Journal of Experimental Psychology: General
2006, Vol. 135, No. 2, 322–326

0096-3445/06/$12.00 DOI: 10.1037/0096-3445.135.2.322

Do We Need Algebraic-Like Computations? A Reply to Bonatti, Peña, Nespor, and Mehler (2006)

Pierre Perruchet and Ronald Peereman
University of Bourgogne

Michael D. Tyler
University of Western Sydney

L. L. Bonatti, M. Peña, M. Nespor, and J. Mehler (2006) argued that P. Perruchet, M. D. Tyler, N. Galland, and R. Peereman (2004) confused the notions of segmentation and generalization by ignoring the evidence for generalization in M. Peña, L. L. Bonatti, M. Nespor, and J. Mehler (2002). In this reply, the authors reformulate and complement their initial arguments, showing that their way of dealing with segmentation and generalization is not due to confusion or ignorance but rather to the fact that the tests used in Peña et al. make it likely that neither segmentation nor generalization were captured in their experiments. Finally, the authors address the challenge posed by Pena et al. of accounting for the whole pattern of their results without invoking rule-based, algebraic-like computations.

Keywords: statistical computation, rule-based computation, associative learning, language

Perruchet, Tyler, Galland, and Peereman (2004) challenged the conclusions of an earlier *Science* article by Peña, Bonatti, Nespor, and Mehler (2002). In their comment, Bonatti, Peña, Nespor, and Mehler (2006) found that our "attacks dissolve after cursory examination" (p. 21). Space limitations prevent an exhaustive examination of the points raised in this comment. Thus, we focus here on the fundamental issue at hand: Does Peña et al.'s (2002) article give evidence for rule-based, algebraic-like computations, construed as qualitatively different from statistical processes?[1]

The recurrent criticism of Bonatti et al. (2006) is that we confused what Peña et al. (2002) called *segmentation* and *generalization* (e.g., pp. 7, 8, 11). They claimed that we have ignored the evidence for generalization (e.g., p. 12), which they posited as an existence proof for algebraic-like processes. In fact, we organized our article around a set of methodological issues common to all of their experiments with the aim of demonstrating that Peña et al. captured neither of those two processes. The Bonatti et al. comment suggests, however, that a more focused critique of the two hypothesized processes may be useful at this point. We thank the authors of Bonatti et al. for bringing this to our attention and for giving us this opportunity to restate our arguments more clearly. In this reply, we discuss successively the problems that we consider to be inherent to the notions of segmentation and generalization.

Pierre Perruchet and Ronald Peereman, Laboratoire d'Etude de l'Apprentissage et du Developpement/Centre National de la Recherche Scientifique (LEAD/CNRS), University of Bourgogne, Dijon, France; Michael D. Tyler, MARCS Auditory Laboratories, University of Western Sydney, Sydney, New South Wales, Australia.

This work was supported by a grant from the French Ministère de la Recherche (Action Cognitique: Ecoles et Sciences Cognitives) to Ronald Peereman, and grants from the CNRS and from the Université de Bourgogne to the LEAD/CNRS, UMR 5022. Michael D. Tyler was supported by a postdoctoral fellowship from the Conseil Régional de Bourgogne and by National Institutes of Health Grant DC00403 to Catherine Best.

Correspondence concerning this article should be addressed to Pierre Perruchet, Université de Bourgogne, LEAD/CNRS, Pole AAFE, Esplanade Erasme, 21000, Dijon, France. E-mail: pierre.perruchet@u-bourgogne.fr

On Segmentation

Participants in Peña et al. (2002) were first exposed to artificial languages composed of the concatenation of *AXC words*—trisyllabic units in which A perfectly predicts C and X is an unrelated variable syllable. This exposure phase was followed by a two-alternative forced-choice test consisting of a word (AXC) and a part word (XCA or CAX), with instructions to select the item that seemed to be more similar to a word of the imaginary language. Peña et al. explicitly reasoned that if participants discovered the relations between A and C, then they would organize the stream into AXC words, hence selecting the words against the part words in the test. Then, from selection of the correct item at test, they inferred that the computation of the nonadjacent statistical relations among syllables (hereafter noted as A_C relations) had occurred. The point of Perruchet et al. (2004) was that this inference in fact requires that remote dependencies are not only one potential source but the exclusive source of information about the AXC words available in the experimental context. As we wrote, "If participants also rely on other cues to segment the auditory string, then measuring the learning of remote dependencies through its effect on word segmentation becomes clearly unwarranted" (p. 575). Most of our article was devoted to showing that Peña et al.'s languages embedded many cues for segmentation other than the nonadjacent dependency structure.

Training-Independent Factors

We called a first category of segmentation cues *training independent*, meaning that these cues were directly exploitable by the participants, presumably due to gestalt-like perceptual mechanisms and/or to participants' prior experience with natural languages. To examine whether training-independent cues may have influenced participants' performance in the Peña et al. (2002) study, we

[1] An exhaustive, point-by-point discussion of the comment is, nevertheless, available on http://www.u-bourgogne.fr/LEAD/people/perruchet.html

exposed participants to a language that keeps unchanged most of the features of Peña et al. languages with the exception that all combinations between the A and C syllables were allowed, thus breaking the nonadjacent dependency. Our reasoning was that correct segmentation of such a modified language would be proof that segmentation cues other than the A_C relationships were available. The data showed, indeed, that participants immediately perceived 52.73% of the words as trisyllabic, and among these trisyllabic words, 72.35% followed the AXC pattern (when chance is at 33.33%, because three patterns are possible: AXC, XCA, and CAX). Onnis, Monaghan, Chater, and Richmond (2005) used the Peña et al. word/part word forced-choice test after familiarizing participants with the same language as ours in their Experiment 3 and obtained results similar to ours.

We found three kinds of arguments against our demonstration in the Bonatti et al. (2006) comment. First, they noted that we failed to find a clear explanation for participants' performance. Indeed, we considered various possible explanations for this part of the results without achieving the isolation of one or a few responsible factors, and in our conclusion we suggested the involvement of multiple segmentation cues interacting together. We maintain, nevertheless, that our inability to isolate individual cues does not invalidate the empirical demonstration that these cues were present—participants segmented the stream into AXC units even without the presence of a rule. More recent experiments have begun to identify the nature of those cues (e.g., the presence of plosive consonants at the beginning of words, studied by Onnis et al., 2005).

A second Bonatti et al. (2006) counterargument is that Peña et al. (2002) themselves provided a control for training-independent factors by keeping the test items constant but inverting the items' status as word/part word. However, the composition of their material made it impossible to achieve a full inversion between words and part words. In fact, as can be seen in Table 1 of Bonatti et al., only 8 items of 18 were fully inverted. Moreover, the rate of correct segmentation was 73% in the main experiment and only 58% in the control experiment. The fact that the difference with chance (50%) fell from 23 percentage points to 8 percentage points in the control experiment, when less than half of the items had been inverted, could just as easily support the claim that participants segmented A_C units because of training-independent factors. Further support for the idea that Peña et al.'s control experiment did not adequately counteract the effect of training-independent factors can be found in Newport and Aslin (2004), who obtained no evidence of learning when they used a language without phonological confounds but observed positive results when they reintroduced the phonetic cues deemed responsible for Peña et al.'s results. Likewise, Onnis et al. (2005) reported chance performance when they counteracted position-specific training-independent factors by using a different random allocation of syllables to words for each participant (Experiment 2), and they found a reverse preference for part words when they inverted the first and second syllables of each word (Experiment 4).

In a third counterargument, Bonatti et al. (2006) argued that participants in our experiment may have learned the AXC words because "AXC words had [a] higher absolute frequency than did any of the part words generated by the concatenation" (p. 317). This is not the case: Our materials were constructed in such a way that the frequency of words and the frequency of any part words were exactly the same—our language was "statistically flat," as indicated in Perruchet et al.

To conclude, we maintain that our results, supported and extended by those of Newport and Aslin (2004) and Onnis et al. (2005; see also Seidenberg, MacDonald, & Saffran, 2002), show that the materials used in Peña et al. (2002) were biased by training-independent factors.

Training-Dependent Factors

In Perruchet et al. (2004), we also considered the (nonexclusive) possibility that the correct segmentation of Peña et al.'s (2002) language may have been due to some unwanted statistical regularities present in the speech flow, beyond the A_C relationships on which the authors focused. Because these regularities need to be learned before performance can be affected, we call them *training dependent*. To test for the existence of such regularities, we entered Peña et al.'s language into PARSER (Perruchet & Vinter, 1998), a computational model that is, by construction, unable to exploit nonadjacent statistical dependencies. Indeed, the only information exploited by PARSER is composed of chunks of adjacent elements. Our reasoning was that if PARSER achieves segmentation, then the input must contain some kind of information about the word structure unrelated to nonadjacent dependencies. PARSER did reach a rate of correct responses roughly similar to those obtained by the actual participants of the Peña et al. experiments.

Bonatti et al.'s (2006) counterargument appears to be that if PARSER is able to segment Peña et al.'s language, then the conclusion should be that PARSER is able to learn the A_C relationship. This ability would proceed from some still-undetected power of the chunking mechanisms, the discovery of which the authors postponed to a future article. However, the reasons why PARSER achieves segmentation are all but mysterious. As an example, in all Peña et al.'s (2002) experiments (except an additional experiment reported in their footnote 16), the AXC words were twice as frequent as any part words during the familiarization phase. PARSER obviously learned thanks to such uncontrolled cues, as reported in Perruchet et al. (2004). PARSER did not learn that three families of AXC words exist, each composed of three members. Instead, it learned nine distinct words without exploiting, in any way, the internal structure of these words—no hidden power of chunking is required to explain PARSER's performance on those materials.

This being said, we do not conceive of this demonstration as evidence that participants performed as PARSER did. It is even clear that they did not: PARSER is immune to any a priori knowledge and biases and hence is insensitive to the training-independent factors that influenced participants' performance. Thus, the uncertainty remains as to whether the confounds that PARSER reveals were actually exploited by the participants. Regardless of the true state of affairs, however, our analysis provides a note of caution for future studies: Even if all of the training-independent factors (mainly phonological) had been carefully controlled, the design implemented in Peña et al. (2002) could not have ruled out the influence of regularities in the language, other than distant relations, that can be captured by elementary memory processes such as those implemented in PARSER.

On Generalization

In the Peña et al. (2002) test of generalization, the part words were the same as those used in the segmentation test. However, the AXC test words now differed from the AXC words heard during the familiarization phase because their X syllable was changed. Peña et al. called the resulting item, hereafter denoted as AX*C, a *rule word*. The postulate of Peña et al. was that selecting AX*C rule words over part words gives evidence of what they call *generalization*.

At a descriptive level, we do not dispute that the term generalization is appropriate; however, when Peña et al. (2002) and Bonatti et al. (2006) wrote about generalization, they referred to the specific process that they construed as responsible for generalization in their experimental setting. This process is identified as the discovery of the rule "If A_i occurs then C_i will follow after an intervening X" (Peña et al., p. 605). For the sake of clarity, we will refer to this as *structural generalization* to clearly distinguish the inferred process from the test that, allegedly, taps this process.

An Unreliable Measure of Structural Generalization

Our criticism of structural generalization parallels that of segmentation. We do not take issue with the fact that the discovery of the rule "If A_i occurs then C_i will follow after an intervening X" (Peña et al., p. 605) may lead to the selection of AX*C rule words over part words during the test phase. However, we stand by our claim that a researcher cannot conclude that discrimination at test reflects the abstraction of algebraic rules unless simpler hypotheses have been ruled out.

In this context, it is easy to show that success on the test of generalization can be achieved through use of a very simple strategy that does not imply rule discovery in any way. As a preliminary note, let us recall that above-chance performance in the generalization test occurred only in a condition in which 25-ms silent pauses were introduced between the AXC words during the study phase. Perruchet et al. (2004) provided experimental evidence that in this condition, the segmented speech flow generally was perceived as a succession of trisyllabic words from the outset. Indeed, in Experiment 1, 87.68% of the items written by the participants, who were asked to write the words that they perceived in the speech flow, were AXC words. Thus, it seems reasonable to assume that, thanks to the presence of pauses, an overwhelming proportion of participants knew the AXC words by the end of the familiarization phase.[2]

The central point is what happens during the test, when participants are asked to say which item, between an AX*C rule word and a part word (XCA or CAX), is more similar to a word of the imaginary language. As an illustration, imagine that participants know the words *puliki* and *taraga* and have to choose between *pubeki* (a rule word, with *be* serving as X*) and *likita* (a part word). Of course, they may select *pubeki* because this item respects the long-range dependency rule "If *pu* then *ki,*" as the authors postulate. But participants can use a far simpler means of achieving success in this test. It is sufficient to consider that *pu* is a correct word beginning, whereas *li* never occurs in this location. Alternatively, it is also possible to consider that *ki* is a correct word ending, whereas *ta* never occurs in this location. More generally, participants may select the rule words while relying exclusively on the A syllables and/or on the C syllables.

The scenario described in the previous paragraph is simpler than that postulated by Peña et al. (2002) because identifying the A syllables and/or the C syllables (the ability that we assume) is a necessary prerequisite for discovering the distant relationships among these syllables (the ability that Peña et al. assumed). In Peña et al.'s interpretation, participants need to have noticed not only either A or C but both A and C and, moreover, the relationships between a specific A syllable and a specific C syllable. In summary, exploiting fragmentary information about the AXC words is sufficient to achieve a perfect discrimination on the generalization test without having learned anything about the A_C dependency rule.

Comparing Pena et al.'s (2002) Test With a More Sound Test

Thus far, we have shown the possibility that the selection of AX*C rule words over part words is based on other, simpler features than the A_C dependency rules. In Perruchet et al. (2004), we demonstrated that this theoretical possibility was effective by comparing the scores obtained in the Peña et al. (2002) test with the scores obtained in a test of nonadjacent dependency learning inspired by Gomez (2002). In the example cited two paragraphs earlier, the test would contrast the rule words *pubeki* or *tabega* with items such as *pubega* or *tabeki* (with *be* as an intermediary, new syllable). We have called the latter items *scrambled words* because they are derived from the words by scrambling the relations between their first and last syllables. Selection of the rule words over the scrambled words cannot rely on participants' fragmentary knowledge of the first or last syllable of each word because these syllables are presented in their correct location in the two kinds of items. The only difference between rule words and scrambled words lies in the relations between a specific A and a specific C, hence the relevance of this comparison.

The performance, as measured by the rule-word-versus-part-word test pairs borrowed from Peña et al.'s (2002) test substantially exceeded the performance assessed by our rule-word-versus-scrambled-word comparison. Thus, our overall conclusion was that the better part of the effect obtained with the comparison

[2] Bonatti et al. (2006) are somewhat difficult to follow on this aspect. On the one hand, they reiterated the Peña et al. (2002) claim according to which the addition of "subliminal" pauses leads to a speech flow "subjectively very similar" (e.g., p. 606) to the continuous speech flow. This claim suggests that the segmented language is not perceived as a succession of discrete words. On the other hand, they acknowledged that pauses provide explicit bracketing cues and that they facilitate the segmentation into words. Moreover, the fact that structural generalization can operate only on a discrete corpus is a cornerstone of their position. "Generalization to structure in syntax and morphology," they wrote, "always appears to require a corpus with constituent structure" (p. 12). We infer that for Bonatti et al., whether a constituent structure is present does not alter subjective experience, and we conclude that our dispute here is only a matter of defining what counts as similar and dissimilar subjective experiences.

borrowed from Peña et al. was due to factors other than genuine nonadjacent dependency learning.[3]

Bonatti et al. (2006) did not challenge the validity of the test that contrasts rule words with scrambled words. However, they argued that our methodology vitiates the comparisons we drew between the tests. They reason, "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" (p.). This criticism is unwarranted. Bonatti et al.'s reasoning would apply if the task were, for instance, a recognition test, in which, indeed, rejecting an item in one comparison is logically incompatible with endorsing the same item as correct in another comparison. However, participants in our experiment were asked for something similar to a word-likeness judgment. Judging that *yas* is more similar to an English word than *yik* is quite compatible with judging that *yik* is more similar to an English word than *ytz*. Our analysis was valid and demonstrates unambiguously that the Peña et al. (2002) measure is sensitive to the effect of variables other than those involved in true nonadjacent dependency learning.

Considering Peña et al.'s (2002) Entire Pattern of Data

Peña et al. (2002) claimed to have shown the need for positing two different processes to account for their data. On one side are statistical processes, which are held to be responsible for the participants' ability to segment a continuous speech flow into trisyllabic words, defined by the relationships between their first and third syllable. We have shown that segmentation may be attributed to a number of factors, both training dependent and training independent, which are unrelated to the A_C relationships. On the other side are algebraic-like processes, which supposedly account for the participants' ability to generalize the A_C relationships to items including other intervening syllables when the speech flow is presegmented into AXC units. We have shown that above-chance performance in the generalization test does not require knowledge about the A_C dependency rule, a conclusion that led us to the somewhat provocative assessment that "What Peña et al. showed is that adding segmentation cues helps segmentation" (p. 581). Bonatti et al. (2006) took issue with this assessment, claiming to have shown instead that adding segmentation cues helps generalization (as opposed to segmentation). Our shortcut was, in fact, intended to mean that once segmentation is given, success in the generalization test may be ensured by elementary, well-known forms of learning, given that, for instance, knowledge that A is a word's beginning (or C is a word's ending) is sufficient to drive the selection of an AX*C rule word over a part word.

Thus, the point we have shown so far is that Peña et al. (2002) gave no compelling evidence for segmentation or structural generalization. However, Bonatti et al. (2006) rightly noted that we did not show how our interpretation accounts for the dramatically different pattern of results that they observed regarding whether the speech flow included pauses. We now address this important challenge. Recall that Peña et al. observed that after undergoing familiarization with a 10-min continuous speech flow, participants failed to discriminate rule words (AX*C) and part words. When training was extended to 30 min, participants even selected part words over rule words. However, when 25-ms silent pauses were introduced between the AXC words, participants selected rule words over part words after 10 min of familiarization. The same level of performance was even reached after only 2 min of familiarization. According to Peña et al., the spectacular reversal of performance with segmented speech was due to the fact that silent gaps "make the stream slightly more similar to natural language" (p. 606). In so doing, they trigger algebraic-like processes that are oriented toward the discovery of the speech structure and allow a nearly immediate discovery of the nonadjacent dependencies.

The observed pattern of results is indeed impressive, and we can see how this pattern may lead researchers to hypothesize the action of qualitatively different processes. But if one considers that success in the generalization test may be due to the exploitation of fragmentary information about the words, then this explanation emerges as a more parsimonious interpretation. Indeed, success in the generalization test after a short exposure, which appears so demonstrative of powerful abstractive processes to Peña et al. (2002), now becomes all but surprising. A reversal of performance with further practice also becomes predictable because, with additional exposure, participants presumably learn more about the words than about their first (or last) syllables. It is noteworthy that this additional knowledge could plausibly impair performance in the generalization test. Choosing AX*C over XCA or CAX is indeed more difficult when participants know the whole AXC words than when they have only partial knowledge of them because this additional knowledge makes participants fully aware of the difference between AX*C (the test rule word) and AXC (the words that participants heard). Therefore, attributing success in the generalization test to the use of partial information about the words explains the early appearance of learning and the reversal of performance when more complete information is acquired about the words.

This explanation does not mean, however, that introducing the 25-ms silent pauses is deprived of any consequence. Peña et al. (2002) showed that after 10 min of familiarization, participants achieved the generalization test only when they were exposed to the segmented speech flow. A natural explanation is that the pauses favor the discovery of the first or last syllables of the words. Moreover, a problem of bootstrapping may arise when the speech is continuous. Indeed, identifying the first (or the last) syllable of any word requires that the word boundaries have been previously identified, an operation that presumably involves the processing of the whole words when the boundaries are not directly given in the auditory signal. Thus, it is quite possible that the continuous speech flow never allows the participants to acquire the form of knowledge that is the most efficient for achieving the generalization test, namely, partial knowledge limited to the beginning or to the end of words.

[3] This does not mean that the learning of nonadjacent dependency is impossible. Perruchet et al. (2004) provided evidence for this form of learning after 3 min of exposure to the speech flow (Peña et al., 2002, used 2 min), which was significant thanks to the use of a large sample of participants ($N = 40$; in Peña et al., $N = 14$). Perruchet et al. proposed an interpretation for their results that does not rely on rule-based processes (pp. 581–582). Whether learning is still possible with a still shorter familiarization phase ought to be explored in future studies.

Although a definitive evaluation of alternative interpretations would need further studies, it is worth stressing that our interpretation naturally accounts for at least one aspect of the Peña et al. (2002) data that is hard to encompass within their own framework. The point of concern is the Peña et al. observation of no significant improvement from 2 min to 10 min of familiarization with the segmented speech stream in the generalization test. This result is consonant with our interpretation, as explained above. However, if one endorses Peña et al.'s framework, one may wonder why the powerful abstractor that results in 67.1% correct responding after 2 min does not reach closer to 100% after 10 min, which it should do if deterministic rules have truly been discovered.

Conclusion

After a careful examination of Bonatti et al.'s (2006) comment, our initial conclusion remains unchanged: All of the data reported in Peña et al.'s (2002) article can be accounted for without assuming the existence of algebraic-like computations. This conclusion stands in striking contrast to that of Peña et al. We are unclear, however, about the extent to which it also contrasts with that of Bonatti et al. Indeed, although Bonatti et al. considered that all of our criticisms were irrelevant, they also noted that the algebraic character of structure extraction "is compatible with, but not mandated by" (p. 8) their results. This expression suggests that the authors acknowledge that other interpretations are possible and, presumably, that those alternative interpretations, such as ours, rely on simpler processes. If this is the case, it raises an important question: What is the interest of advocating a rule-based framework when simpler alternatives are available? We surmise that an advocate of a rule-based framework must present data that require (instead of being only compatible with) the sophisticated computations inherent in this framework because more parsimonious interpretations have failed. As we have shown in Perruchet et al. (2004) and in this reply, simpler explanations exist. Therefore, we maintain that performance of participants in Peña et al. was not due to the extraction of rule-like regularities.

References

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2006). How to hit Scylla without avoiding Charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004). *Journal of Experimental Psychology: General, 135,* 314–321.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

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

Onnis, L., Monaghan, P., Chater, N., & Richmond, K. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53,* 225–237.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002, October 18). Signal-driven computations in speech processing. *Science, 298,* 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 for word segmentation. *Journal of Memory and Language, 39,* 246–263.

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

Received October 3, 2005
Revision received February 22, 2006
Accepted February 22, 2006 ■