to the speaker. “Ambiguous” responses involved cases where subjects turned
during the response period, but in a direction that was neither clearly toward nor away
from the speaker. For on-line coding, ambiguous responses were coded as positive
responses, whereas for off-line coding they were kept as ambiguous; trials recoded as
ambiguous during the final three habituation trials would require rerunning the experi-
ment, whereas ambiguous test trial responses were eliminated from the final data set.
Lastly, bad responses occurred when subjects oriented toward the speaker immediately
before playback, thereby making it difficult to code the response. These trials were treated
in the same way as the ambiguous responses for both on- and off-line coding.
We maintained subjects on the habituation series until they failed to respond (“no” responses) on three consecutive trials; prior research suggests that following three consecutive no-responses, the probability of a spontaneous orienting response toward the speaker is extremely low (Weiss, Garibaldi, & Hauser, 2001). Following habituation, each subject received eight test trials, alternating different exemplars of two and three tones, counterbalanced for order; half the colony started with the same number as in the habituation series (e.g. habituate to two syllables, test with two tones then three tones), and half with the different number. We use non-parametric tests throughout and a two-tailed distribution for evaluating P levels.

Although we scored responses on-line during the experiment, and thus were aware of the conditions, we recoded the last three habituation trials and the eight test trials blind to condition by using a digitized video acquisition system. In brief, each of these trials were acquired and assigned a file name; this file name was listed next to the experimental condition and this list constituted the master file. Before analyzing a trial, an experimenter marked each digitized record with three flags, one for the onset of the playback, one for the offset, and one for a 2 s response window following the offset of the playback. Once all files were marked, an experimenter stepped through each file, frame by frame, and coded responses blind to condition by simply turning off the audio portion of the file. To establish inter-observer reliabilities, at least two coders scored 30 trials, with a reliability score of 0.92. All habituation trials were recoded as no-responses, and only five of the 26 test trials were scored differently from the on-line scores; for these trials, we entered the on-line coding response.

3. Results

Subjects required a mean number of 26.5 trials (SD = 9.1, min = 15, max = 45) to habituate. As a result, subjects were exposed to a considerable variety of exemplars from the particular numerical class. Although some subjects habituated more rapidly than others (Fig. 2), there was no evidence that subjects habituated to two-syllable exemplars (mean = 30.0, SD = 16.9, n = 6) differed from those habituated to three-syllable exemplars (mean = 27.1, SD = 8.1, n = 7; Mann–Whitney U = 19.5, P = 0.83).

Following habituation, there was a highly significant overall difference in responses to two-tone and three-tone stimuli depending upon whether subjects were habituated to two or three syllables. Specifically, subjects were more likely to respond when number changed (mean = 82.69%, SE = 6.23) from habituation to test than when number stayed the same (mean = 43.62%, SE = 4.27; Wilcoxon signed rank test = 3.04, P = 0.002). As illustrated in Fig. 3, there was no evidence that subjects habituated to two syllables were more or less likely to respond in the test trials than subjects habituated to three syllables.

4. Discussion

The results presented here provide the first controlled test of number discrimination in the auditory domain using untrained subjects, and the proper controls for continuous variables that can confound tests of number quantification. Although the set of habituation stimuli comprised a highly variable mixture of different speech sounds with variable
fundamental frequencies, intensities, and phonemic content, cotton-top tamarins readily habituated to the repeated presentation of either two- or three-speech syllables. Furthermore, the subsequent presentation of novel stimuli consisting of two or three tones led to renewed orienting, which was about twice as likely to occur if the number of tones differed from the previous number of syllables, than if it was identical. Because the stimuli were controlled for various non-numerical parameters such as duration and tempo, this orienting response to numerically different trials indicates that tamarins are sensitive to the number of auditory stimuli that they hear.

Our data speak to the abstractness of the numerical representations deployed in this task. The habituation phase consisted entirely of speech stimuli, while the test phase comprised
only tone stimuli. The content of the stimuli, their duration, and their tempo varied considerably. The fact that habituation transferred across such variations indicates that tamarins can discriminate auditory stimuli based on the number of events or elements, a capacity that abstracts away from changes in stimulus identity and temporal parameters. Tamarins appear able to represent the occurrence of two or three auditory events, independently of their content and timing. Our experiments leave open the exact nature of the representation underlying this behavioral performance. One possibility is that tamarins possess an auditory event file, analogous to the object file or indexing system (Kahneman et al., 1992; Pylyshyn & Storm, 1998; Hommel, 1998), which would list recently perceived auditory stimuli in short-term memory. Numerosity would then be represented only implicitly as the number of open event files, and with a limit of three or four events. Alternatively, tamarins might possess an explicitly numerical representation based on an accumulator which would tally the approximate number of stimuli (Dehaene & Changeux, 1993; Gallistel & Gelman, 2000; Meck & Church, 1983). This representation is not limited to small numerosities, but would follow Weber’s law and hence would become more imprecise for larger numerosities. Either of these representations might be specific to the auditory modality or highly abstract, extending cross-modally. Further experiments should test these alternative models by extending the current paradigm to larger numbers, as well as to cross-modal paradigms that involve habituation within one modality and testing in a different modality.

The most important conclusion from our study, however, is that auditory numerosities are spontaneously discriminated in cotton-top tamarins. The capacity to discriminate based on numerosities is sometimes considered as a ‘last resort hypothesis’ (Davis & Perusse, 1988), which should only be adopted once all other alternative possibilities have been rejected. Similarly, some researchers assume that genuinely numerical representations are normally not available to non-linguistic organisms, and that the construction of numerical representations from a “brain without numbers” (Simon, 1999) results from a slow process of learning and education. Our results and others (Feigenson, Carey, & Hauser, 2002; Xu & Spelke, 2000), by contrast, suggest that numerical representations are widely available in preverbal organisms and spontaneously evolve and mature without explicit training.

Acknowledgements

We thank the reviewers of Cognition for comments on the manuscript. For funding, M.D.H. thanks the Mind, Brain and Behavior initiative at Harvard, and a grant from the NSF (ROLE). S.D. was supported by a McDonnell Centennial Fellowship.

References


