Journal of Experimental Psychology: General
2006, Vol. 135, No. 2, 314–321

0096-3445/06/$12.00 DOI: 10.1037/0096-3445.135.2.314

COMMENTS

How to Hit Scylla Without Avoiding Charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004)

Luca L. Bonatti
International School for Advanced Studies and
University of Paris VIII at Saint Denis

Marcela Peña
International School for Advanced Studies and
Pontificia Universidad Católica de Chile

Marina Nespor
University of Ferrara

Jacques Mehler
International School for Advanced Studies and
Centre National de la Recherche Scientifique

M. Peña, L. L. Bonatti, M. Nespor, and J. Mehler (2002) argued that humans compute nonadjacent statistical relations among syllables in a continuous artificial speech stream to extract words, but they use other computations to determine the structural properties of words. Instead, when participants are familiarized with a segmented stream, structural generalizations about words are quickly established. P. Perruchet, M. D. Tyler, N. Galland, and R. Peereman (2004) criticized M. Peña et al.'s work and dismissed their results. In this article, the authors show that P. Perruchet et al.'s criticisms are groundless.

Keywords: language, statistical learning, rule learning, parsing

Infants and adults can use statistical properties of unsegmented speech to identify lexical candidates. They can segment a continuous artificial speech stream when word boundaries are marked by transition probabilities (TPs) between adjacent syllables—that is, the conditional probability of encountering one syllable after another (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, & Tunick, 1997; Newport & Aslin, 2004); however, a language is more than a lexicon. Sensitivity to structural relations is fundamental. For instance, in the sentence "The children use the sand in the garden to play," the final verb does not refer to its adjacent constituent but rather to the initial one. Also, in morphology, comparable phenomena occur. For example, although the verb *imbiancare* (paint) exists, neither the adjective *imbianco* nor the verb *biancare* do. The verb is created by simultaneously adding a prefix and a suffix to the adjectival morphological root *bianco*, a process that generalizes to many verbs. Both examples illustrate that nonadjacent dependencies play an important role in natural language.

Thus, tracking adjacent relations among constituents is not sufficient for language acquisition. Some researchers have argued that if humans could track higher order statistics such as nonadjacent TPs, even learning at least some syntax might be accomplished through the same statistical mechanism extracting words from continuous speech (Altmann, 2002; Bates & Elman, 1996; Elman, 1999, 2001; Elman et al., 1996). Other researchers have argued instead that human statistical abilities are limited to computations between adjacent items (e.g., Newport & Aslin, 2004).

In Peña, Bonatti, Nespor, and Mehler (2002), we studied how learners extract structural properties from speech by comparing how learners can segment streams lacking clear statistical information among adjacent syllables and how they extract elementary structural generalizations about constituents. Our aims were to assess the power of human statistical abilities, to determine whether such power could attain linguistic generalizations, and to explore whether one single mechanism could account for both speech segmentation and structural generalizations. We proposed a series of conclusions that we summarize below.

Luca L. Bonatti, Sector of Cognitive Neuroscience, International School for Advanced Studies, Trieste, Italy, and Department of Psychology, University of Paris VIII at Saint Denis, Saint Denis, France; Marcela Peña, Sector of Cognitive Neuroscience, International School for Advanced Studies and Laboratorio de Neurosciencias, Escuela de Psicología, Pontificia Universidad Católica de Chile, Macul, Chile; Marina Nespor, Department of Humanities, University of Ferrara, Ferrara, Italy; Jacques Mehler, Sector of Cognitive Neuroscience, International School for Advanced Studies and Laboratoire de Sciences Cognitives et de Psycholinguistique, Centre National de la Recherche Scientifique, Paris, France.

This research was funded by the European Science Foundation Origin of Man, Language and Languages Grant UE NEST 012778 CALACEI awarded to Marina Nespor, the Human Frontier Science Program Grant RGP 68/2002 awarded to Jacques Mehler, and Friuli-Venezia-Giulia Grant LR 3/9 awarded to Jacques Mehler. We thank Ansgar Endress for discussions and suggestions.

Correspondence concerning this article should be addressed to Luca L. Bonatti, Sector of Cognitive Neuroscience, International School for Advanced Studies, 2-4 Via Beirut, Trieste 34014, Italy. E-mail: lucabonatti@mac.com

Segmentation and Generalization: Peña et al.'s (2002) Claims

Generally, TPs between adjacent syllables are higher within words than across word edges; therefore, a system tracking TPs has good chances of retrieving words from unsegmented speech. Several studies showed that adults and infants can segment words by exploiting TPs between adjacent syllables (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran et al., 1997). Similar computations occur in other modalities (Fiser & Aslin, 2001, 2002; Saffran, Johnson, Aslin, & Newport, 1999).

Since many dependencies in language are not adjacent, if learners were sensitive to nonadjacent relations *between individual tokens,* in principle, the same computations used to segment words from a continuous stream could uncover long-distance relations. Can participants perform such computations? In Peña et al. (2002), we explored this question by assessing whether participants could parse a continuous stream with low TPs among adjacent syllables but with high TPs among nonadjacent syllables. In Experiment 1, the first syllable of each pseudoword predicted the last syllable with certainty, whereas the middle syllables varied. The language contained nine trisyllabic words belonging to three families, with constant syllables at the edges but a varying middle syllable (hence, having the form AXC).

We found that after undergoing a 10-min familiarization, participants successfully segment the stream, as indicated by their preference for words over *part words* (defined as trisyllabic items spanning across two words). We ruled out that this result was due to idiosyncratic features of the material with a control experiment. When we changed the probability relations among the syllables in the familiarization, most words and part-words of Experiment 1 became, respectively, part words and words of the control. If low-level features of the items or previous language experience were responsible for the observed preferences for words in Experiment 1, participants should still prefer those items even if they are part words in the control; however, again they preferred words (see Table 1; Peña, 2002; Peña et al., 2002, footnote 17).

We concluded that "participants appear to take advantage of nonadjacent statistical dependencies between consonant-vowel syllables to automatically segment a continuous stream" (Peña et al., 2002, p. 605). It is important to keep in mind that our conclusion concerned the relation among syllables and not among syllable chunks. That the relation between *A* and *C* in *ABC* is nonadjacent is clear by definition; the question was whether learners can compute such dependencies among syllables in continuous speech, however they do it.

In Peña et al. (2002), we separated the issue of token segmentation out of a continuum from a second issue: the identification of the structure instantiated in such tokens. We showed that even if participants appear to segment a continuum on the basis of second order probabilities, this is insufficient to identify the structural relations characterizing words. In Experiment 2, we familiarized participants with the same 10-min continuous stream of Experiment 1 and evaluated whether they preferred an item that never appeared in the stream but that had AXC structure (a rule word such as PU*BE*KI, where *BE* never occurred in that position during familiarization) to a part word that appeared in the stream but that lacked this structure (such as LIKITA, composed of the two last syllables of one word and the first syllable of another word). Participants had no preference, suggesting that they did not realize that the words they could extract from the stream conformed to a common structural description. However, when familiarized with a stream that was phenomenologically similar to that of Experiment 1 but that contained small gaps (25 ms) at the edges of words, participants switched their preference toward rule words (Peña et al., 2002; Experiments 3 and 5), even if they were consciously unable to detect the gaps (Peña et al., 2002, footnote 22). These results led us to suggest that subtle properties of the stimulus trigger two mechanisms, each with a different purpose: One mechanism is dedicated to breaking continua into their components, and the other mechanism is dedicated to projecting conjectures about the underlying structure of the components.

Table 1
Test Items Used in Peña et al. (2002)

Experiment 1		Control	
Words	Part words	Words	Part words
PULIKI	DUBERA	**RAGAPU**	**PULIKI**
PURAKI	*FOGAPU*	*RADUPU*	**PUFOKI**
PUFOKI	GAPUFO	*RAKIPU*	**BERAGA**
BELIGA	KITAFO	*FOGATA*	**BEFOGA**
BERAGA	KITARA	*FODUTA*	**TALIDU**
BEFOGA	LIDUBE	*FOKITA*	**TARADU**
TALIDU	LIKITA	*LIGABE*	*GAPULI*
TARADU	*RADUBE*	**LIDUBE**	*KIBERA*
TAFODU	*RAGAPU*	**LIKIBE**	*DUPUFO*
Preference for words: 73% ($p < .0005$)		Preference for words: 58% ($p < .02$)	

Note. This table presents the test items used in Peña et al.'s (2002) Experiment 1 and its control in Footnote 17. Because words are organized into families, it is not possible to exactly invert all words and part words. The control uses the closest possible match to a full inversion. Items in bold invert their word/part-word status exactly from the experiment to its control; items in italics change, at most, one syllable in one position from the experiment to its control. Results of the experiments are reported at the bottom of the table.

Another issue concerns the kind of information these systems exploit. Conceivably, both systems may look for some statistical relations (such as TPs) by extensively sampling the input. We showed that this possibility is unlikely: The time courses of segmentation and generalization are different. Exposure for 2 min to the subliminally segmented stream suffices to trigger generalizations (Experiment 5), whereas exposure for 30 min to the continuous stream does not (Experiment 4), and, indeed, induces preference for part words rather than rule words. Thus, increasing exposure to the continuous stream results only in a better recall of the items actually heard in the stream, to the detriment of those that conform to the structure.

These results led us to conclude that different procedures mediate the segmentation of continuous speech and the extraction of the structure of the recovered items. In particular, we suggested that parsing is essentially statistical in nature, whereas structural generalizations may not be. Structure is conjectured on the basis of sparse evidence; it does not operate on a continuum but only on a discrete corpus; and it is easily overwhelmed by memory consolidation.

These statements are, in brief, our major claims. Recently, Perruchet, Tyler, Galland, and Peereman (2004) argued that our experiments and conclusions were "flawed by deep methodological inadequacies" (p. 573), that our data were "invalid" (p. 573) and "unreliable" (p. 579), and that our work contained a "major logical pitfall" (p. 574), and they concluded that "What Peña et al. [2002] showed is that adding segmentation cues helps segmentation, and that is all" (p. 581). In this article, we show that Perruchet et al. failed to grasp our work and that, if anything, the data that they provided support our conclusions.

Peña et al.'s (2002) Logical Mistakes

According to our critics, we made a logical mistake. Supposedly, we claimed that the remote dependency between syllables is the exclusive source of learning about their relation. But as we introduced cues to segment streams in Experiments 3 and 5, we measured "the learning of remote dependencies through its effect on word segmentation" (Peña et al., 2002, p. 574). Thus, we would have shown only that adding segmentation cues helps segmentation.

Notice that remote dependency learning is an ambiguous formulation that covers segmentation and generalization—two processes that we distinguished and studied in detail. We showed that participants use remote dependencies to segment words in a speech stream but fail to generalize to word structure by exploiting these dependencies (see Peña et al., 2002, Experiments 2 and 5). Obviously, if participants segment a continuous stream when A predicts C, a fortiori they will do so when segmentation cues are inserted. But our point was different: Such cues help identifying the constituents of the stream, hence releasing the learners from the need to parse it and directing attention to its structure. Indeed, without gaps, participants failed to capture structural relations.

To repeat, we argued that subliminal segmentation cues help learners generalize structural properties of the items, whereas probability relations among syllable*s* do not promote generalization but, rather, promote segmentation. Perruchet et al. (2004) confused the two, as their summary of our work clearly shows:

> Peña et al. [2002] (1) observed that word segmentation is not possible when AXC words are displayed as a continuous speech stream, (2) introduced an explicit cue for segmentation between the AXC words, (3) observed that this cue indeed helps segmentation, (4) inferred from this observation that it is possible to detect the remote dependencies between A and C when segmentation cues are displayed, and (5) attributed this achievement to the fact that the segmentation cues made the stream more similar to natural language, hence triggering language-specific algebraic-like computations. (p. 581)

In fact, Peña et al.'s (2002) Experiment 1 shows the opposite of what our critics claim: Word segmentation is indeed possible with AXC words in a continuous speech stream. Because our critics fail to distinguish between segmentation and generalization, they make statements that our experiments directly contradict.

In their summary, Perruchet et al. (2004) also confused other issues that we kept separate. They conflated the language-specific nature of structure extraction (for which we did not argue); its algebraic character, which is compatible with but not mandated by our results; and the similarity between natural language and a subliminally segmented artificial speech stream (highlighted by the constituents and not by the presence of pauses). But Perruchet et al.'s main confusion was about the nature of our argument. They argued that our conclusion does not logically follow from our alleged premises. It obviously does not: We never proposed a logical argument to begin with but, rather, presented empirical data and empirical hypotheses that Perruchet et al. failed to keep apart.

Peña et al.'s (2002) Methodological Errors

Perruchet et al. (2004) claimed that our methodology—the same as that used by many researchers in this area (e.g., Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996; Saffran et al., 1997)—was deeply flawed, our result being vitiated by training-independent and training-dependent factors. *Training-independent factors* are the result of perceptual biases that individuals acquire when learning the maternal language. *Training-dependent factors* are "cues other than distant dependencies" contained in the familiarization of our experiments (Perruchet et al, p. 574). Let us examine how these factors would conspire to produce what Perruchet et al. labeled our "invalid data" (pp. 573, 575).

Training-Independent Factors, Segmentation, and Generalization

Every participant who comes to the experimental room has years of experience with his or her maternal language. This truism, which Perruchet et al. (2004) held against our experiments, applies to all psycholinguistic experiments with adults; the real issue is how to control these factors most effectively. For example, in some of our experiments, the trisyllabic words shared the consonantal structure *stop–continuous–stop* (e.g., PULIKI) whereas part words do not (Newport & Aslin, 2004; Seidenberg, MacDonald, & Saffran, 2002). In their Experiment 1, Perruchet et al. investigated whether this pattern, as well as other factors, could have accounted for our results. For example, they considered the base rate distribution of the word's initial syllables in French, the distribution in trisyllabic French words of the consonantal patterns that we used, the parameters of the synthesizer Mbrola, and several other factors. They found that none of these factors could have lead participants to make the choices that we reported. Although the obvious conclusion should be that our experiments were not vitiated by such factors, our critics concluded that the perceptual biases generated by the familiarization "certainly originate from a variety of factors, the respective contributions of which should warrant further studies" (Perruchet et al., p. 579).

In fact, the control described above already controls for the most plausible training-independent factors: By modifying the probability relations among syllables during familiarization, we kept the test items constant but inverted their status as words/part words. Because participants still preferred items according to their role as words or part words, training-independent factors, low-level phonological or phonotactic features, or intrinsic preference for some test items could not account for the pattern of results that we reported. Moreover, in several experiments conducted in our laboratory, we found that testing other familiarization streams with different phonological and phonotactic properties or even testing speakers of other languages (see Peña, 2002, Experiment 6) did not change the results. In all cases, the probability relations contained in the familiarization streams determined participants' preferences. Peña et al.'s (2002) Experiment 1 and its control offer an even richer conclusion: The particular choice of the series of syllables in the familiarization does modulate participants' preferences but does not explain them.[1] This information is available in our article and in Peña (2002).

Yet Perruchet et al. (2004), interested in showing that preference for items of AXC structure in our Experiment 1 was due to uncontrolled training-independent factors, pressed other criticisms. They obliterated distant TPs in the familiarization of their Experiment 1 by increasing the number of words in the stream, allowing all combinations among first and last syllables. Because participants tend to remember trisyllabic items after undergoing familiarization with continuous or discontinuous streams, Perruchet et al. concluded that participants "were sensitive from the start to some prosodic or phonological aspect of the auditory strings" (p. 578). Although this sensitivity may happen with the particular stream that the researchers used, we have shown that these factors cannot explain participants' preferences in our segmentation tasks.

More generally, it is doubtful that the continuous-stream condition of Perruchet et al.'s (2004) Experiment 1 demonstrates the influence of training-independent factors. Although our critics correctly pointed out that the absolute frequency of words in the stream of our Experiment 1 may explain participants' preferences, they failed to point out that the same factor (which we controlled for in a separate experiment but they did not) affected their (vs. our) Experiment 1. Because they obtained their stream by concatenating 27 AXC words, the AXC words had higher absolute frequency than did any of the part words generated by the concatenation; therefore, the finding that participants recalled AXC items after undergoing familiarization with a random concatenation of AXC words is not surprising and may be explained by the absolute frequency of the items in the familiarization stream. This observation in no way belittles Peña et al.'s (2002) results, which show that even when the absolute frequency of words and part words is equated, participants still prefer words even when they cannot rely on statistical information among adjacent syllables.

Training-Dependent Factors and Generalization

The Subliminal Segmentation Cues

The second factor that would invalidate our work—the subliminal segmentation cues (Experiments 3 and 5)—is training dependent. Our critics commented on our results, writing that because segmentation cues segment the stream, it is no surprise that people segment; however, our experiments with subliminal cues were not intended to prove segmentation, but rather were intended to prove generalization, from a speech stream. Because Perruchet et al. (2004) confused segmentation with generalization, they missed the point of our experiments.

Our critics also seemed skeptical about the subliminal status of the 25-ms gaps. In preparing our material, we piloted several different pause durations until we found a level at which participants became unaware of their presence. Of course such cues, although failing to reach awareness, were processed, as the term *subliminal* itself suggests. Indeed, they induce a sense of constituency helping segmentation, just as rhythm and intonation do in natural speech. As we wrote in Peña et al. (2002), "The gaps contained in the stream may help participants to segment" because they provide "explicit bracketing cues" (p. 606). Because this is what we wrote, we are surprised that our critics presented the fact that gaps induce a sense of segmentation as their original discovery.

Our critics replicated our results about the subliminal nature of gaps but, they wrote, "The crucial point is not whether participants detected the pauses, but whether they perceived the auditory string including the pauses as a succession of words" (p. 577). In one condition of their Experiment 1, Perruchet et al. (2004) familiarized participants with a stream containing gaps but deprived of nonadjacent dependencies among syllables and found that participants remembered mainly trisyllabic items. Because a similar result was obtained with a segmented stream containing nonadjacent dependencies, our critics concluded that remote dependencies do not help segmentation because segmentation cues are already helping word identification. For some reason, they interpreted this result as a blow to Peña et al. (2002), but this conclusion is precisely our claim: Silences are undetected, yet they provide a subliminal cue that facilitates the segmentation into words. We did not use the subliminal pauses to establish segmentation: Instead, we showed that participants captured the generalizations that had eluded them when the stream had no segmentation cues. Perruchet et al. ignored this difference.

Generalizations to structure in syntax and morphology always seem to require a corpus with constituent structure. We showed that even in the case of artificial grammar learning, constituent structure (promoted by gaps) is a prerequisite for generalization. Perruchet et al. (2004) criticized our claim that pauses make the stream closer to natural language, but our critics overlooked its context. Natural language does not contain systematic pauses between words; however, prosodic cues, often perceived as pauses, signal constituency (Nespor & Vogel, 1986; Selkirk, 1984). Hence, an artificial stream with segmentation cues does share with natural language the relevant property that we intended to test: Namely, it contains information about the basic units of analysis over which the generalizations need to be drawn. Of course, speech itself does not contain systematic 25-ms pauses between words, but this fact is irrelevant to our argument: We only made the point that without segmentation cues in the signal, generalizations are not projected.

On two other related points, Perruchet et al. (2004) were factually wrong. First, they insisted that the sense of trisyllabicity induced by our streams, per se, explains our results; however, trisyllabicity is, to a large extent, irrelevant. Perruchet et al. overlooked the fact that Peña et al. (2002) showed that participants exposed to a trisyllabic segmented stream prefer a quadrisyllabic item provided that it has the correct A–C structure (Peña et al., 2002, Footnote 29). The second mistake clearly appeared when they wrote:

[1] Participants expressed a 73% preference for words over part words when exposed to the continuous familiarization stream of our Experiment 1, but they switched their preference to the new words (indeed, the previous part words) when they were exposed to the continuous stream of its control with a 58% mean preference. The difference among the two experiments (significant, as we wrote in Pena et al., 2002) allows one to gauge the effects due to the peculiarities of the streams. Differences in familiarizations not related to their probability relations exist but are insufficient to explain participants' preferences.

> If the speech stream was perceived directly as a set of AXC words, it follows that the part words used during the test *were not even encoded* during the familiarization phase. In those conditions, the fact that participants failed to recognize them becomes trivial. (Perruchet et al., 2004, p. 579, italics added)

This may appear trivial to Perruchet et al. (2004) but not to the individuals who participated in our Experiment 4 who, exposed to a 30-min stream, selected part words over rule words. This result shows that part words are indeed encoded during familiarization and are even preferred to rule words when extra exposure strengthens their memory traces. Once again, Perruchet et al. made statements that are contradicted by our experiments.

Pauses and the Extraction of Generalizations

In Peña et al. (2002), we explored whether participants who had been familiarized with AB_1C, AB_2C, and AB_3C words could capture the AXC relation, accepting as legal items such as AB_*C, where B_* never occurred in that position during familiarization. We showed that a short familiarization with a segmented stream induces a preference for novel items with the structure AXC over familiar part words.

Our critics focused on the generalization issue only briefly, claiming that our test for generalization is invalid. According to them, the proper test would compare rule words against what they call scrambled words—that is, items that draw the A and C syllables from different families. Thus, if a word in the stream has the structure A_1BC_1 and another word in the stream has the structure A_2BC_2, a scrambled word would have the structure $A_1B_*C_2$.

After using a 2-min discontinuous stream similar to that used in Peña et al.'s (2002) Experiment 5, in their second experiment our critics tested three kinds of comparisons: rule words vs. part words, scrambled words vs. part words, and scrambled words vs. rule words. The first comparison was similar to the test conditions of our Experiments 3 and 5, although it was less sensitive because of the scant number of comparisons tested (9 pairs per participant vs. 36 pairs per participant in Peña et al.). Nevertheless, they replicated our results that, with little exposure, participants select rule words over part words.[2] However, Perruchet et al. (2004) ignored this and claimed that the only "correct comparison" (p. 580) is between scrambled words and part words. Because participants preferred scrambled words, Perruchet et al. concluded that our data are invalid.

This conclusion does not follow for two reasons. First, Perruchet et al.'s (2004) methodology vitiates the results. Perruchet et al. tested the three comparisons within participants as if these comparisons were independent; however, they are not. A participant who chooses a rule word α over a part word β may then be confronted with a choice between a scrambled word γ and the same part word β. Thus, the individual might select γ, not because he or she prefers γ but because he or she knows that he or she had already rejected β. Therefore, preference for scrambled words over part words might reflect a preference for rule words over part words. With Perruchet et al.'s use of the nonindependent alternative forced-choice design, it is impossible to interpret their results and accept their statistical analyses.

Second, the comparison is indeed interesting, but the conclusions that it suggests are not the ones that Perruchet et al. (2004) advance. With methodologically sound tests, Endress (2005) independently explored this comparison. What our critics call *scrambled words* and what Endress calls *class words* are instances of a lawful relation among syllables occurring in the first and last position. Indeed, participants may learn that the first and last positions are interconnected, creating classes of items that occurred in those positions during familiarization. Preference for class words may indicate that participants captured this generalization. Many phenomena in natural language are reminiscent of class words. For example, roots of words remain constant, but prefixes and suffixes can vary independently when words are derived or inflected. Endress systematically explored sensitivity to class word structure and showed that under certain conditions, participants capture it.

Endress (2005) also showed that even when participants prefer class words to part words, they still prefer rule words to class words. When directly comparing class words with rule words, Perruchet et al. (2004) reached the same conclusion; however, this fact puts our critics in a paradoxical situation. Recall that during familiarization, neither rule words nor class words ever appeared and that their middle syllable never appeared in that position either. Therefore, to prefer rule words to class words, participants must be able to consider the relation between the first *and* last syllables of test items simultaneously, abstracting over their middle position. Showing this ability is tantamount to establishing that participants are sensitive to the abstract structure of distant elements in words by virtue of their ability to generalize. But this was exactly Peña et al.'s (2002) conclusion: Participants who are exposed to a short, subliminally segmented stream can generalize to structural information. Our critics did not realize that their result directly confirms our thesis and refutes their claim that no such abstract dependencies are captured.

Having replicated our results, instead of concluding that Peña et al.'s (2002) findings are not "invalid" after all, our critics argued that our results are, at most, small- to-medium-size effects (Cohen's convention); however, their experiment was not suited for assessment of effect size. In testing three conditions within participants and reducing the number of items as compared with the number used in Peña et al., the experiment is bound to underestimate effect sizes. Thus, Perruchet et al. (2004) reported a preference for rule words over class words with a d of 0.38 (presumably coming from the restricted analysis of one of their three within-participant conditions); however, Endress (2005) tested the same contrast in an experiment in which 20 participants were tested exclusively with 27 rule-word pairs vs. class-word pairs. After 2 min of familiarization with the subliminally segmented stream, participants still preferred rule words (hence confirming the learning of long-distance relationships) but with a large effect size (Cohen's $d = 0.98$).

In summary, our critics argued that our data are invalid, yet they replicated them, including the phenomena whose existence they denied. Thus, Perruchet et al. (2004) began by boldly stating that

[2] Because Perruchet et al. (2004) tested two different familiarization streams (between groups) and found no interaction between familiarization and test results, this experiment also shows that Peña et al's (2002) original experiments were not vitiated by the choice of the specific material used, as Perruchet et al. claimed only few lines above in their paper.

our results are deeply flawed and ended with a failed attempt at reassessing their effect size. At the same time, they claimed that a theory based on simple and ubiquitous associative mechanisms can "exactly" accommodate them (Perruchet et al., p. 573) but offered no hint as to how this theory would explain the generalization to novel items which they also attested. Thus, how Perruchet et al. might explain the results of their own experiment, let alone ours, remains a mystery.

Statistics and Speed of Learning Generalizations

So far, we have shown that Perruchet et al.'s (2004) Experiment 2 (barring its multiple methodological problems) replicated our findings that participants project generalizations after exposure to a short discontinuous speech stream. However, our critics neglected other crucial findings that we reported—for example, that prolonging exposure to the continuous stream induces a preference for part words over rule words (Peña et al., 2002, Experiment 4). It was the contrast between this result and the preference for structurally correct but previously unheard items observed after short familiarization with a segmented stream that led us to favor a dual-mechanism account. After noting that Perruchet et al. also attested to the existence of fast learning of structural relations, we wonder what alternative explanation they propose. We were expecting to see this issue discussed in an article that promises to show that there is "no need" for the kinds of computations that Peña et al. proposed; however, we found only two passing observations about how structural information can be quickly extracted.

In the first attempt, Perruchet et al. (2004) argued that Peña et al.'s (2002) long-distance relation effect arises because the middle element of words is not encoded:

> Both experimental evidence and everyday experience suggest that the start and the end of a sequence captures [sic] more attention than the intermediary events. Thus it is likely that when the auditory stream is perceived as a succession of artificial words, participants pay more attention to their first and last syllables than to their middle one, and then encode those syllables as well as the relevant positional information. This prompts the formation of AXC units, where A and C are specific syllables and X stands for unnoticed events. (p. 582)

Indeed, we suspect that edge phenomena may be involved in the quick projection of rules (Endress, Scholl, & Mehler, 2005); however, the suggestion that our results reduce to the fact that the internal syllable goes unnoticed is wrong. To exclude that the generalizations from short streams are due to the forgetting of the middle element, Peña (2002, Experiment 16) compared rule words against actual words after giving participants 2 min of exposure to a segmented stream. As rule words and actual words differ precisely in their middle element, had this internal syllable gone unnoticed, participants should have shown no preference. In fact, they preferred words, showing that the middle element is indeed encoded. Hence, our critics are wrong.

In their second attempt, Perruchet et al. (2004) hinted at a putative statistical explanation of fast learning of structural generalizations:

> The assertion that associative learning proceeds slowly does not stand up to empirical observations. . . . Saffran, Aslin, and Newport (1996) observed that babies were able to segment an artificial language presented as a continuous speech flow after only 2 min of exposure. Now, this phenomenon is commonly attributed to statistical mechanisms, even by Peña et al. Overall, the learning of the A–C relationships proceeded at a rate that, roughly speaking, is quite compatible with a statistical or distributional approach. (p. 582)

Here, Perruchet et al. (2004) compared the ability of adults to rapidly extract structural information reported in Peña et al. (2002) with Saffran, Aslin, and Newport's (1996) infant experiments on word identification as if both sets of experiments were the same thing. Perruchet et al. seemed to believe that because the exposure duration is identical in both sets of studies, participants must be doing the same computations: Thus, if Saffran, Aslin, and Newport's infants learn by statistics, so must Peña et al.'s adults.

Perruchet et al.'s (2004) proposed explanation compares the extraction of structural information after exposure to a discontinuous 2-min stream, as in Peña et al.'s (2002) Experiment 5, with the segmentation of words after a 2-min continuous exposure, as in Saffran, Aslin, and Newport (1996). However, even the most cursory reading of Peña et al. should suffice to persuade researchers that no extraction of structural information is observed after participants are familiarized with a continuous stream: not with 2 min (Peña, 2002, Experiment 13), not with 10 min (Peña et al., 2002, Experiment 2), and not with 30 min (Peña et al., 2002, Experiment 5).

Perruchet et al. (2004) also neglected the fact that in Saffran, Aslin, and Newport (1996), infants had to learn four words defined by high TPs among adjacent syllables. In Peña et al.'s (2002) relevant experiments, adults had to extract nine words requiring the computation of TPs among nonadjacent syllables. The experience needed to extract nine words exploiting syllable TPs in a continuous stream can be assessed rather accurately. Saffran et al. (1997) used 21-min familiarization with adults to establish parsing of a stream with six items. Peña (2002) and Peña et al. ran extensive pilot studies to assess the minimal exposure needed to identify such a lexicon, when defined by both adjacent and nonadjacent TPs, and found that under about 10 min of exposure, no segmentation occurs. Thus, comparing Peña et al.'s adult experiments with Saffran, Aslin, and Newport's (1996) infant experiments is meaningless: With 2 min of exposure to a continuous stream, a six- to nine-word lexicon is not extracted.

But even disregarding these misrepresentations, a substantive issue remains. We never denied that a theory based on statistical learning might account for fast learning; however, the thesis that all learning can be explained by statistical computations is empty unless our critics can propose a single mechanism that is capable of simultaneously explaining (a) segmentation of words after exposure to a long familiarization (but not to a short familiarization) with a continuous stream, (b) extraction of structural information after a short familiarization with a discontinuous stream, and (c) failure to extract the same information after familiarization with any continuous stream. Without explaining the phenomena they so sharply criticize, Perruchet et al. (2004) claimed that statistically based alternatives to Peña et al.'s (2002) dual system may exist, yet they provide no credible alternative.

Generalization, Segmentation, and Simulations

Perruchet et al. (2004) asserted that our conclusions are flawed because their model of segmentation, PARSER, simulates Peña et al.'s (2002) Experiment 1 and the frequency control used in that experiment (Peña et al., 2002, Footnote 16; Perruchet & Vinter, 1998). We leave it to a future article to assess how PARSER fits the data of these two experiments. Here, we only draw the reader's attention to two simple mistakes in this argument.

First, our critics argue that because we used the same sequences of words during the study phase of all of our experiments, Perruchet et al.'s (2004) simulation of Experiment 1 generalizes to all of our data; however, by construction, PARSER can extract only chunks that actually occurred in a stream. Because the model has no way to account for the extraction of generalizations, that is, for data that are central to the dual-model mechanism that our critics reject (while replicating our observations), their argument that simple associative mechanisms such as PARSER fit all the data is groundless.

Second, Perruchet et al. (2004) misrepresented the purpose of Peña et al.'s (2002) Experiment 1 and its control. These experiments show that participants can capture second-order relations among *syllables.* Distant TPs among single syllables may explain this ability. Possibly, the computation of frequency of *chunks* of different size might also explain this ability—in principle, it trivially can because a chunking mechanism is computationally more powerful than is a mechanism based on TPs. The point is that these are different means by which language learners may exploit regularities in a stream to break a continuum when adjacent relations among syllables do not allow them to do so. Thus, if PARSER can simulate our Experiment 1, the conclusion that Perruchet et al. should have drawn is the same conclusion that Peña et al. drew—namely, that learners are able to track long-distance relations among syllables in a stream.

In fact, there is no reason to assume, as Perruchet et al. (2004) did, that nonadjacent TPs and chunking are alternatives. In all likelihood, a parsing algorithm will exploit either set of information to achieve segmentation; however, considerations of simplicity should invite our critics to take a more sober attitude toward their model. PARSER contains several free parameters (never studied in detail by its authors) that can be adjusted to simulate experiments ad lib.[3] A mechanism based on TPs accounts for Peña et al.'s (2002) results and Saffran and her collaborators' (e.g., Saffran, Newport, & Aslin, 1996; Saffran et al., 1997) results on segmentation without the need of adjusting parameters. In the absence of further evidence, the preference should be for the simpler theory. Hence, even granting the soundness of Perruchet et al.'s simulations, much remains to be done before researchers can conclude that humans rely on chunking, as opposed to computing distant TPs, to capture nonadjacent relations among components of a continuous stream.

In short, simulations with PARSER, if successful, formally confirm the possibility to compute second-order dependencies among syllables in a stream. Therefore, far from showing that our data are flawed, Perruchet et al. (2004) supported the phenomenon that we documented.

Conclusion

Quite early in life, humans can acquire and master many of the complex structures involved in natural languages. Sensitivity to statistical properties in the linguistic data that the infant receives is an important source of information. A mechanism that projects generalizations after sparse linguistic data also may play a salient role during language acquisition. In Peña et al. (2002), we proposed a series of empirical theses, backed up by data replicated by our critics, supporting the existence of a dual mechanism. Yet, other less explored mechanisms may contribute to language acquisition. It is premature to conclude that a single, all-encompassing mechanism can accomplish such a task. Indeed, there is no reason to think that language learning will be less diverse and complex than other well-studied biological functions (Gallistel, 1993).

Bound to sail uncharted waters, scientists always run the risk of hitting some unknown rocks. It is all the more important, then, to stay clear of the visible ones. In their critical review, Perruchet et al. (2004) hit Scylla without being able to avoid Charybdis. They misconstrued our aims and conclusions, ignoring the evidence for structural generalizations that motivated them. They replicated most of our results, including those whose existence they denied, and neglected data that were problematic for their theses. Despite their efforts, we found that their attacks dissolve after cursory examination. Although, undoubtedly, the issues we raised in Peña et al. (2002) are far from being settled, Perruchet et al.'s reconstruction of our claims and their proposed alternative model, in our opinion, do little to advance our knowledge of language acquisition.

[3] For example, in their simulation of Saffran, Newport, and Aslin's (1996) experiments on adjacent probability learning, Perruchet and Vinter (1998) set forgetting to 0.05 and interference to 0.005. However, to simulate Saffran, Aslin, and Newport's (1996) results with infants, forgetting is increased from 0.05 to 0.08 because, as Perruchet and Vinter write, "this change is consonant with the common belief about the limited efficiency of infant memory with regard to adults" (p. 257). Instead, in the current article, forgetting and interference are both dramatically adjusted to 0.005 and 0.001, respectively, thus reducing the role of interference and increasing recalling beyond the reasonable. Taking seriously Perruchet and Vinter's logic that halving the chunk sensitivity epochs reflects the limited efficiency of infants' memory, we should conclude that humans will be able to succeed in Peña et al's (2002) segmentation experiments after memory efficiency has increased tenfold.

References

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

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

Bates, E., & Elman, J. (1996, December 13). Learning rediscovered. *Science, 274*(5294), 1849–1850.

Elman, J. L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 1–27). Mahwah, NJ: Erlbaum.

Elman, J. L. (2001). Connectionism and language acquisition. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings* (pp. 295–306). Malden, MA: Blackwell.

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

Endress, A. (2005). *Limitations perceptives et représentationnelles des* *computations symboliques mentales* [Perceptual and representational limitations of symbolic mental computations]. Unpublished doctoral dissertation, École des Hautes Études en Sciences Sociales, Paris.

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

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

Fiser, J., & Aslin, R., N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467.

Gallistel, C. R. (1993). *The organization of learning (learning, development, and conceptual change)*. Cambridge, MA: MIT Press.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology.* Dordrecht, the Netherlands: Foris.

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

Peña, M. (2002). *Rôle du calcul statistique dans l'acquisition du langage* [Role of statistical computations in language acquisition]. Unpublished doctoral dissertation, École des Hautes Études en Sciences Sociales, Paris.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002, October 18). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133,* 573–583.

Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language, 39,* 246–263.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

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

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

Saffran, J. R., Newport, E. L., Aslin, R. N., & Tunick, R. A. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101–105.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002, October 18). Does grammar start where statistics stop? *Science, 298*(5593), 553–554.

Selkirk, E. O. (1984). Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press.

Accepted August 31, 2005 ■