Available online at www.sciencedirect.com
ScienceDirect

Cognition 105 (2007) 577–614

COGNITION

www.elsevier.com/locate/COGNIT

Perceptual constraints and the learnability of simple grammars ☆

Ansgar D. Endress [a,b,*], Ghislaine Dehaene-Lambertz [c,d], Jacques Mehler [a,b]

[a] *International School for Advanced Studies, Trieste, Italy*
[b] *Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-ENS-CNRS, Paris, France*
[c] *INSERM U562, Orsay, France*
[d] *Service de Neuropédiatrie, CHU Bicêtre, France*

Received 29 June 2005; revised 10 October 2006; accepted 8 December 2006

Abstract

Cognitive processes are often attributed to statistical or symbolic general-purpose mechanisms. Here we show that some spontaneous generalizations are driven by specialized, highly constrained symbolic operations. We explore how two types of artificial grammars are acquired, one based on repetitions and the other on characteristic relations between tones ("ordinal" grammars). Whereas participants readily acquire repetition-based grammars, displaying early electrophysiological responses to grammar violations, they perform poorly with ordinal grammars, displaying no such electrophysiological responses. This outcome is problematic for both general symbolic and statistical models, which predict that both types of

☆ The study was supported in part by an Allocation de Moniteur Polytechnicien fellowship to ADE, the HFSP Grant RGP 68/2002 and the Regione Friuli-Venezia-Giulia Grant (L.R. 3/98) to JM. The research was also supported in the framework of the European Science Foundation EUROCORES program "The Origin of Man, Language and Languages", and by two McDonnell foundation Grants to JM (21002089) and GDL. We are indebted to three anonymous reviewers for their detailed and insightful comments, and to L. Bonatti, E. Dupoux, H. Gleitman, L. Gleitman, T. Gliga, J. Pierrehumbert, F. Ramus, M. Shukla and A. Treves for discussions and suggestions.

* Corresponding author. Present address: Cognitive Neuroscience Sector, International School for Advanced Studies, Via Beirut n.2-4, 34014 Trieste, Italy. Tel.: +39 040 3787 603; fax: +39 040 3787 615.

E-mail address: ansgar.endress@m4x.org (A.D. Endress).

doi:10.1016/j.cognition.2006.12.014

grammars should be processed equally easily. This suggests that some simple grammars are acquired using *perceptual primitives* rather than general-purpose mechanisms; such primitives may be elements of a "toolbox" of specialized computational heuristics, which may ultimately allow constructing a psychological theory of symbol manipulation.

Keywords: Grammar acquisition; Perceptual primitives; Symbol manipulation; Statistical learning; Modularity; Learnability; Connectionism; Rule learning

1. Introduction

The idea that various aspects of cognition involve formal symbolic operations over variables is arguably one of the most important yet controversial themes in cognitive science. This is because the existence of symbolic mental operations implies the existence of innate representational constraints, since any finite set of examples is compatible with an infinite number of generalizations (e.g., Goodman, 1955; Hume, 1739/2003; Wittgenstein, 1953). Without such constraints, a learner would have infinitely more chances of picking a wrong generalization than of picking the correct one. For example, when learning from examples the rule "add 2", what prevents a learner from learning the rule "add 2 up to 1000; add 4 otherwise" – when *both* rules are compatible with the observed examples? Intuitively, the former solution is more "natural", and therefore preferable. However, to paraphrase a point made by Morgan (1986), completely unbiased learners cannot have any notion of "naturalness": for them, one generalization is as good as any other as long as it is compatible with the examples. Learners can thus be guaranteed to end up with the "correct" solution only if they are equipped with a reasonably rich set of constraints on the hypotheses they will consider.[1]

[1] More recently, several authors have proposed that a learner may choose a solution by appealing to its "simplicity" rather than to its "naturalness" (e.g., Chater, 1996, 1999; Chater & Vitányi, 2003; Pothos & Chater, 2002). The problem of "simplicity", however, is exactly the same as that of naturalness: It crucially depends on the constraints of the computational system which has to select a solution. For example, spatial rotations are easy for humans while divisions (e.g., 117/3 = 39) are hard, whereas the opposite is true for a computer. It thus seems that also simplicity is underspecified – unless the constraints of the learner are taken into account. This problem actually follows from the formal definition of "simplicity" used, for example, by Chater and Vitányi (2003), namely Kolmogorov complexity (KC). KC is essentially the length of the shortest program used to describe an object. Chater and Vitányi (2003) implicitly admit this problem by stating that "the choice of programming language does not matter [for computing KC], up to a constant additive factor." In fact, only the *difference* between the KCs in different languages is bounded by a constant; that is, if $K_1(x)$ and $K_2(x)$ are the KCs in two languages L_1 and L_2, then $|K_1(x) - K_2(x)| < C$, where C is a constant. The reason is that, with general-purpose programming languages, it is possible to simulate L_1 in L_2 and vice-versa, adding a constant overhead to the length of a program. This, however, does not guarantee that the two objects have the same *relative* KCs in two languages. Clearly, if L_1 has operations for multiplication and addition but L_2 only for addition, a program using multiplication will be longer than a program using addition in L_2 (because on has to write the multiplication operation first), but both programs will be equally long in L_1. Simplicity can thus not be defined a priori without deriving experimentally what actually counts as "simple" for the human computational apparatus.

In a more formal vein, learnability theory pursued similar questions. Gold (1967) showed for instance that languages simpler than natural ones cannot be learned from examples alone. Despite more recent results showing that some interesting classes of language are learnable from examples (e.g., Angluin, 1980; Morgan, 1986; Osherson, Stob, & Weinstein, 1984; Stabler, 1998; Wexler & Cullicover, 1980), the adjective "learnable" is in fact misleading: the learnability of a class of grammars implies only that there *exists* a learning procedure (that is, an idealization of the learners' predispositions) that can learn all grammars of this class – but not that an arbitrary learning procedure would succeed. Clearly, a learner who maps all possible inputs to the language that contains only the word "parrot" (which is a perfectly conceivable learning function) would have a hard time learning natural languages. Hence, language learning (and, in fact, any kind of learning) can only succeed if the learner is endowed with appropriate constraints.

Here we ask what kinds of constraints govern the acquisition of simple grammars. While many authors have suggested that learning may occur through monolithic general-purpose mechanisms that may be either statistical (such as neural networks; see e.g. Elman et al., 1996; McClelland, Rumelhart, & The PDP Research Group, 1986; Rumelhart, McClelland & The PDP Research Group, 1986; Seidenberg, 1997) or symbolic (such as digital computers; see e.g. Anderson, 1993; Marcus, 2001; Newell, 1980)[2], other authors have viewed the mind as a collection of heuristics (e.g., Gigerenzer, Todd, & The ABC Group, 1999) or as a "bag of tricks" (e.g., Ramachandran, 1990); in line with these views, non-human animals seem to be equipped with specialized computational mechanisms to solve the problems they face in their environment (e.g., Gallistel, 1990, 2000; Gould & Marler, 1987). In this paper, we investigate the possibility that such specialized computational devices provide constraints also for how human adults learn even simple grammars; rather than using statistical or symbolic general-purpose mechanisms, they may use simpler specialized *perceptual primitives* to extract structure from the stimuli.

The debate about the existence of mental rules has been invigorated by experimental demonstrations that infants and adults can deploy sophisticated statistical computational capacities. For example, they can cut a continuous speech signal down into (nonce) words when the only available cues to word boundaries are statistical properties of the syllable distribution (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). Sensitivity to statistical cues is ubiquitous. It arises in domains as different as speech-like stimuli, tones (e.g., Creel, Newport, & Aslin, 2004; Saffran, Johnson, Aslin, & Newport, 1999), timbres (Tillmann & McAdams, 2004) and visual configurations

[2] It is useful to clarify what we mean by "symbolic." According to Gallistel (2001), a "system of symbols is isomorphic to another system (the represented system) so that conclusions drawn through the processing of the symbols in the representing system *constitute valid inferences about the represented system.* […] A symbolic system contains […] symbols, rules that govern the manipulation of those symbols, and measuring processes [that] relate the numerical values of the symbols to the [quantities] to which they refer" (emphasis added). In other words, symbolic systems have also representations of *relations* among the representations of the items in the environment, although the former are not directly observable.

(Fiser & Aslin, 2001, 2002; Turk-Browne, Jungé, & Scholl, 2005), and it is shared with other mammals (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005). Still, such operations are computationally constrained; for example, they can be applied to consonants but not to vowels (Bonatti, Peña, Nespor, & Mehler, 2005, but see Newport & Aslin, 2004).

Other studies claimed to have found symbolic-like behaviors in young infants. Indeed, seven-month olds can learn the "grammars" ABB and AAB, where A and B represent different syllables that were duplicated either at the left or the right edge of a triplet (Marcus, Vijayan, Rao, & Vishton, 1999; see also Gómez & Gerken, 1999, for similar experiments). Participants were familiarized with items like "ga-na-na", and tested with items using *novel* syllables conforming either to AAB or to ABB; the infants discriminated familiar from novel grammars with the new items, from which Marcus et al. (1999) concluded that these repetition-based grammars were generalized using a symbolic mechanism.

Just like statistical computations such as those demonstrated by Saffran et al. (1996) have been taken as confirmations of proposals that cognition may rely on statistical general-purpose mechanisms (e.g., Bates & Elman, 1996), the acquisition of repetition-based grammars like in Marcus et al.'s experiments might confirm proposals attributing to the mind symbolic general-purpose computations like those in a digital computer; the learning of repetition-based grammars may thus be due to a mechanism that represents serial positions as variables and discovers relations between these variables. Indeed, Marcus et al. (1999) proposed that "infants extract abstract algebra-like rules that represent relationships between placeholders (variables), such as 'the first item X is the same as the third item Y', or more generally, that 'item I is the same as item J'" (p. 79). Such proposals fit well with the view that the mind is essentially analogous to a digital computer (e.g., Anderson, 1993; Marcus, 2001; Newell, 1980). For example, observing that "registers are central to digital computers", Marcus (2001) proposed that "registers are central to human cognition as well" (p. 55), and discussed how registers could be built into the nervous system (pp. 55–58).

Such results raise the possibility that infants may be endowed with inductive general-purpose machinery. However, as outlined above, computational general-purpose mechanisms cannot be guaranteed to choose the correct grammar among the infinity of grammars compatible with the input – which all infant learners clearly do (except in pathological cases): they learn English when exposed to English and French Sign Language when exposed to French Sign Language. In contrast to these general-purpose mechanisms, non-human animals may be endowed with a great variety of specialized symbolic computations (e.g., Gallistel, 1990, 2000; Gould & Marler, 1987). Honeybees, for example, learn the solar ephemeris (e.g., Dyer & Dickinson, 1994), different insects perform path integration (e.g., Collett & Collett, 2000), different birds compute dominance relations even using transitivity (that is, they deduce from $A < B$ and $B < C$ that $A < C$; e.g. Paz-Y-Miño C, Bond, Kamil, & Balda, 2004) and have specific predispositions for song learning (e.g., Gardner, Naef, & Nottebohm, 2005; Marler, 1997), and many animals show time scale invariance in conditioning experiments (Gallistel & Gibbon, 2000, 2002). Also the human mind may thus be endowed with specialized computational abilities; specialized mechanisms do not face learnability problems – since they presumably can acquire just the grammars they evolved to acquire.

Here we ask whether general-purpose computations appropriately describe the spontaneous acquisition of simple grammars in humans, or whether more specialized operations are required. As a case study, we ask whether repetitions like in the experiments of Marcus et al. (1999) are identified by symbolic general-purpose computations – or rather by simpler specialized low-level mechanisms. Indeed, different results suggest that repetition-based grammars may be generalized by a simpler mechanism. First, cotton-top tamarins can learn such grammars (Hauser et al., 2001), and even bees can learn identity and non-identity relations (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001).

Second, there is evidence from the Artificial Grammar Learning literature that repetitions may have a special status. In such experiments, participants are generally exposed to consonant strings that are generated from a finite-state grammar; after such a familiarization, they can classify new strings as grammatical or ungrammatical. In itself, such results do not speak to the question of how children learn natural languages, because computing associations among the consonants is sufficient to "learn" these finite-state grammars (e.g., Cleeremans & McClelland, 1991; Dienes, Broadbent, & Berry, 1991; Kinder, 2000; Kinder & Assmann, 2000). However, if participants could transfer their knowledge of such grammars to consonants that differ from those used during training, then one would be in a position to conclude that participants have acquired more abstract knowledge than associations among consonants. Indeed, participants can classify consonant strings with *new* consonants as grammatical or ungrammatical (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gómez, Gerken, & Schvaneveldt, 2000; Knowlton & Squire, 1996; Meulemans & van der Linden, 1997; Reber, 1969; Tunney & Altmann, 2001). However, subsequent research has shown that this transfer depends crucially on repetition patterns that arise through the grammars; no transfer occurs when the grammars avoid such repetition patterns (see e.g. Gómez et al., 2000; Tunney & Altmann, 2001; see also Brooks & Vokey, 1991). It is thus plausible that the use of repetitions may have facilitated generalizations also in Marcus et al.'s (1999) experiments.

In our experiments, participants had to learn grammars from piano tone triplets with varying pitch (see Fig. 1). (We use the term "grammar" just to indicate that participants had to learn structural relations, but, of course, the "grammars" we use are very different from those used by the language faculty.) In Experiment 1, participants had to learn the repetition-based grammars ABA and ABB (Fig. 1a and b), and in Experiments 2 and 3 the grammars Low-High-Middle (LHM) and Middle-High-Low (MHL), henceforth called *ordinal* grammars (Fig. 1c and d). We recorded the participants' responses and their evoked potentials with a high-density system of 129 channels.

We used an AAAAX paradigm, which is particularly well suited to study discrimination responses with ERPs. In this design, the first stimuli of the trials create a grammatical context to which participants must compare the last item to decide whether it follows the same grammar or not. Violations of the regularities in

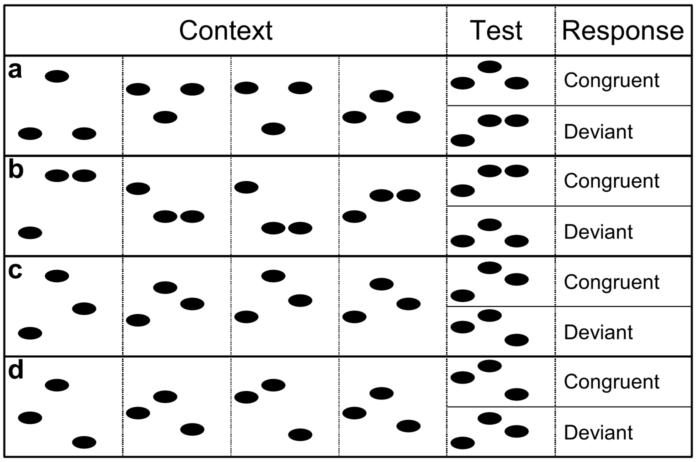

Fig. 1. Trials comprised of five triplets of piano tones. The first four triplets conformed to a common grammar, while the fifth one followed either the same grammar (congruent trials) or a different grammar (deviant trials). Participants had to indicate with button presses whether the fifth triplet followed the same grammar as the preceding triplets, and received feedback after each trial. In Experiment 1, the first four triplets conformed either to the grammar ABA (a) or to ABB (b); in both types of trials, the fifth triplet could follow either grammar. In Experiment 2 and 3, the first four triplets conformed either to the grammar Low-High-Middle (c) or to Middle-High-Low (d); in both types of trials, the fifth triplet could follow either grammar.

auditory stimuli by a deviant item are known to induce a particular electrical component, called the Mismatch Negativity (MMN). While it was first observed after changes of simple acoustic features such as the duration or the pitch of items (e.g., Giard et al., 1995; Näätänen & Alho, 1995), this component was observed also after changes based on more complex representations, such as conjunctions of features (Takegata, Paavilainen, Näätänen, & Winkler, 1999), the lexical and grammatical status of words (Shtyrov & Pulvermüller, 2002a, 2002b), and violations of arbitrary rules (Horváth, Czigler, Sussman, & Winkler, 2001). The MMN seems to be elicited irrespectively of whether the participants' attention is directed towards the auditory stimuli (Näätänen & Alho, 1995), and its duration is generally correlated with discrimination performance (Tiitinen, May, Reinikainen, & Näätänen, 1994). However, in a training task on subtle phonetic differences, Tremblay, Kraus, and McGee (1998) reported a change in the MMN *before* a behavioral performance improvement, suggesting that, in some situations, the MMN may be a more sensitive measure of learning than overt performance. Here, we combined behavioral measures with ERPs recordings to take advantage of their potentially greater sensitivity to learning. As we anticipated that some grammars may be more difficult to learn than others, we asked whether neurophysiological correlates of learning might be observed nevertheless, and, if so, how early they would arise in terms of their latency.

Before presenting Experiments 1–3, it is worth stressing that our predictions may clash with the intuition that repetition-based grammars appear to be easier to process than ordinal grammars. (After having presented the data, we will show that symbolic and statistical general-purpose mechanisms should learn the ordinal grammars better than, or at least as well as, the repetition-based grammars.) In fact, the aim of our experiments was to show that repetition-based grammars are computed via a specialized mechanism rather than via general-purpose mechanisms, and thus that general-purpose models do *not* account for our results. It is therefore possible that we have the intuition that repetition-based grammars may be simpler because we may be equipped with a specialized operation processing repetitions. The salience of repetitions, however, does not fall out of the formal or statistical structure of the stimuli – unless the intuition that repetitions are "special" is explicitly incorporated into a model.

2. Experiment 1: Acquiring repetition-based grammars

In this experiment, we studied the learning of the repetition-based grammars ABA and ABB. If repetitions are perceptually salient, such grammars should be learned by relatively early representations and elicit a MMN; in fact, violations of simple alternating rules based on the direction of pitch changes (up and down) have been shown to elicit a MMN (Horváth et al., 2001; Korzyukov, Winkler, Gumenyuk, & Alho, 2003; Paavilainen, Jaramillo, Näätänen, & Winkler, 1999). In contrast, if such grammars are learned by a symbolic general-purpose mechanism (e.g., Marcus, 2001; Marcus et al., 1999), one would expect later responses, as such general processes presumably cannot be accommodated by sensory representations.

2.1. Materials and methods

2.1.1. Participants

Twelve Italian participants (9 females, 3 males, mean age 23.5, range 19–28) were tested individually after giving written informed consent. They were all right-handed according to self-report and the Edinburgh inventory; they reported neither hearing deficits nor neurological or psychiatric diseases. Two participants had to be excluded because of too many trials contaminated by movement or eye blink artifacts.

2.1.2. Materials

Ten piano tones were computer-generated with duration of 400 ms. They were first generated as MIDI files using a custom program, and then converted to wave files using TiMidity++ (http://timidity.sourceforge.net/); we used stereo files with a sampling rate of 16,000 Hz, 32 bit sample width and signed linear encoding. The lowest tone was A^b at 103.8 Hz and the other tones multiples of quarts above it in Lydian mode. The Lydian scale is identical to the major scale, except that its fourth tone is raised by a semi-tone. We selected this mode because it is rare, and may thus well be perceived as atonal, especially in the way our stimuli were generated.

Dowling and Fujitani (1971) showed that atonal melodies sharing their contour[3] are difficult to discriminate (see also Trehub, Schellenberg, & Kamenetsky, 1999, for similar results with infants). This is not the result of perceptual problems, but rather of how relations between tones in "usual" melodies tend to be encoded; it thus allows us to assess how different *arbitrary* relations among tones are learned. If similar difficulties apply also to extracting the grammars of triplets, participants may perform better for the repetition-based grammars than for the ordinal grammars (as both ordinal grammars have the same contour).

Tones were combined by three into triplets with no silence between tones. (Repeated tones were still perceived as two tones rather than a single, long, tone, because the amplitude of piano tones decays over time.) Triplets followed the grammars ABA or ABB. Intervals could be both raising and falling. Because ERPs are very sensitive to low-levels properties, the same test triplets were used in the congruent and in the deviant condition. We thus reserved four triplets conforming to each grammar as test triplets. Statistical cues to the grammars were controlled for by making sure that the tones and intervals that occurred in the test triplets had occurred equally often in triplets conforming to either grammar. For example, the tone G (that was used in test triplets) occurred equally often in ABA and ABB triplets. Hence, neither pitch nor intervals were associated with one grammar or the other.

2.1.3. Procedure

The procedure is illustrated in Fig. 1a and b. Trials comprised of five triplets of piano tones separated by a silence of 400 ms. The first four triplets constituted the context and conformed to a common grammar (e.g. ABA), while the fifth one, the test triplet, followed either the same grammar (in congruent trials) or a different grammar (in deviant trials). Participants had to indicate with button presses whether the test triplet followed the same grammar as the preceding triplets or not. They received feedback after each trial. Deviant and congruent trials were randomly presented; the context triplets could conform to both grammars and were randomly intermixed throughout the experiment. A computer screen indicated the number of the context triplets (1, 2, 3 or 4), showing a question mark for test triplets. Participants completed 312 trials; trial order was randomized between participants. Stimuli were presented over loudspeakers. The experiment was run on a PC running the Windows™ 98 operating system using the E-Prime software package (Psychology Software Tools, Pittsburgh, PA).

2.1.4. Recording system and data analysis

EEG was recorded from 129 carbon electrodes (EGI recording system) referenced to the vertex. Scalp voltages were amplified, low-pass filtered at 100 Hz and digitized at 250 Hz. The signal was then digitally filtered between 0.5 and 20 Hz. Epochs starting 200 ms before the onset of the last tone of each test triplet and ending 800 ms after

[3] The contour of a melody is the sequence of the direction of the pitch changes. For example, in the melody "1 5 4 7 6 3 2 4" (where the numbers denote pitch levels), the contour would be "up down up down down down up."

it were extracted from the continuous signal. These epochs were automatically edited to reject trials contaminated by eye or body movements (voltage exceeding thresholds of 30 μV on electrodes surrounding the eyes, 30 μV (local deviations) and 80 μV (global deviation) on the other electrodes). The artifact-free trials were averaged for each subject in two conditions, congruent and deviant. Finally averages were baseline corrected (200 ms) and transformed into reference-independent values using the average reference method. Two-dimensional reconstructions of scalp voltage at each time step were computed using a spherical spline interpolation.

The synchronized activity of columns of neurons that are at the origin of the scalp event-related responses can be characterized as electric dipoles. High-density scalp recordings can usually distinguish the positive and negative poles of the corresponding electric fields because of their dense spatial sampling, and because using the average voltage as the reference sets the average voltage on the scalp to zero. Therefore, in order to study event-related responses, clusters of electrodes at the maxima of a topographical dipole configuration are chosen (Michel et al., 2004).

As the test stimuli were identical in congruent and deviant trials, any significant difference between the waveforms would indicate that, in deviant trials, a change was detected with respect to the context stimuli. We thus inspected the time-course of two-dimensional reconstructions of the *t*-test values in the comparison of deviant and congruent trials in order to isolate the time-windows in which a significant dipolar response was present. This was the case for two time windows (100–200 and 408–588 ms). For the two time-windows separately, voltage was averaged across two groups of nine contiguous electrodes chosen at the maxima of the negativity and the positivity of the dipole topography, and their symmetrical counterparts on the other hemisphere, and then entered in an analysis of variance (ANOVA) with congruence (deviant and congruent), electrode group (anterior: around FZ and posterior: around the occipital regions 01 and 02) and hemisphere (left and right) as within-participant factors. Because of the voltage inversion between the selected electrodes, a main effect of condition is not interpretable. Therefore, only interactions between Congruence, Electrodes and Hemisphere were examined. As we never found a main effect of Hemisphere nor an interaction with this factor, we will report below only the interaction between Congruence and Electrodes.

2.2. Results and discussion

As shown in Fig. 2, participants successfully learned the grammars ABA and ABB (percentage of correct responses: $M = 78.7\%$, $SD = 13.3\%$), $t(11) = 7.5$, $p = 0.00001$. (Statistical tests are two-tailed throughout this article with a chance level of 50%.) In order to test whether participants learned both grammars, we analyzed trials with the context grammar ABA and trials with the context grammar ABB separately; participants performed well above chance both for ABA ($M = 76.2\%$, $SD = 15.4\%$), $t(11) = 5.9$, $p < 0.0001$, and for ABB ($M = 81.1$, $SD = 12.0\%$), $t(11) = 9.0$, $p = 2.07 \times 10^{-6}$.

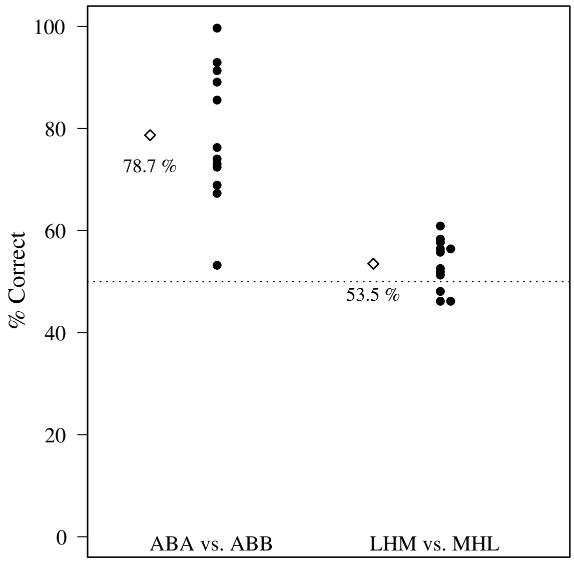

Fig. 2. Participants learned the grammars ABA and ABB (Experiment 1, left) better than the grammars LHM and MHL ("ordinal grammars"; Experiment 2, right). Dots represent the means of individual participants, diamonds population averages and the dotted line the chance level of 50%.

As shown in Fig. 3a and c, we found significant brain responses to grammar changes from 100 to 200 ms after the onset of the third tone of the test triplet, $F(1,9) = 11.5$, $p = 0.008$, and from 408 to 588 ms after the onset of the third tone of the test triplet, $F(1,9) = 11.0$, $p = 0.009$; both responses yielded an anterior negativity and a posterior positivity. The latency and the topography (negativity over the frontal areas with a reverse of polarity along the temporal axis) of the first effect is compatible with a mismatch negativity (e.g., Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). As such responses seem to be independent of attention and to reflect relatively automatic processes, the extraction of repetition-based grammars may possibly be embedded in relatively early sensory representations.[4]

[4] Some test items in Experiment 1 occurred with different frequencies in different conditions; we thus replicated this experiment with a new counterbalancing. Again, participants performed significantly above chance ($M = 83.9\%$, $SD = 17.9\%$), $t(16) = 7.8$, $p < 0.001$, both when the context grammar was ABA ($M = 83.6\%$, $SD = 18.0\%$), $t(16) = 7.7$, $p < 0.001$, and when it was ABB ($M = 84.2\%$, $SD = 18.1\%$), $t(16) = 7.8$, $p < 0.001$. We also observed an early electrophysiological response to grammar changes. This response was significant in the time window from 50 to 180 ms after the last tone of the test triplet, $F(1,10) = 15.7$, $p < 0.003$. The response occurred earlier and lasted longer than in the original experiment, probably because the participants' reaction times were faster and more variable (mean reaction times 860 vs. 959 ms after the last tone of the test triplet; SDs: 322 vs. 184 ms, $F(16,11) = 2.99$, $p = 0.036$). Below, we will report simulations of our experiments; also the simulation results for the replication are similar to those for the original experiment.

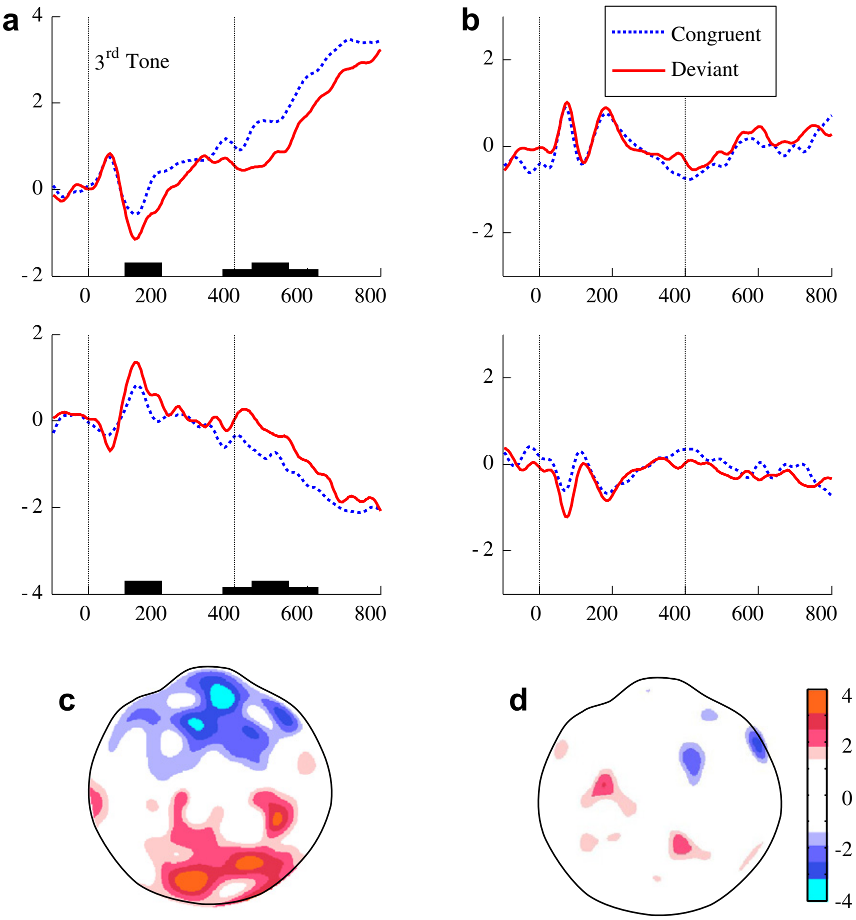

Fig. 3. (a,b) Mean voltage of the electrode groups used for statistical analysis in Experiments 1(a) and 2(b). Upper and lower figures represent anterior and posterior electrode groups, respectively. Epochs start 200 ms before the onset of the *third* tone of test triplets and end 800 ms afterwards. Black bars represent statistical significance. (c,d) Cartography of *t*-values 120 ms after the onset of the third tone in Experiment 1 (c) and 2 (d).

3. Experiment 2: Acquiring ordinal grammars

Experiment 1 suggests that participants readily detect changes in repetition-based grammars, and that such changes elicit a response at an early processing stage. In Experiment 2 below, we asked whether such results would hold also for other simple grammars based on “ordinal” relations not entailing repetitions. If the repetition-based grammars are processed by a symbolic general-purpose mechanism, the ordinal grammars should be processed as readily as the repetition-based grammars; in contrast, if participants are endowed with a specialized mechanism processing repetitions, they should experience more difficulties for processing the ordinal grammars.

3.1. Materials and methods

Experiment 2 was identical to Experiment 1 except that participants had to learn the ordinal grammars instead of the repetition-based grammars, and that it comprised of only 156 trials.

3.1.1. Participants

Twelve French participants (6 females, 6 males, mean age 22.8, range 20–33) were tested individually after giving written informed consent. They were all right-handed according to self-report and the Edinburgh inventory; they reported neither hearing deficits nor neurological or psychiatric diseases. One participant had to be excluded because of too many trials contaminated by movement or eye blink artifacts.

3.1.2. Materials

Triplets were generated as in Experiment 1 but followed the grammars Low-High-Middle or Middle-High-Low. That is, triplets were created using ten computer-generated piano tones of 400 ms, with the lowest tone at 103.8 Hz and the other tones multiples of quarts above it in Lydian mode. Three triplets conforming to each grammar were reserved as test triplets. As in Experiment 1, statistical cues to the grammars were controlled for by making sure that the tones and intervals that occurred in the test triplets had occurred equally often in context triplets conforming to either of the grammars.

3.1.3. Procedure

The procedure was identical to the one used in Experiment 1, except that the experiment comprised of 156 trials, and that the experiment was run on a PC under MS Dos with the EXPE software package (Pallier, Dupoux, & Jeannin, 1997).

3.1.4. Recording system

The recording system was the same as in Experiment 1.

3.1.5. Data analysis

The data analysis procedure was the same as in Experiment 1.

3.2. Results and discussion

As shown in Fig. 2, the mean percentage of correct responses was better than chance ($M = 53.5\%$, $SD = 4.9\%$), $t(11) = 2.5$, $p = 0.032$, but substantially worse than in Experiment 1, $F(1,22) = 38.1$, $p = 0.000003$. When compared to Experiment 1, participants performed worse for the ordinal grammars both relative to trials with context grammar ABA, $F(1,22) = 23.7$, $p < 0.0001$, and relative to trials with context grammar ABB, $F(1,22) = 54.9$, $p = 2.04 \times 10^{-7}$. No reliable electrophysiological responses to grammar changes were observed. These results seem to suggest that it is easier to learn repetition-based grammars than ordinal grammars.

Before accepting this conclusion, however, it is necessary to rule out another explanation of the differences between Experiments 1 and 2. First, Experiment 1 consisted of 312 trials while Experiment 2 had only 156 trials; participants may thus have learned the repetition-based grammars better because Experiment 1 included more trials. However, when considering only the first 156 of Experiment 1, participants still performed better in Experiment 1 than in Experiment 2, $F(1,22) = 44.8$, $p = 9.93 \times 10^{-7}$. Fig. 4 shows the average performance for sliding windows of 20 trials during the first 156 trials in Experiments 1 and 2. Except for the very first trials, the performance was always better in Experiment 1 than in Experiment 2. Hence, the different number of trials in these experiments does not explain the advantage for repetition-based grammars.

Another possible confound is that participants may experience psychophysical or attentional difficulties when faced with the ordinal grammars. However, the intervals we used were several orders of magnitude above the discrimination thresholds (e.g., Sinnott & Aslin, 1985); it thus seems rather unlikely that participants would experience such difficulties. Moreover, one still would have to explain why these difficulties should exist only for the ordinal grammars but not for the repetition-based grammars, as the tones and the intervals used were exactly the same. Still, in order to rule out this possibility, we included in Experiment 3 a

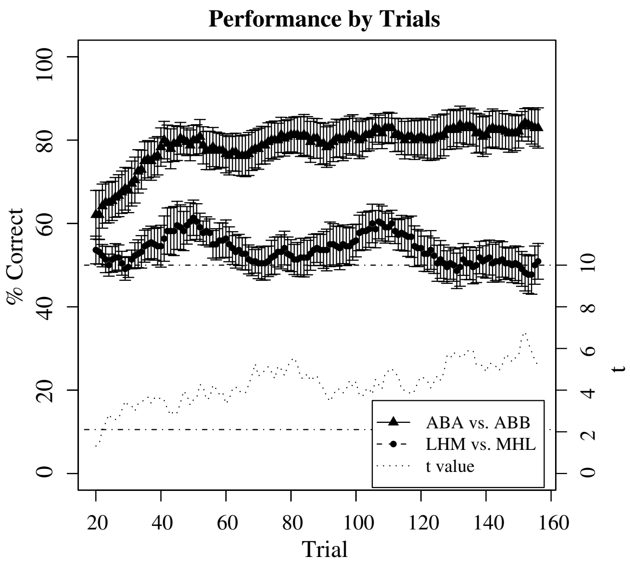

Fig. 4. Percentage of correct responses averaged over sliding windows of 20 trials. The *t*-value corresponds to the *t*-test between the performance for repetition-based grammars and ordinal grammars at each time point. The slash-dotted lines represent the chance level of 50% and the significance threshold for the *t*-test; the error bars represent the standard error. Except for the very first trials, the performance is always better for the repetition-based grammars than for the ordinal grammars.

condition where participants had to *discriminate* triplets from each other without the requirement to generalize their grammar; if the advantage for repetition-based grammars were due to psychophysical or attentional problems (that would have to occur only with the ordinal grammars but not with the repetition-based grammars), participants should experience such difficulties also when discriminating triplets.

Finally, the advantage for repetition-based grammars may possibly be due to the *similarity* of the intervals in the triplets. In the grammar ABB, the last interval is always the same; hence, two triplets conforming to the grammar ABB have very similar second intervals (in fact, they are identical). In contrast, the ordinal grammars (but also ABA) allow different intervals in the second position. To rule out this confound, we include in Experiment 3 a condition where the context and the test triplets were *transpositions* of each other, that is, the intervals in the triplets were kept constant throughout congruent trials. (In deviant trials, the test triplet was obviously not a transposition of the context triplets.) If the similarity of the intervals in the context and the test triplets were crucial to the good performance in Experiment 1, we would expect participants to perform well also in this transposition condition.

4. Experiment 3: Acquiring ordinal grammars under simplified conditions

The joint results of Experiments 1 and 2 suggest that participants process repetition-based grammars more easily than ordinal grammars. Experiment 3 was designed to address a number of alternative interpretations of the outcome of Experiments 1 and 2, and to assess whether participants would perform better with ordinal structures under more simplified conditions. First, we attempted to facilitate learning of the ordinal grammars by blocking the trials by the grammar of their context triplets; that is, for half of the participants, the grammar of the context triplets was LHM in the first half of the experiment and MHL in the second half. The order of the context grammars was counterbalanced across participants. Second, we controlled for attentional and psychophysical problems by including the *Identity Condition*. In this condition, the context triplets in a trial were physically identical; the test triplet was then either identical to the context triplets, or it conformed to the other grammar. Participants thus had just to *discriminate* triplets, and were not required to acquire their grammars. Third, we controlled for the possibility that the advantage for repetition-based grammars in Experiments 1 and 2 may have been due to the similarity between the intervals in the context triplets and the test triplets. For this purpose, we included the *Transposition* Condition, where the context triplets were transpositions of one another (i.e., the intervals within the triplets remained constant), and the test triplet either was again a transposition of the context triplets, or conformed to the other grammar. Finally, the *Grammar Condition* was analogous to Experiment 2: The context triplets shared only the grammar whereas the test triplet either conformed to the same grammar or not.

4.1. Materials and methods

4.1.1. Participants

Sixteen French-speaking participants (10 females, 6 males, mean age 23.6, range 21–27) were tested individually after giving written informed consent. They were all right-handed according to self-report and the Edinburgh inventory. None of them had a history of neurological, or psychiatric disease or a hearing deficit. Four participants had to be excluded from electrophysiological analysis due to too many movement or eye blink artifacts or technical problems.

4.1.2. Procedure

The procedure was similar to the one of Experiment 2; only differences will be described here. In this Experiment, participants had to discriminate the melodic grammars *Low-High-Middle* (*LHM*) and *Middle-High-Low* (*MHL*). The paradigm was again an AAAAX paradigm. In contrast to Experiment 2, participants were exposed to three intermingled experimental conditions: In the *Identity Condition*, the four context triplets in a trial were physically identical; the test triplet was either also identical to the context triplets, or conformed to the opposite grammar (e.g., MHL if the context items conformed to LHM).

In the *Transposition Condition*, the context triplets were transpositions of one another (i.e., the intervals between the tones in a triplet remained constant); the test triplet was either another transposition of the context triplets, or it conformed to the opposite grammar. The context triplets were generated by transposing the triplets reserved as test triplets (see below) by 7 to 11 semi-tones upward or downward.

In the *Grammar Condition*, the context triplets shared only the grammar and varied in their tones and intervals; the test triplet either conformed to the same grammar as the context triplets or to the opposite one. This condition was analogous to Experiment 2.

The trials were blocked by context grammar. That is, the context triplets in the first half of the experiment conformed to a common grammar, and to another grammar in the second half of the experiment; the order of the context grammars was counterbalanced across participants. This was intended to facilitate the learning of the grammars. Five triplets were reserved as test triplets. The experiment comprised of 120 trials per condition. All conditions were intermingled in the two blocks. Stimuli were presented over loudspeakers. The experiment was run on a PC under MS Dos with the EXPE software package (Pallier et al., 1997).

4.1.3. Recording system

The recording system was the same as in Experiment 1 and 2, except that the data were digitized at 125 Hz.

4.1.4. Data analysis

The procedure for data analysis was the same as in Experiment 2 except that trials were segmented into epochs starting 200 ms before the onset of the *first* tone of each test triplet and ending 1800 ms after it. While participants in Experiments 1 or 2 could notice a grammar change only after the third tone of the test triplet (because it was only then that they could know whether the grammar was ABA and ABB), they could notice it already after the *first* tone of the test triplet in the Identity condition of Experiment 3, as this tone was already deviant.

4.2. Results

As shown in Fig. 5, the condition (Identity, Transposition or Grammar) yielded a significant main effect of the proportion of correct responses, $F(2,30) = 241.7$, $p < 2.2 \times 10^{-16}$ (repeated-measure ANOVA). Participants performed better in the Identity Condition ($M = 95.2\%$, $SD = 5.1\%$) than in the Transposition Condition ($M = 59.7\%$, $SD = 7.7\%$), $t(15) = 17.5$, $p = 2.2 \times 10^{-11}$ (paired t-test), or the Grammar Condition ($M = 54.5\%$, $SD = 5.9\%$), $t(15) = 19.8$, $p = 3.5 \times 10^{-12}$ (paired t-test); there was no difference between the latter two conditions, $t(15) = 2.8$, $p = 0.053$, ns. In all cases, participants performed better than chance (Identity Condition: $t(15) = 35.2$, $p = 7.7 \times 10^{-16}$; Transposition Condition: $t(15) = 5.1$, $p < 0.002$; Grammar Condition: $t(15) = 3.1$, $p < 0.009$). Crucially, participants performed better in Experiment 1 than in the Grammar Condition of Experiment 3, $F(1,26) = 42.6$, $p = 6.5 \times 10^{-7}$, and there was no difference between the Grammar Condition of this experiment and Experiment 2, $F(1,26) = 0.08$, $p = 0.784$, ns. Finally, participants performed better in Experiment 1 than in the Transposition Condition of Experiment 3, $F(1,26) = 22.7$, $p < 6.3 \times 10^{-5}$.

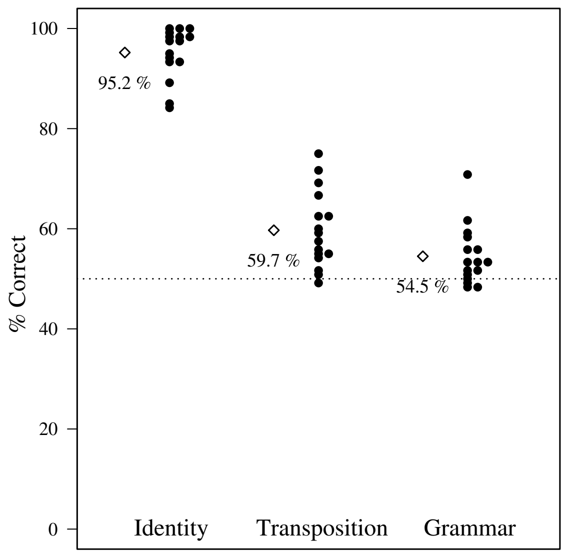

Fig. 5. Behavioral results in Experiment 3. Dots represent the means of individual participants, diamonds population averages and the dotted line the chance level of 50%. In the Identity Condition, where participants just had to discriminate triplets, the participants' performance was almost perfect. In contrast, participants performed much worse in the Transposition Condition (where context and test triplets were transpositions of each other) or in the Grammar Condition; there was no difference between the latter two conditions.

In the ERPs, the first tone of the test triplet was, as expected, sufficient to detect a mismatch in the Identity Condition, and a first brain response to a triplet change was observed starting at 200 ms after the onset of the first tone of the test triplet, $F(1,11) = 8.3$, $p = 0.015$. Other responses with a similar topography occurred 144 ms after the onset of the second tone, $F(1,11) = 7.5$, $p = 0.019$, and 88 ms after the onset of the third tone, $F(1,11) = 8.0$, $p = 0.016$. A MMN was thus present after each tone, with decreasing latencies as the deviance increased. A late difference was also recorded 464 ms after the onset of the third tone, $F(1,11) = 8.0$, $p < 0.001$. In the Transposition Condition, the grammar change cannot be detected before the second tone, because this tone is necessary to compute the interval between the first two tones, which signals a change in this condition. Brain responses to changes were observed 328 ms after the onset of the second tone, $F(1,11) = 9.0$, $p = 0.012$, 320 ms after the onset of the third tone, $F(1,11) = 16.1$, $p = 0.002$, and 568 ms after the onset of the third tone, $F(1,11) = 13.2$, $p = 0.004$. In the Grammar Condition, we observed no reliable response to a grammar change.

4.3. Discussion

Like in Experiment 2, the Grammar condition required participants to learn the ordinal grammars. Again, they performed only slightly above chance and displayed no reliable brain responses to a grammar change. These observations are difficult to explain by psychophysical or attentional problems because participants were virtually perfect in the Identity Condition, and, in any case, the intervals we used were several orders of magnitude above the discrimination thresholds (e.g., Sinnott & Aslin, 1985). Likewise, the Transposition Condition showed that the advantage for repetition-based grammars did not arise from a higher similarity among intervals in the grammar ABB than in the ordinal grammars; in fact, in each trial, the intervals in the context triplets and in the test triplet (at least in congruent trials) were *identical* (and thus as similar as they could possibly be); hence, if participants learned the grammars by computing the similarity among intervals in context and test triplets, they should have performed even better than in Experiment 1, but, in fact, the performance in Experiment 1 was better than in the Transposition Condition.

The Transposition Condition controls also for a related possibility. In the grammars ABA and ABB, the last tone is predictable from the first two ones, while the first tones predict only a *range* of tones in the ordinal grammars. In the Transposition Condition, however, the third tone is predictable from the first two; hence, if the advantage for repetition-based grammars were due to the predictability of the last tone in the repetition-based grammars, participants should perform as well in the Transposition Condition as in Experiment 1, which is not what we found.

ERP results paralleled the behavioral results, showing clear MMNs in the Identity condition, while we did not observe any difference between deviant and congruent trials in the Grammar Condition. In the Transposition Condition, in contrast, although the behavioral performance was similar to the Grammar Condition, the ERP results suggest that participants detected some regularity (although they could not use it for improving their overt responses). Dissociations between an early sensory detection and a lack of overt performance have already been described, for example for the attentional blink, where conscious processing of a second stimulus is blocked by the processing of a first stimulus (e.g., Luck, Vogel, & Shapiro, 1996; Sergent, Baillet, & Dehaene, 2005). Whatever may explain this discrepancy between the behavioral and the ERP results in the Transposition Condition, repetition-based grammars seem to be sufficiently robust to withstand such problems, suggesting again that they may be processed by a specialized operation.

5. Rationale of computing the predictions of general-purpose mechanisms

The results of Experiments 1–3 suggest that repetition-based grammars are learned more readily than ordinal grammars. We will now compute the predictions of symbolic and statistical general-purpose mechanisms for these experiments. Our operational definition of "general-purpose mechanism" is that such a mechanism should behave in some sense "optimally." For symbolic general-purpose mechanisms, this means that they should find the simplest (symbolic) relations among tones that can be used to distinguish the grammars in an experiment. The simplicity of a solution can be measured by the number of required operations. Statistical general-purpose mechanisms, in contrast, should exploit the statistical dependencies that are present in the stimuli; the "stronger" these dependencies are, the better the performance. It is worth stressing that our models are not intended to be psychologically valid; we just use them to ask what kinds of computational principles may be required to explain our data. In fact, the predictions of these models may clash with the intuition that repetition-based structures may be particularly simple. This, however, does not vitiate their predictions; rather, it highlights that such models cannot explain why one would have this intuition in the first place, suggesting again that repetitions may be processed by a specialized mechanism.

6. Predictions of symbolic general-purpose mechanisms

A rule-extraction mechanism representing sequential positions as variables and establishing diagnostic relations between them (Marcus et al., 1999) should represent the three tones in a triplet as a sequence of three variables XYZ, and find relations between these variables that characterize the grammars. Such a mechanism should discover that ABA can be defined by the relation $X=Z$, and that the relation $X<Z$ is sufficient to distinguish LHM and MHL. Hence, it should discover that it is sufficient to compute one relation to discriminate the grammars in either experiment.[5] A symbolic

[5] This conclusion is obviously valid only if humans actually can process lower-than/higher-than relations. This, however, seems to be the case; indeed, violations of a sequence of intervals sharing their direction (up or down) seem to elicit a Mismatch Negativity (e.g., Horváth et al., 2001; Korzyukov et al., 2003; Paavilainen et al., 1999), which is thought to reflect pre-attentive processing. Hence, lower-than/higher-than relations can be extracted pair-wise, and all relevant comparison operations should thus be available.

general-purpose mechanism endowed with these operations should therefore learn the grammars independently of whether they entail repetitions or not.

Still, it may be easier to process identity relations than other intervals, for example because the third tone in triplets with the grammar ABB is predicted with certainty from the first two tones, while the grammar Low-High-Middle allows a *range* of tones in the last position. This possibility is certainly intuitive, but it is misleading. In fact, it may just as well be harder to process *identical* tones, for example because detecting the identity of two tones always requires checking more precisely whether the tones may differ by a small amount, although they were perceived as roughly the same. None of these intuitions, however, follows from the formal structure of the relations. For a general-purpose mechanism such as a computer, both types of grammars just entail applying one operation to two of the tones, and to ignore a third one: For discriminating the repetition-based grammars, it is sufficient to compute an "equal-to" relation ("Is $X = Z$?"), while it is sufficient to compute a "less-than" relation ("Is $X < Z$?") to discriminate the ordinal grammars. Hence, both types of grammars should be equally easy to learn. In fact, in a computer, the "less-than", "greater-than" and "equal-to" operations are usually implemented in exactly the same way, namely by checking the sign of the difference "$X–Z$," which is less than, equal to or greater than 0 if X is less than, equal to or greater than Z, respectively. (On some systems, the equality relation is actually more complex than the other two relations, because "equal-to (x,y)" is implemented as the negation of "less-than (x,y) OR greater-than (x,y)".) Hence, a symbolic general-purpose mechanism should learn the ordinal grammars as well as the repetition-based grammars.

7. Predictions of statistical general-purpose mechanisms

We will now show that statistical general-purpose mechanisms should either learn the ordinal grammars more readily than the repetition-based grammars, or perform equally well for all grammars. While the exact predictions vary somewhat depending on the models, it seems fair to conclude that there is no a priori reason why such mechanisms should process the repetition-based grammars better than the ordinal grammars.

Statistical general-purpose mechanisms do not have other representational constraints than the ability to encode the pitch in some way, and to compute associations among pitch levels and the responses; they learn by detecting statistical relations among their inputs, that is, departures from statistical independence. We will first calculate the predictions of such models using mutual information, a very general measure of statistical dependence (e.g., Schneidman, Bialek, & Berry, 2003). Then, we will consider models that have been proposed to account for Marcus et al.'s (1999) results Fig. 6.

7.1. Mutual information between pitch levels and grammars

We obtained a rather model-independent prediction for statistical mechanisms by computing the mutual information (MI) between the grammars and

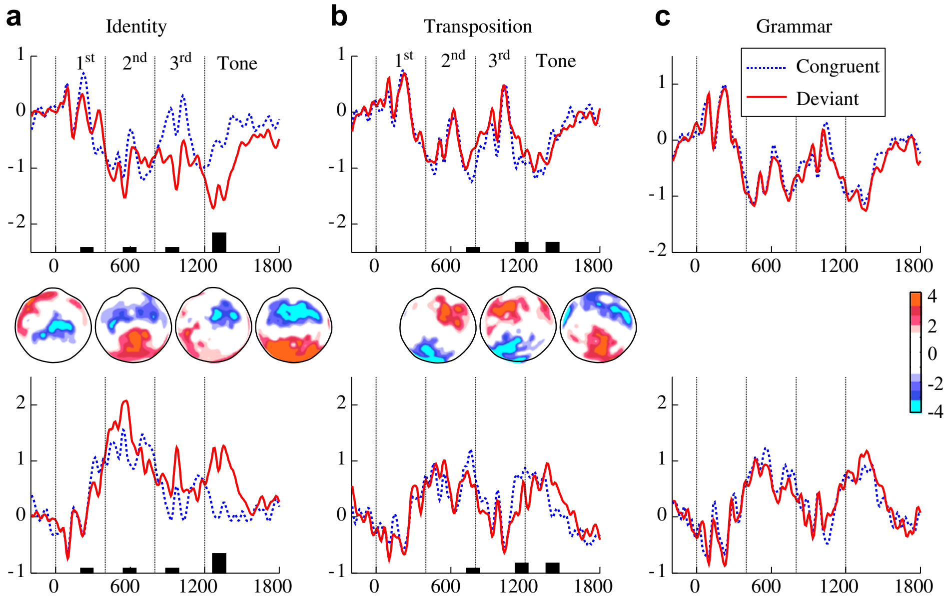

Fig. 6. Electrophysiological results of Experiment 3. The upper curves show the mean voltage for an anterior electrode group and the lower curves for a posterior electrode group. The electrode groups were the optimal groups for the response to a triplet change in the Identity Condition 888 ms after the onset of the first tone of the test triplet, but were found to illustrate also the observed responses in the other conditions. Voltages are given in μV. The topographies show the distributions of *t*-values on the scalp for the maxima of the responses in the different conditions, (a) Identity Condition, (b) Transposition Condition, (c) Grammar Condition.

arbitrary combinations of pitch levels in the three tones of a triplet. MI indicates how well the pitch levels predict the grammar of the corresponding triplet.[6] (In our case, the MI is already normalized by its maximal value (the entropy of the grammars), because this entropy is 1.) Since participants might have monitored not only the presence of pitch levels but also their absence, we evaluated combinations of pitch levels in addition to the individual pitch levels. For example, participants might have monitored whether the lowest pitch level occurred as the first tone, whether the highest pitch occurred as the second tone, or whether any of all ten pitch levels could occur as the second tone. We computed the MI

[6] Consider a system with two variables X and Y, representing the states (on or off) of a lamp and the corresponding switch. There are four (logically) possible states of the system (switch on/off × lamp on/off), and hence two yes/no questions to be asked to determine its state. But when considering the variables together, there are only two possible states because the state of the switch predicts the state of the lamp. Hence, one has to ask only *one* yes/no question in order to determine the state of the system. The saving of one yes/no by considering the variables together is their MI.

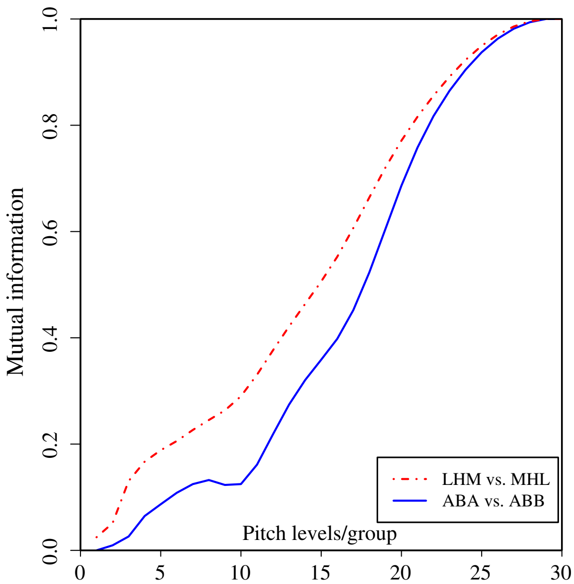

Fig. 7. The mutual information between grammars and combinations of pitch levels is higher for the grammars Low-High-Middle (LHM) and Middle-High-Low (MHL; red slash-dotted line) than for the grammars ABA and ABB (blue continuous line) for all group sizes. Hence, the tones are more predictive of the ordinal grammars than of the repetition-based grammars; statistical mechanisms should thus process the ordinal grammars better than the repetition-based grammars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper)

between all possible combinations of pitch levels and the grammars. For example, we computed the MI between the first pitch level of the first tone (which could be present or absent in a given triplet) and the grammar the triplet obeyed to. We computed the MI explicitly for combinations with 1 and 2 pitch levels, and ran Monte-Carlo simulations for all other group sizes. As shown in Fig. 7, for all group sizes, the MI was higher for the ordinal grammars than for the repetition-based grammars ($t(29) = 3.16$, $p < 0.004$ and $t(434) = 13.1$, $p < 2.2 \times 10^{-16}$, paired t-tests, for group sizes 1 and 2, respectively, and smaller p-values for the other group-sizes). Similar results were obtained for the MI between the three *positions* in triplets (1st, 2nd, 3rd) and the grammars when the pitch levels were considered as the possible states of the three positional variables (e.g. the possible values of the variable “first position” were all pitch levels): the MI between all positions and the repetition-based grammars was 0, while the MI between the first and the third position and the ordinal grammars was 0.334 and the between the second position and the ordinal grammars 0. Hence, the pitch levels are more predictive of the ordinal grammars than of the repetition-based grammars, and statistical general-purpose mechanisms (for which statistical relations between pitch levels are the *only* source of information) should perform better for the ordinal grammars than for the repetition-based grammars.[7,8]

7.2. Other models

Immediately after the publication of Marcus et al.'s (1999) experiments, a wealth of statistical models has been devised to account for their data without postulating symbolic operations (e.g., Altmann & Dienes, 1999; Christiansen & Curtin, 1999; Christiansen, Conway, & Curtin, 2000; Gasser & Colunga, 2000; McClelland & Plaut, 1999; Negishi, 1999; Seidenberg & Elman, 1999a, 1999b); it is therefore important to ask whether these models may explain also our data. However, each of these models has been criticized on various grounds (see Marcus, 1999a, 1999b, 1999c, 1999d, 1999e, for individual replies; see Shultz & Bale, 2001, for a similar list of criticisms). We will not review the models and their problems here (see Shultz & Bale, 2001, for a review), but these criticisms suggest that the models do not provide an alternative explanation for Marcus et al.'s (1999) data. Instead, we will focus on two more recent models that have not yet been evaluated (Altmann, 2002; Shultz & Bale, 2001). In short, we will show that, by their own criteria, Shultz and Bale's (2001) model implements a way to learn relations among variables. We then show that Altmann's (2002) model cannot reproduce the participants' behavior.

Shultz and Bale (2001) trained their model to reproduce the activation of the input units on the output units using a version of the cascade correlation algorithm (Fahlman & Lebiere, 1990). They encoded sentences as vectors of six continuous-valued elements, corresponding to the sonority of the six phonemes of a sentence. They thus assume that the position of each phoneme in a sentence is encoded by a dedicated neuron, whose activations have quasi-continuous values. However, they have criticized that such a coding scheme actually encodes phoneme positions as variables in the case of Negishi's (1999) model (see also Marcus, 1999b, for the same criticism); by their own account, their model thus does not provide an alternative to processing relations among variables either. We will thus not explore this model further.

---

[7] We hypothesized that the advantage for ordinal grammars may be due to the existence of a lowest tone. For example, if the first tone of a triplet is the lowest possible tone, it can occur in both ABA and ABB, but only in LHM and not in MHL (because, in the grammar MHL, the last tone needs to be lower than the first one, which is not possible if the first tone is already the lowest tone). In order to investigate this possibility, we arranged the tones on a ring (which eliminates "boundaries" of the pitch range), that is, the highest tone of the scale was considered as lower than the lowest tones. In order to keep the topological relations "higher than" and "lower than", we used a range of 20 possible tones (as opposed to 10 possible tones in the previous simulations) and a maximal interval between tones in a triplet of 10 tones. After this manipulation, all pitch levels in all positions appear an equal number of times in either of the repetition-based or the ordinal grammars; hence, the mutual information between pitch levels and grammars is 0, independently of whether the pitch levels are considered individually or as values of the three positional variables.

[8] It is interesting to note that these results imply that the performance of an Ideal Observer as defined in Signal Detection Theory (that is, an observer that chooses the most probable response, given the stimuli) should be better for the ordinal grammars than for the repetitions based grammars; in fact, as our experiments contain two possible choices, higher mutual information between the tones and the grammars implies a higher performance of an Ideal Observer (see Thomson & Kristan, 2005, equations 4.19 and 4.20).

Altmann (2002) hypothesized that the generalizations in Marcus et al.'s (1999) Experiment 1 were due to what the infants had learned *before* participating in the experiments. He thus pre-trained his model on a corpus of simplified sentences before simulating Marcus et al.'s (1999) Experiment 1, using a modified Simple Recurrent Network (Elman, 1990). The network's task was to predict the next word in its input. The model was pre-trained on a subset of a corpus of simplified sentences such as "boy chase cat" (Elman, 1990), and then on the familiarization sentences from Marcus et al.'s (1999) Experiment 1. In half of the simulations, the network was trained on sentences conforming to the grammar ABA; in the other simulations, it was trained on the grammar ABB. Finally, Altmann (2002) tested the model on the test items from Marcus et al.'s (1999) Experiment 1; during test, the network was trained on each test item for 20 iterations, and then the prediction for the last word of the test item was recorded. Altmann (2002) reported that the predictions for consistent test items were better than the predictions for inconsistent test items; hence, the network appeared to learn the grammars ABA and ABB.

While the model may well learn the distinction between the grammars ABA and ABB, it has difficulties accounting for the distinction between AAB and ABB (that was used in Marcus et al.'s (1999) crucial Experiments 2 and 3). In Appendix A, we train the model on AAB and ABB in the same way as Altmann (2002), and then test it on AAB, ABB, but also on *AAA*. In contrast to human participants (see Appendix A for the experiment), the network, after being trained on either AAB or ABB, shows a strong preference for AAA compared to the grammar it has been trained on; hence, while Altmann's (2002) model clearly learns something, it seems to learn something different from what real participants learn, and thus can account neither for Marcus et al.'s (1999) data nor for our own.

Can a model such as Altmann's (2002) account for the advantage for repetition-based grammars with respect to ordinal grammars? A first problem is that sequence-learning models such as Altmann's (2002) cannot learn two grammars in the same experiment, and thus seem unable to simulate our experiments. Indeed, since grammars (e.g., ABA and ABB during Experiment 1) were presented equally often throughout the experiment, there is no way to predict the third tone of a triplet from the first two tones: it can be the same as the first one if the grammar is ABA or the same as the second one if the grammar is ABB, and one can only know the grammar *after* having encountered the third tone. (A similar argument applies to the ordinal grammars.) Still, one could simulate the experiments trial-wise. Recall that, in each trial of our experiments, participants first listened to four context triplets, and then had to judge whether a fifth triplet had the same grammar as the previous four triplets or not. Similarly, the model could be trained in each trial on the context triplets, and be tested on the test triplets. In this way, the task of the network would also mirror more closely the task our participants faced.

We ran the simulations with a version of Altmann's (2002) model where the number of input and output units was adapted to the number of pitch levels in our experiments (see Appendix B for details). We sampled the model's parameter space using 990 parameter sets; with each parameter set, we ran a simulated experiment with 14 participants. We ran the network in one condition/experiment where it had to learn the repetition-based grammars (like in Experiment 1), and in another condition/experiment where it had to learn the ordinal grammars (like in Experiment 2). Then we compared the performance in these two conditions/experiments. (For ease of exposition, we will consider the simulations of Experiments 1 and 2 as two conditions of the same experiment; in reality, however, the simulations were run independently.) As for our experiments, we also compared separately the performance for trials with the context grammars ABA and ABB to the performance on the ordinal grammars, respectively. The results are shown in Fig. 8. This figure shows the percentage of simulated experiments showing an advantage for the ordinal grammars compared to trials with the context grammar ABA (left two columns) and compared to trials with the context grammar ABB (right two columns). The first and the third column from the left show the percentages with respect to all experiments independently of

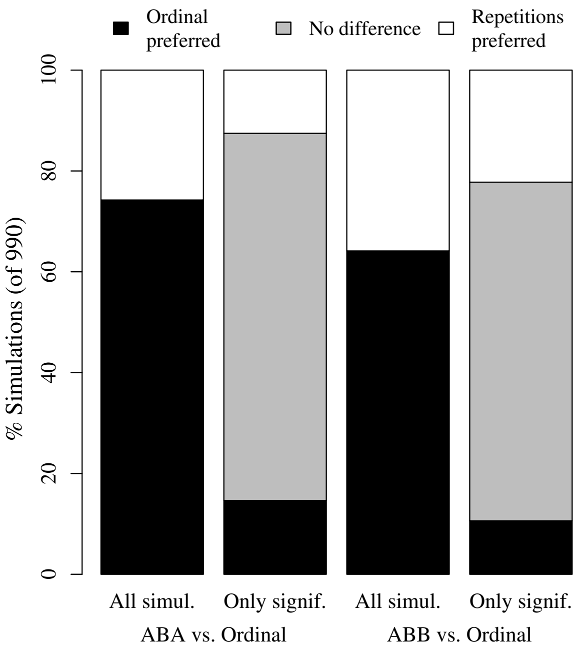

Fig. 8. Results of the trial-wise simulations with a version of a Simple Recurrent Network. The bars show the percentage of simulated experiments performing better for the repetition-based grammars (white shading), the ordinal grammars (black shading), or without such a preference (gray shading). The first and the third bar from the left show these results for all simulated experiments, while the se cond and the fourth bar from the left count only those simulated experiments where the difference between the ordinal and the repetition-based grammars reached significance. The two bars on the left compare trials with the context grammar ABA to the performance on the ordinal grammars; the two bars on the right compare trials with the context grammar ABB to the performance on the ordinal grammars. The network performed better for the ordinal grammars than for the grammar ABA, both when all simulations were considered irrespectively of whether this difference was significant, and when only significant differences between the grammar types were considered. In contrast, when comparing the grammar ABB to the ordinal grammars, the network performed better for the grammar ABB, at least when only significant differences between the grammar types were considered.

whether the advantage reached significance or not; the second and the fourth columns show the percentages of experiments where the advantage reached significance.

When comparing trials with the context grammar ABA to trials with the ordinal grammars, 145 experiments showed a significant advantage for the ordinal grammars, and 124 for the repetition-based grammars. Considering all simulations (irrespectively of significance), the network performed better for the ordinal grammars in 735 out of 990 experiments (74.2%). The pattern was different when comparing trials with the context grammar ABB to trials with the ordinal grammars. In this case, 220 experiments showed a significant advantage for the repetition-based grammars and 105 for the ordinal grammars; when taking into account all experiments (irrespectively of significance), however, 635 out of 990 experiments (64.1%) showed an advantage for the *ordinal* grammars.

Considering only significant differences, the model thus predicts that repetition-based grammars and ordinal grammars should be processed roughly equally well, and it predicts an asymmetry between the grammars ABA and ABB that is not found in our data: The ordinal grammars should be easier to process than ABA, but ABB should be easier to process than the ordinal grammars. In our experiments, in contrast, both repetition-based grammars were processed better than the ordinal grammars. When taking into account all experiments, the model predicts that the ordinal grammars should be processed better. Hence, also this model cannot account for our data.

Obviously, we cannot conclude that no statistical model could account for our data. However, the fact that no viable statistical model generalizing repetition-based grammars has been developed since Marcus et al.'s (1999) publication suggests that it will prove a formidable challenge to develop a statistical model that (i) generalizes repetition-based grammars and (ii) acquires repetition-based grammars better than ordinal grammars. Moreover, such a model would have to account for the results reported by Endress, Scholl, and Mehler (2005), who observed a dissociation between the ability to process particular items and the ability to generalize their grammars (see General Discussion for more details); such a dissociation seems to be problematic for generalizations by statistical models. We leave it therefore as a challenge to the modeler community to develop a simple but *psychologically realistic* statistical model without innate constraints that accounts for all aspects of our data; until such a model has been developed, we feel entitled to conclude that repetition-based grammars may be generalized by a specialized but symbolic operation.

8. General discussion

While cognition is often characterized in terms of computational symbolic or statistical general-purpose machinery (e.g., Anderson, 1993; Elman et al., 1996; Marcus, 2001; Marcus et al., 1999; McClelland et al., 1986; Newell, 1980; Rumelhart et al., 1986; Seidenberg, 1997), other authors proposed that it may rely on many specialized operations (e.g., Gallistel, 1990, 2000). This question is particularly important, because it constrains the forms that theories of learning can take. In this paper, we asked whether the acquisition of some simple grammars can be described more readily by general-purpose mechanisms, or whether they are better accounted for by more specialized operations. Using repetition-based grammars as a case study, participants had to learn either repetition-based grammars, or *ordinal* grammars not entailing repetitions. They readily learned the repetition-based grammars but performed poorly with the ordinal grammars. Control conditions showed that the difficulties with ordinal grammars cannot be entirely attributed to global attentional or psychophysical problems (in fact, the intervals we used were several orders of magnitude above the discrimination threshold, see e.g. Sinnott & Aslin, 1985), nor to similarity-based computations. We then analyzed the predictions of symbolic and statistical general-purpose mechanisms, showing that they cannot explain the advantage for repetition-based grammars either. We conclude that human adults may be endowed with a specialized symbolic mechanism that detects repetitions *independently* of their constituent tones; we speculate that this repetition-detecting mechanism is one of several types of elementary operators that we call *perceptual primitives*.

8.1. A symbolic repetition-detector?

Our explorations of possible statistical models of the generalizations suggest that such mechanisms cannot account for the advantage for repetition-based grammars compared to ordinal grammars. The conclusion that repetition-based grammars are generalized by a specialized operations is also supported by other experiments. Indeed, Endress et al. (2005) showed that participants learn repetition-based grammars much better when the repetitions are located in salient positions than when they are located in less salient positions; they readily learn the grammar ABCDE*FF* (instantiated by syllable sequences like /zØfesapitu<u>koko</u>/) but perform poorly with the grammar ABC*DD*EF (carried by syllable sequences like /zØfesa<u>pipi</u>tuko/). Since control conditions showed that participants can process sequence-medial syllables perfectly well when they are not required to generalize grammars, it seems reasonable to conclude that the difficulties for learning grammars with sequence-medial repetitions were specific to the processes generalizing grammars from a small sample of exemplars.

A statistical mechanism, however, cannot exhibit this dissociation between the ability to process particular items and the ability to learn their structure. For such a mechanism, learning a grammar just amounts to processing the items instantiating it – precisely because the mechanism has no distinct representation of the grammar. Hence, it should learn the grammar when it can process the corresponding items and vice versa. In contrast, if the grammars are represented *independently* of the items that instantiate them, constraints can apply also to the grammars themselves, independently of the perceptual constraints that apply to the items that *instantiate* the grammars; this is simply because the representations of the grammars can be constrained independently only if they exist in the first place. Such a view would thus explain the results mentioned above. It is therefore unlikely that a statistical mechanism could explain why the participants in the aforementioned experiments could process particular items but not their underlying structure.

Of course, the fact that no current model can account for our result does not imply that one could not construct a statistical model without "innate" constraints that does, and we leave it as a challenge to the modeler community to devise a simple and *psychologically plausible* statistical model without innate constraints that explains our data; until such a model is developed, however, we feel entitled to conclude that repetitions are generalized by specialized but symbolic perceptual primitive.[9]

8.2. A specialized repetition-detector?

The above considerations suggest that repetition-based grammars may be learned by a specialized symbolic operation. This conclusion, however, may be challenged by the possibility that participants may not perceive all intervals equally well. As mentioned above, this (quite plausible) possibility is actually compatible with our conclusion, but it cannot explain our data. In fact, it does not follow in any way from the formal structure of the grammars. In a computer, "equal-to" relations (that define the repetition-based grammars) are implemented in exactly the same way as "less-than" or "greater-than" relations (that define the ordinal grammars); for example, the conditional expressions "$x = y$" and "$x < y$" are usually tested by checking the sign of the difference "$x - y$." Likewise, statistical general-purpose mechanisms cannot even have a notion of intervals, because intervals are by definition *relations* between tones; for example, a fifth can occur between many different tones, but the interval is the same irrespectively of its constituent tones.

Moreover, the possibility that participants may not process all intervals equally well is unlikely to explain our data. Let us assume first that the interval between the first and the last tone is not computed. If participants have problems determining intervals exactly, they should learn neither the ordinal grammars nor the *repetition-based* grammar ABA. However, informal debriefing suggested that participants in Experiment 1 were well aware of *both* repetition-based grammars, and participants were well above chance in trials where the context triplets conformed to the grammar ABA; hence, problems for processing intervals between adjacent tones cannot explain the advantage for repetition-based grammars. Let us assume now that also the interval between the first and the third tone is computed. If so, participants should be able to process this interval in the ordinal grammars, as they just have to decide whether it is rising or falling. If the triplets were presented without the second tone, participants would be able to generalize the structure of the remaining tones, but the presence of the irrelevant middle tone appears to prevent participants from drawing these generalizations. In contrast,

[9] Our conclusion is also compatible with previous proposals in the context of Artificial Grammar Learning. Some authors proposed that transfer of a finite state grammar between two consonant sets might be mediated by "Abstract Analogy" (e.g., Brooks & Vokey, 1991; Vokey & Brooks, 1994). Importantly, however, these authors used their concept of abstract analogy to criticize the proposal that participants acquire a *global* knowledge of the entire grammar (rather than the more restricted knowledge afforded by abstract analogy; see e.g. Reber, 1969, 1989). There is thus no contradiction between our results and theirs; rather, both make a similar point, and, indeed, the perceptual primitive generalizing repetitions may be one of the psychological mechanisms that lead to transfer by abstract analogy.

the middle tone does not appear to prevent participants from learning the grammar ABA. Moreover, there is no reason to assume that a general-purpose mechanism should be unable to "ignore" irrelevant information. Hence, if identity relations are easier to process than other intervals, the most likely reason is that repetitions are processed by some specialized but symbolic operator.[10]

The performance for ordinal grammars may possibly improve if the triplets were created using another set of tones and intervals; as mentioned above, we chose our stimuli precisely because of this possibility, as these stimuli allow to uncover the advantage for repetition-based grammars without being hindered due to ceiling effects. Under these conditions, identity relations are readily extracted, while other relations are not, suggesting that participants may be equipped with a specialized mechanism processing repetitions.

8.3. The psychology of symbol manipulation

Our results suggest that humans are equipped with a specialized symbolic operation processing repetitions. It may be more precise to call this operation an *identity-detector* rather than a repetition-detector, as it may plausibly detect identity relations also at some distance and not only repetitions (though we cannot distinguish these possibilities from our data). We used the term "repetition-detector" nevertheless because it lacks the formal connotations of identity-relations. Likewise, such a repetition-detector does not need to be a binary operation, which "responds" if and only if two items are exactly identical; it may well show a graded response depending on how "similar" two items are.

This operation does not seem to require high-level processing, as we recorded neurophysiological responses to a grammar change at latencies that have often been observed with processes that may be relatively automatic (e.g., Näätänen et al., 2001). These results are compatible with the view that the human mind may be endowed with some specialized operations; they are also in line with research on animal learning, where specialized operations fulfill the computational needs an organism faces in its environment (e.g., Gallistel, 1990, 2000). We thus propose that the mind may be equipped with a "computational toolbox" of such specialized *perceptual primitives*, which may solve some of the computational problems it faces.

We chose the term "perceptual primitives" to distinguish these operations from more general primitives such as those in a computer, and to highlight their intrinsic constraints. We certainly do not claim that all mental operations are "perceptual"; rather, we suggest that Gestalt-like specialized and constrained operations such as the

[10] Still another criticism of our conclusions may be that triplets conforming to the repetition-based grammars contain only two different items, while triplets conforming to the ordinal grammars contain three different items. However, in order to notice that triplets conforming to repetition-based grammars contain only two different items, participants have to learn the grammars in the first place. Indeed, all grammars are instantiated by triplets, and thus by three items; if participants have learned that there are actually only two different items in the repetition-based grammars, they have learned these grammars – because this is precisely what defines them. Hence, if it is easier to learn a grammar with identical elements than other grammars, something must make the former grammars easier to learn. We propose that this is because participants may be equipped with a specialized mechanism processing repetitions.

repetition-detector we studied may be found in different modules and at different levels of abstraction. In fact, the examples reviewed below suggest that specialized computations can be observed also in non-perceptual domains. Our point is thus not to stress the "perceptualness" of such operations – but rather their specialization and limitations, and that they seem to be recruited rather effortlessly.

In fact, specific processing constraints are well documented in other domains for both humans and other animals. For example, human adults (at least non-musicians) and infants discriminate tonal but not atonal melodies sharing their contour (e.g., Dowling & Fujitani, 1971; Trehub et al., 1999). Other processing constraints have been observed in the visual domain. For example, it is easy to detect that the lines of a contour are symmetrical, that is, that one line can be transformed into the other by mirroring and translating it; in contrast, it is hard to detect that the lines of a contour are "repeated", that is, that one line can be transformed into the other just by translation *without* mirroring it (e.g., Baylis & Driver, 1994, 1995; Bruce & Morgan, 1975; Corballis & Roldan, 1974), even though symmetry detection is formally more complex (since it entails mirroring in addition to translation). Likewise, one can easily see a 3D object in, say, a Necker cube but not in other, very similar figures such as Kopfermann figures (e.g., Hoffman, 1998). Processing constraints have also been observed in less "perceptual" domains like numerosity. Infants and monkeys discriminate the numerosity 1 from 2 and 2 from 3 but not 2 from 4 or even 3 from 6, although the Weber ratios in these latter examples are the same as or higher than in the former examples (e.g., Feigenson, Carey, & Hauser, 2002; Feigenson, Dehaene, & Spelke, 2004; Hauser & Carey, 2003), and also human adults seem to use different mechanisms for processing numerosities below and above 3 or 4 (e.g., Dehaene, Dehaene-Lambertz, & Cohen, 1998; Trick & Pylyshyn, 1994). It seems therefore that computational constraints are not limited to the distinction between repetition-based and ordinal grammars; rather, many mental computations seem to be specialized and therefore applicable only in a limited domain.

Clearly, there is a huge gap between general capacities like studying high-energy physics and the limited operations that we describe. Nevertheless, some linguistic processes may in fact rely on such simple operations. For example, most languages place grammatical morphemes in salient positions (i.e., the word-edges), leading to affixation like in "walk-*ed*", and edges of constituents have to be aligned (McCarthy & Prince, 1993); likewise, in Semitic languages, the sequence of consonants in a word (irrespectively of the intervening vowels) cannot have repeated consonants (at least in word-initial positions; e.g. Frisch, Pierrehumbert, & Broe, 2004; McCarthy, 1979). While these simple operations may contrast with the more complex computations that underlie language, it is at least well known that a combination of very simple components can yield a powerful computational system, just like a Turing machine is "built" from very simple components. It is therefore possible that very simple, local components like perceptual primitives can be parts of powerful global computational systems like language and, maybe, horizontal faculties.

One may arguably be pessimistic about the possibility of even *studying* horizontal faculties (e.g., Fodor, 1983, 2000). However, if a psychological theory of symbol manipulation can be constructed at all, viewing symbol manipulation as the result of a "computational toolbox" of specialized and constrained operations rather than as a general all-in-one rule-extraction mechanism may be a first step towards such a theory, because it would allow decomposing the problem into modular and investigable pieces (if it can be investigated at all, see also Fodor, 1983).[11]

8.4. Integrating statistical and symbolic models of cognition?

A computational toolbox of perceptual primitives may also help integrating symbolic and statistical models of the mind. Rather than being in opposition, symbolic and statistical processes may contribute different ingredients to learning. Symbolic operations – such as perceptual primitives – may supply representational constraints upon which statistical learning mechanisms can operate. For example, models of inflectional morphology that allow for statistical learning over symbolic representations (e.g., Albright & Hayes, 2003; Albright, 2002; Bybee, 1995) can account for phenomena that cannot be explained by more "classical" models, where statistical and symbolic processing is seen as mutually exclusive (e.g., Hahn & Nakisa, 2000; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Pinker, 1991; Rumelhart & McClelland, 1986); statistical learning over symbolic representations has also produced powerful models in computational linguistics (e.g., Bod & Scha, 1997). Such a complementary view of symbolic and statistical operations may contribute to integrate associationist and symbolic models of the mind, and the notion of perceptual primitives may be one of the ingredients that may help achieving such a synthesis.

Appendix A. Explorations with Altmann's (2002) Model

A.1. Architecture and pretraining

We used the same network architecture and training procedure as in Altmann (2002). The network was a modified Simple Recurrent Network (Elman, 1990) with an additional "recoding" layer between the input layer and the recurrent layer. The input layer consisted of 47 units (29 units coding for the words from Elman's (1990) corpus, 8 for the training syllables in Marcus et al.'s (1999) Experiment 1, 8 for the

[11] There are still other reasons to speculate that perceptual primitives may play some role in mental computations. Compositionality, for example, appears to be a crucial property of human cognition (e.g., Fodor & Pylyshyn, 1988). It reflects the ability of the output of some computations to be the input of other computations. Note first that detecting a repetition may be a compositional operation, as repetitions can be detected only of other representations, which are the result of other computations in turn. Second, also the output of repetitions can be used as an input to other operations, at least if the latter operation is an association. Monkeys, for example, can be trained to respond to an identity-relation when presented with one cue, and to a non-identity-relation when presented with another cue (e.g., Wallis, Anderson, & Miller, 2001). These results can be described in terms of using the result of the repetition-detecting operation as an input to the association with the visual cue; this may be considered as a simple example of a compositional computation. Of course, it is an open question whether the properties of cognition highlighted by Fodor and Pylyshyn (1988) can be explained by properties of perceptual primitives such as a repetition-detector, but the results above suggest that it is at least worthwhile pursuing this question.

test syllables in this experiment, and 2 for utterance boundaries). The recoding layer and the recurrent layer comprised of 20 and 25 neurons, respectively. The output layer contained again 47 units. The network's task was to predict the next element in its input sequences. We used a learning rate of 0.2 and a momentum of 0.01.

Altmann (2002) used a corpus of 252 sentences, but, in fact, Elman's (1990) grammar generates 1040 sentences, excluding sentences with repetitions such as "boy chase boy" (which were excluded also by Altmann (2002) in a control simulation). We thus generated all sentences from Elman's (1990) corpus, and used 100 random subsets of 252 sentences. As in Altmann's (2002) simulations, we concatenated these corpora 40 times, and trained the network on this overall corpus for six cycles.

A.2. Replication of Altmann's (2002) simulations

We first attempted to replicate Altmann's (2002) results. After the pre-training phase, the network was trained on the 16 familiarization sentences from Marcus et al.'s (1999) Experiment 1; the syllables were encoded in eight input units that had not been used so far. The network was trained three times on the corpus of 16 sentences for 50 cycles. For each pre-training corpus (see below), we ran an "experiment" with 16 "participants" pitting the grammar ABA against the grammar ABB. We call an experiment a set of simulations with the same pre-training corpus; a participant is one simulation in an experiment. Participants differed by the initializations of the networks. In each experiment, eight participants were familiarized with ABA sentences, and eight with ABB sentences.

Finally, as in Marcus et al.'s (1999) Experiment 1, the network was tested on triplets conforming to ABA or ABB. In each "trial", the network was trained on a test item for twenty cycles; during the last cycle, the prediction for the last syllable of the test item was recorded. We used two different test items for each grammar, and implemented them using four different syllables.

As Altmann (2002), we evaluated the predictions for the last syllable of the test items by computing the correlation between the target output and the actual output. These scores were then evaluated with an ANOVA using congruence (consistent/inconsistent) as within-subjects factor and familiarization type (ABA or ABB) as between-subject factor. The model was implemented using SNNS[12], and the results were automatically analyzed by a set of Perl and R scripts.

We implemented Altmann's (2002) simulations in different ways, but we generally had difficulties replicating his results. We will thus only report the results for the best simulation parameters, namely when non-orthogonal vectors were used (that is, the strongest activation in each input vector was 0.9 while all other activations were set to 0.1), sentences in the pre-training corpus, in the familiarization sequences and in the test sequences were separated by extra-symbols representing silences, and all activations were reset to 0 after each triplet during familiarization and test. Using these parameters, 28 out of 100 simulated experiments showed a significant advantage for

[12] Version 4.2 is available from http://www-ra.informatik.uni-tuebingen.de/SNNS/. We modified the source to reset the hidden units after utterance boundaries.

consistent test items. While not particularly strong, these results essentially replicate those reported by Altmann (2002).

A.3. Simulations with the grammars AAB and ABB

The simulations were identical to the previous ones, except that the network was trained on the grammars AAB and ABB, and then tested on AAB, ABB, but also AAA items with new "syllables." Instead of using the prediction for the last syllable for evaluating the model's performance, we concatenated the targets and the predictions for all three syllables, and evaluated the correlation between those two vectors. This modification of the evaluation scheme was necessary because the model is generally evaluated using its *predictions*; however, in AAB and ABB, different syllables are predictable, that is, the second one in AAB and the third one in ABB. Hence, it is not possible just to record the predictions for the second or the third syllable; rather, one has to choose an evaluation scheme that is appropriate both for AAB and ABB.

We performed two analyses. In the first, we compared the model's performance on consistent items to the performance on inconsistent items of the form AAB (if the model had been familiarized with ABB) and ABB (if it had been familiarized with AAB). The results were comparable to those with the grammars ABA and ABB: In 22 out 100 simulated experiments, the model performed significantly better for consistent items than for inconsistent items.

The second analysis compared consistent items to foils of the form AAA. In contrast to the previous analysis, where only 22% of the simulated experiments yielded an advantage for consistent items, here the network performed much better for *inconsistent* items than for consistent ones in *all* simulated experiments; even in the least significant experiment, this effect was strong, $F(1,14)=94.8$, $p<0.001$. Hence, the network predicts that, when familiarized with AAB or ABB, participants should prefer consistent items when choosing between AAB and ABB, but they should also show a strong preference for AAA. We will investigate this prediction in the next section.

A.4. Empirical falsification of the predictions

A.4.1. Materials and methods

A.4.1.1. Participants. 10 native speakers of Italian (6 females, 4 males, mean age 24.4, 21-34) took part in this experiment.

A.4.1.2. Familiarization. Participants were presented with the "sentences" from Marcus et al.'s (1999) Experiment 3 synthesized using the fr2 diphone base of Mbrola (Dutoit, Pagel, Pierret, Bataille, & Vreken, 1996). They were instructed to memorize the sentences. Each sentence was played three times in random order; participants could proceed to the next sentence by pressing a key. Half of the participants were familiarized with AAB sentences and half with ABB sentences.

A.4.1.3. Test. After familiarization, participants were informed that the sentences contained some regularity; they were told that they would hear new sequences, and had to judge whether these new sequences conformed to the regularity of the sequences they had memorized during familiarization. Then they were presented with sentences with the structure they had been familiarized with (AAB or ABB), and with sentences of the form AAA; we used the syllables of the test phase of Marcus et al.'s (1999) Experiment 3. This yielded four sentences for each structure, each presented twice in random order.

A.4.1.4. Results and discussion. We recorded the proportion of endorsements of legal items (AAB or ABB, depending on the familiarization) and of foils (that is, AAA sentences). Participants endorsed legal items ($M = 58.8\%$, $SD = 28.9\%$) more often than AAA items ($M = 27.5\%$, $SD = 36.2\%$). An ANOVA with familiarization grammar (AAB or ABB) as between-subject factor and grammaticality (legal items or AAA) as within-subject factor yielded significant main effect of grammaticality, $F(1,8) = 6.1$, $p = 0.039$, but not of familiarization grammar, $F(1,8) = 3.4$, $p = 0.102$, ns, nor an interaction between these factors, $F(1,8) = 4.3$, $p = 0.072$, ns. Hence, in contrast to the predictions of Altmann's (2002) network, participants endorsed legal items more than foils of the form AAA.

Appendix B. Trial-wise simulations with the network from Altmann (2002)

The architecture was the same as in Altmann (2002), except that it was adapted to our stimuli. The input and output layers that contained 11 neurons each (10 for the pitch levels, one for the silence); the recoding layer and the recurrent layer comprised of 10 neurons each. We sampled the parameter space of the model, using 990 combinations of learning rate, momentum and number of cycles: Learning rate: $\{1, 5, 9\} \times \{0.00001, 0.0001, 0.001, 0.01, 0.1\}$; Momentum: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1; Number of training cycles: 10, 50, 90, 100, 500, 900. For each parameter set, we ran 14 simulations with different initializations, representing 14 participants; each set of 14 simulations will be called an "experiment."

Each "participant" in an experiment was run in two independent conditions, corresponding to Experiments 1 and 2, respectively. (We consider that each participant took part in two conditions because we used the same random seed.) The trial order was randomized between participants. In each trial, the network was trained on the four context triplets of the trial (that is, it had to predict the next element in the sequence) for a given number of training cycles; all activations were reset after each triplet. It was then trained with the test triplet for 20 iterations; we then recorded the biggest activation for the last tone, and scored this activation as 1 if the prediction was consistent with the grammar of the test triplet, and as 0 otherwise. For each participant, we averaged this score over trials separately for congruent and deviant trials. For each experiment, these averages were compared using a paired t-test; the experiments were compared using a repeated-measure ANOVA using congruence (congruent vs. deviant) and grammar type (repetition-based vs. ordinal) as within subject factors; if the network performed better for the repetition-based grammars than for the ordinal grammars, we should observe an interaction between congruence and grammar type, and the t-value of the comparison between congruent and deviant trials should be greater for the repetition-based grammars than for the ordinal grammars.

References

Albright, A. (2002). Islands of reliability for regular morphology: evidence from Italian. *Language, 78*, 684–709.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition, 90*(2), 119–161.

Altmann, G. T. (2002). Learning and development in neural networks – the importance of prior experience. *Cognition, 85*(2), B43–B50.

Altmann, G. T., & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science, 284*, 875a.

Altmann, G. T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology – Learning Memory and Cognition, 21*(4), 899–912.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Information and Control, 45*(2), 117–135.

Aslin, R. N., Saffran, J., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Bates, E., & Elman, J. L. (1996). Learning rediscovered. *Science, 274*(5294), 1849–1850.

Baylis, G. C., & Driver, J. (1994). Parallel computation of symmetry but not repetition in single visual objects. *Visual Cognition, 1*, 337–400.

Baylis, G. C., & Driver, J. (1995). Obligatory edge assignment in vision: the role of figure and part segmentation in symmetry detection. *Journal of Experimental Psychology – Human Perception and Performance, 21*(6), 1323–1343.

Bod, R., & Scha, R. (1997). Data-oriented language processing. In S. Young & G. Bloothooft (Eds.), *Corpus-based methods in language and speech processing* (pp. 137–173). Boston, MA: Kluwer.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(8).

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychological – General, 120*(3), 316–323.

Bruce, V. G., & Morgan, M. J. (1975). Violations of symmetry and repetition in visual patterns. *Perception, 4*(3), 239–249.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425–455.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103*(3), 566–581.

Chater, N. (1999). The search for simplicity: a fundamental cognitive principle? *Quarterly Journal of Experimental Psychology Section-A, 52*(2), 273–302.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*(1), 19–22.

Christiansen, M., Conway, C., & Curtin, S. (2000). A connectionist single-mechanism account of rule-like behavior in infancy. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 83–88). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M., & Curtin, S. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends in Cognitive Sciences, 3*(8), 289–290.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology – General, 120*(3), 235–253.

Collett, T., & Collett, M. (2000). Path integration in insects. *Current Opinion in Neurobiology, 10*(6), 757–762.

Corballis, M. C., & Roldan, C. E. (1974). On the perception of symmetrical and repeated patterns. *Perception & Psychophysics, 16*(1), 136–142.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology – Learning Memogy and Cognition, 30*(5), 1119–1130.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience, 21*(8), 355–361.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology – Learning Memory and Cognition, 17*(5), 875–887.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of Acoustical Society of America, 49*(2), 524–531.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Vreken, O. van der. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (Vol. 3, pp. 1393–1396). Philadelphia.

Dyer, F., & Dickinson, J. (1994). Development of sun compensation by honeybees: how partially experienced bees estimate the sun's course. *Proceedings of the National Academy of Sciences of the United States of America, 91*(10), 4471–4474.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211.

Elman, J. L., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology – General, 134*(3), 406–419.

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in neural information processing systems* (Vol. 2, pp. 524–532). Denver 1989: Morgan Kaufman.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: object files versus analog magnitudes. *Psychological Science, 13*(2), 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822–15826.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition, 28*(1–2), 3–71.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *National Language & Linguistic Theory, 22*(1), 179–228.

Gallistel, C. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gallistel, C. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 1179–1191). Cambridge, MA: MIT Press.

Gallistel, C. (2001). Mental representations, psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 9691–9695). Amsterdam, The Netherlands: Elsevier.

Gallistel, C., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107*(2), 289–344.

Gallistel, C., & Gibbon, J. (2002). *The symbolic foundations of conditioned behavior*. Mahwah, NJ: Lawrence Erlbaum.

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science, 308*(5724), 1046–1049.

Gasser, M., & Colunga, E. (2000). Babies, variables, and relational correlations. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 160–165).

Giard, M., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., et al. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *Journal of Cognition Neuroscience, 7*(2), 133–143.

Gigerenzer, G., & Todd, P., The ABC Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature, 410*(6831), 930–933.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10*, 447–474.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109–135.

Gómez, R. L., Gerken, L., & Schvaneveldt, R. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition, 28*(2), 253–263.

Goodman, N. (1955). *Fact, fiction and forecast*. Cambridge: Harvard University Press.

Gould, J. L., & Marler, P. (1987). Learning by instinct. *Scientific American, 256*(1), 74–85.

Hahn, U., & Nakisa, R. (2000). German inflection: single route or dual route? *Cognitive Psychology, 41*(4), 313–360.

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: examinations of content and format. *Cognitive Psychology, 47*(4), 367–401.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53–B64.

Hoffman, D. D. (1998). *Visual intelligence: How we create what we see*. New York, NY: W.W. Norton & Co.

Horváth, J., Czigler, I., Sussman, E., & Winkler, I. (2001). Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Research Cognitive Brain Research, 12*(1), 131–144.

Hume, D. (1739/2003). A treatise of human nature. *Project Gutenberg*.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: testing the predictions of two connectionist models. *Psychological Review, 63*(2), 95–105.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: no evidence for the acquisition of rules. *Memory & Cognition, 28*(8), 1321–1332.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology – Learning Memory and Cognition, 22*(1), 169–181.

Korzyukov, O. A., Winkler, I., Gumenyuk, V. I., & Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *Neuroimage, 20*(4), 2245–2258.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616–618.

Marcus, G. F. (1999a). Connectionism: with or without rules? response to J.L. McClelland and D.C. Plaut (1999). *Trends in Cognitive Sciences, 3*(5), 168–170.

Marcus, G. F. (1999b). Do infants learn grammar with algebra or statistics? Reply to Seidenberg and Elman, Eimas and Negishi. *Science, 284*, 436–437.

Marcus, G. F. (1999c). Reply to Christiansen and Curtin. *Trends in Cognitive Sciences, 3*(8), 290–291.

Marcus, G. F. (1999d). Reply to Seidenberg and Elman. *Trends in Cognitive Sciences, 3*(8), 289.

Marcus, G. F. (1999e). Rule learning by seven-month-old infants and neural networks. Response to Altmann and Dienes. *Science, 284*, 875a.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: the exception that proves the rule. *Cognitive Psychology, 29*(3), 189–256.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.

Marler, P. (1997). Three models of song learning: evidence from behavior. *Journal of Neurobiology, 33*(5), 501–516.

McCarthy, J.J. (1979). Formal problems in Semitic phonology and morphology. Doctoral dissertation, MIT, Cambridge, MA. (Distributed by Indiana University Linguistics Club, Bloomington, IN. Published by Garland Press, New York, 1985).

McCarthy, J. J., & Prince, A. (1993). Generalized alignment. In G. Booij & J. van Marie (Eds.), *Yearbook of morphology 1993* (pp. 79–153). Boston, MA: Kluwer.

McClelland, J. L., & Plaut, D. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences, 3*(5), 166–168.

McClelland, J. L., Rumelhart, D. E., & The PDP Research Group (Eds.). (1986). Parallel distributed processing (Vol. 2: Psychological and Biological Models). Cambridge, MA: MIT Press.

Meulemans, T., & van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology – Learning Memory and Cognition, 23*(4), 1007–1028.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Peralta, R. G. de. (2004). EEG source imaging. *Clinical Neurophysiology, 115*(10), 2195–2222.

Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Näätänen, R., & Alho, K. (1995). Mismatch negativity – a unique measure of sensory processing in audition. *International Journal of Neuroscience, 80*(1–4), 317–337.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neuroscience, 24*(5), 283–288.

Negishi, M. (1999). Do infants learn grammar with algebra or statistics? *Science, 284*(5413), 435.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135–183.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.

Osherson, D., Stob, M., & Weinstein, S. (1984). Learning theory and natural language. *Cognition, 17*(1), 1–28.

Paavilainen, P., Jaramillo, M., Näätänen, R., & Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters, 265*(3), 179–182.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: an expandable programming language for online psychological experiments. *Behavior Research Methods Instruments & Computers, 29*, 322–327.

Paz-Y-Miño C, G., Bond, A. B., Kamil, A. C., & Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature, 430*(7001), 778–781.

Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530–535.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*(3), 303–343.

Ramachandran, V. (1990). Interactions between motion, depth, color and form: The utilitarian theory of perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 346–360). New York: Cambridge University Press.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*(1), 115–119.

Reber, A. S. (1989). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology – General, 118*(3), 219–235.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 2, Psychological and Biological Models pp. 216–71). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & The PDP Research Group (Eds.). (1986). Parallel distributed processing (Vol. 1: Foundations). Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Schneidman, E., Bialek, W., & Berry, M. J. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience, 23*(37), 11539–11553.

Seidenberg, M. S. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science, 275*(5306), 1599–1603.

Seidenberg, M. S., & Elman, J. (1999a). Do infants learn grammar with algebra or statistics? *Science, 284*(5413), 433.

Seidenberg, M. S., & Elman, J. (1999b). Networks are not 'hidden rules'. *Trends in Cognitive Sciences, 3*(8), 288–289.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391–1400.

Shtyrov, Y., & Pulvermüller, F. (2002a). Memory traces for inflectional affixes as shown by mismatch negativity. *European Journal of Neuroscience, 15*(6), 1085–1091.

Shtyrov, Y., & Pulvermuller, F. (2002b). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport, 13*(4), 521–525.

Shultz, T. R., & Bale, A. C. (2001). Neural network simulation of infant familiarization to artificial sentences: rule-like behavior without explicit rules and variables. *Infancy, 2*(4), 501–536.

Sinnott, J., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78*(6), 1986–1992.

Stabler, E. P. (1998). Acquiring languages with movement. *Syntax, 1*(1), 72.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters, 266*(2), 109–112.

Thomson, E. E., & Kristan, W. B. (2005). Quantifying stimulus discriminability: a comparison of information theory and ideal observer analysis. *Neural Computation, 17*(4), 741–778.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*(6501), 90–92.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology – Learning Memory and Cognition, 30*(5), 1131–1142.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*(5), 867–875.

Trehub, S., Schellenberg, E., & Kamenetsky, S. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology – Human Perception and Performance, 25*(4), 965–975.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport, 9*(16), 3557–3560.

Trick, L., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80–102.

Tunney, R., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology – Learning Memory and Cognition, 27*(3), 614–639.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology – General, 134*(4), 552–564.

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology – Learning Memory and Cognition, 20*(6), 1504–1510.

Wallis, J., Anderson, K., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411*(6840), 953–956.

Wexler, K., & Cullicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.