

WARNING: PROVISIOANL VERSION
(Some argument are still lacking, and English style will be improved)

The text below offers detailed remarks about the following comment (in press in JEP:G) on Perruchet, Tyler, Galland, & Peereman, 2004:

Generalization, segmentation and language learning : How to hit Scylla without avoiding Charybdis

by Luca L. Bonatti, Marcela Peña, Marina Nespor, and Jacques Mehler

The comment of Bonatti et al. is exhaustively reproduced in italics (except the introductory part, which only recall the authors' thesis).

Recently, Perruchet et al. argued that our experiments and conclusions are "flawed by deep methodological inadequacies" (p. 573), that our data are "invalid"(573) and "unreliable" (579), that our work contains a "major logical pitfall" (574), and concluded that "What Peña et al. showed is that adding segmentation cues helps segmentation, and that is all"(581)..

Peña et al.'s 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.

We have never attributed this claim to Pena et al. Our point was that postulating that *the remote dependency between syllables is the exclusive source of learning about their relation* is a condition to make the Pena et al.'s method valid. This is an undisputable methodological assertion. Why do we talk about a logical error in our paper, rather than about a methodological drawback? Because Pena et al tentatively justified their method with an argument that is logically flawed. As we wrote: "The authors' explicit reasoning was that if participants discover the A_C relationships, then they would organize the stream into AXC words. However, a stronger condition is in fact required. Inferring the discovery of the A_C relationships from the segmentation into AXC words requires that remote dependencies are, not only one source, but the *exclusive* source of information exploited by participants to parse the speech flow into the AXC words. 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). Authors' reasoning amounts to infer "q therefore p" from the premise "if p then q". For the inference to be valid, the premise must be "if p *and only p* then q"

By stating, in our original paper, that "a stronger condition is in fact required", beyond "the authors' explicit reasoning", we thought having made clear that the authors did not realize themselves the need for this postulate.

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" (p. 574). Thus, we would have only shown that adding segmentation cues helps segmentation.

Notice that "remote dependency learning" is an ambiguous formulation that covers both 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.'s 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 generalize structural properties of the items, whereas probability relations among syllables do not promote generalization but rather segmentation. Perruchet et al. confuse the two, as their summary of our work clearly shows: [Peña et al.] (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 experiment 1 shows the opposite of what our critics claim: word segmentation is 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.

The point is discussed in our reply. The authors charge us of not having made ours their own distinction between two kinds of processes. But if one does not borrow authors' conceptual framework and terminology, it is not necessarily because one fails to understand their proposal. It may be also because one judges this conceptual framework and the related terminology unwarranted. It was precisely our point to show that Pena et al. measured neither of the two alleged processes.

In their summary, Perruchet et al. also confuse other issues that we kept separate. They conflate the language-specific nature of structure extraction (which we did not argue for);

Pena et al. claimed that structure extraction was possible thanks to the fact that their procedure (adding silent pauses between words) made "the stream slightly more similar to natural language" (Pena et al, p. 606). We indeed (mistakenly, according to Bonatti et al.) construed this claim as an argument for the language-specific nature of structure extraction.

its algebraic character, which is compatible with, but not mandated, by our results;

Pena et al. were not so cautious in their conclusions. We are pleased to see that the authors now acknowledge that other interpretations of their results are possible. For a remark about this claim, see the conclusion of our reply.

and the similarity between natural language and a subliminally segmented artificial speech stream (highlighted by the constituents, and not by the presence of pauses).

Pena et al wrote: "our silent gaps may be the last resort that this system [a system looking for structure in speech] exploits to make a stream more "natural"" (p.606). Again, we (mistakenly, according to Bonatti et al.) construed this claim as an argument for the fact that the system exploits the presence of pauses.

But Perruchet et al.'s main confusion is about the nature of our argument. They argue that our conclusion does not logically follow from our alleged premises. It obviously does not: we never proposed a logical argument to begin with,

We agree on this point. The Pena et al.'s failure to realize the logical pitfall inherent to their methodology is even the starting point of our paper. See our reply.

but presented empirical data and empirical hypotheses that Perruchet et al. fail to keep apart.

Indeed, we did not see in the Pena et al's Science paper a set of "empirical data and empirical hypotheses". The fact that the authors are seemingly unaware of their strong theoretical entrenchment could explain their difficulty to understand our paper.

Peña et al.'s methodological errors

Perruchet et al. claim that our methodology -- the same used by many researchers in this area (e.g., Saffran, Aslin and Newport, 1996; Saffran, Newport, Aslin and Tunick, 1997; Newport and Aslin, 2004) --

The objective of the famous Saffran et al. studies was to measure word segmentation. In this case, we indeed have no concern with the use of a word segmentation test, namely a word vs. part-word comparison. Pena et al. study was aimed at studying the learning of nonadjacent dependencies, and, as we showed in our paper, the word vs. part-word comparison is no longer adequate.

is deeply flawed, our result being vitiated by training independent and dependent factors. Training independent factors are due to perceptual biases acquired when learning the maternal language. Training dependent factors are "cues other than distant dependencies" contained in the familiarization of our experiments. Let us examine how such factors would conspire to produce what Perruchet et al. label our "invalid data" (p. 573 and 575).

Training independent factors, segmentation and generalization

Every participant who comes to the experimental room has years of experience with her maternal language. This truism, which Perruchet et al. hold 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 share the consonantal structure stop - continuous - stop (e.g., PULIKI), whereas part-words do not (Seidenberg, MacDonald, and Saffran, 2002; Newport and Aslin, 2004).

In the test above, the emphasis on "some" (of their experiments) is from the authors. In fact, *all* the experiments reported in the main text of Pena et al. shared this feature, as noted by Seidenberg et al. (2002) The unique (and partial) exception is reported in their footnote 17, discussed in our reply.

In their experiment 1, Perruchet et al. investigate whether this pattern, as well as other factors, could account for our results. For example, they consider the base rate distribution of the word initial syllables in French; the distribution in trisyllabic French words of the consonantal patterns we used; the parameters of the synthesizer Mbrola; and several other factors. They find that none of them could lead participants to make the choices we reported.

This presentation is truncated, because it passes over in silence the results of our Experiment 1. These results demonstrated unambiguously that the test of segmentation used by Pena et al. does not measure the learning of non adjacent dependency. Indeed, participants segmented the speech flow even though we had removed the non adjacent dependency (while keeping all the other features of the Pena et al's materials). It remains somewhat unclear why participants performed as they did, but this in no way undermines our demonstration that achieving the test of segmentation cannot be taken as a measure of nonadjacent dependency.

While the obvious conclusion should be that our experiments were not vitiated by such factors, our critics conclude that the perceptual biases generated by the familiarization "certainly originate from a variety of factors, the respective contributions of which should warrant further studies" (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-word. Because participants still prefer items according to their role as words or part-words, neither training independent factors, nor low-level phonological or phonotactic features, nor intrinsic preference for some test items, can account for the pattern of results 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) does not change the results. In all cases, the probability relations contained in the familiarization streams determine participants' preferences. Peña et al.'s 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. This information is available both in our paper and in Peña (2002).

We show in our reply that what Bonatti call their "control" ran in fact against their conclusions . Regarding the other (unpublished) experiments from their laboratory, which Bonatti et al often present as an undisputable evidence, we wait for their publication before commenting on them.

Yet, interested in showing that preference for items of AXC structure in our experiment 1 is due to uncontrolled training independent factors, Perruchet et al. press other criticisms. They obliterate 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. As participants tend to remember trisyllabic items after familiarization to continuous or discontinuous streams, Perruchet et al. conclude that participants "were sensitive from the start to some prosodic or phonological aspect of the auditory strings" (p. 578). While this may happen with the particular stream they used,

The "particular stream" we used was in fact patterned following that used in the five experiments reported in Pena et al. (except the removal of nonadjacent dependencies).

we have shown that such factors cannot explain participants' preferences in our segmentation tasks.

If the authors allude to what they called their "control", see our reply

More generally, it is doubtful that the continuous stream condition of Perruchet et al.'s experiment 1 demonstrates the influence of training independent factors. While our critics correctly point out that the absolute frequency of words in the stream of our experiment 1 may explain participants' preferences, they fail to point out that the same factor (which we controlled for in a separate experiment, but they did not) affects their experiment 1. Because their stream is obtained by concatenating 27 AXC words, AXC words have higher absolute frequency than any of the part-words generated by their concatenation. Therefore, finding that participants recall AXC items after familiarization to 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 in no way belittles Peña et al's 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.

On the authors' error about this argument, see our reply

Training dependent factors and generalization

The subliminal segmentation cues

The second factor that would invalidate our work is training dependent: the subliminal segmentation cues (experiments 3 and 5).

The subliminal segmentation cues were training *independent* factors. They are natural cues for segmentation, and their role is not learned during training with a specific language. Training dependent factors are the statistical regularities present in the familiarisation stream, which are extracted during the training phase.

Our critics comment our results writing that that since segmentation cues segment the stream, it's no surprise that people segment. However, our experiments with subliminal cues are not intended to prove segmentation, but generalization from a speech stream. As Perruchet et al. confuse segmentation with generalization they miss the point of our experiments.

Again, Bonatti et al. consider as a confusion our main claim, namely that Pena et al. failed to measure both "segmentation" and "generalisation" (in the sense they give to these concepts)

Our critics seem also 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., "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 present the fact that gaps induce a sense of segmentation as their original discovery.

In fact, the Pena et al. quotation that Bonatti et al. put forth as a proof of our misreading, was also reproduced in our original paper. Indeed, we wrote in p.574 of our paper: "Pena et al. acknowledged that "the gaps contained in the stream may help participants to segment", because they provide "explicit bracketing cues" (p.606)". We wonder why Bonatti et al claimed that we present the effect of gaps as our original discovery. But this childish attack overshadows the main question. In other part of the Pena et al. paper, it is claimed that the two conditions were "subjectively very similar" (p.606). Bonatti et al. echoed this assertion, claiming that the streams with and without pauses were "phenomenologically very similar" (e.g. p.5). We are puzzled. Do the authors conceive the fact of perceiving a speech stream either as a long set of syllables or as a sequence of discrete words, actually irrelevant for our subjective or phenomenological experience?

Our critics replicate our results on the subliminal nature of gaps but, they write, "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". In one condition of their experiment 1, Perruchet et al. familiarize participants to a stream containing gaps but deprived of non-adjacent dependencies among syllables, and find that participants remember mainly trisyllabic items.

Bonatti et al. pass over in silence an additional element in their description of our results: Among the trisyllabic items, participants reported 96.97 percent of the AXC words (when chance is at 33.33 percent)!

Because a similar result is obtained with a segmented stream containing non-adjacent dependencies, our critics conclude that remote dependencies do not help segmentation because segmentation cues were already helping word identification. For some reason, they take this result as a blow to Peña et al., but this 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 capture the generalizations that eluded them when the stream had no segmentation cues. Perruchet et al. ignore this.

Again, our alleged confusion between "segmentation" and "generalisation" is taken as responsible for the irrelevance of our remarks. What is missing here is the understanding of our key argument, namely that it is sufficient to segment the speech flow into words to achieve the test they construed as measuring generalisation. The reason was described in our original paper, and, given Bonatti et al's misunderstanding, is presented again in our reply. In a nutshell, when participants have segmented the speech flow into AXC words, and have to choose between AX*C and XCA or CAX (the test that the authors construed as measuring the learning of the dependency rule), they can choose the rule words on the basis of fragmentary information (e.g.: A is the beginning of a word), unrelated to the dependency rule. For instance, relying only on the syllables allowed at the beginning of the words would lead participants to 100% correct, in a test supposedly measuring the relations between the first and the last syllables.

Generalizations to structure in syntax and morphology always appears 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. criticize our claim that pauses make the stream closer to natural language, but they overlook its context. Natural language does not contain systematic pauses between words, but prosodic cues, often perceived as pauses, signal constituency (Nespor and 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 is irrelevant to our argument: we only made the point that without segmentation cues in the signal, generalizations are not projected.

We only made the point that talking about "generalisation" from their experiments is unwarranted

On two other related points Perruchet et al. are factually wrong. First, they insist that the sense of trisyllabicity induced by our streams per se explains our results. But trisyllabicity is to a large extent irrelevant. Perruchet et al. overlooked that Peña et al. 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).

Note that Bonatti et al. suggest that participants preferred quadrisyllabic item over trisyllabic items. This is misleading. In fact, Pena et al. wrote in footnote 29 that participants "listened to pairs of quadrisyllabic items" in the forced-choice test. The correct result is that, between two quadrisyllabic items, participants selected those obtained from the words over those obtained from the part-words. This result in no way undermines our argument. For instance, a strategy based on the knowledge of the first letter of the words heard during familiarisation still leads to 100 percent correct with quadrisyllabic items. A strategy based on the last letter works as well. In both cases, a perfect score is obtained without any knowledge about the dependency rule.

The second mistake clearly appears when they write: 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. (p.579, italics ours) It may appear trivial to Perruchet et al., but not to the participants to our experiment 4 who, exposed to a 30 min stream, selected part-words over rule-words. This shows that part-words are encoded during familiarization and even preferred to rule-words when extra exposure strengthens their memory traces. Once again, Perruchet et al. make statements that are contradicted by our experiments.

Was "ry/tra" displayed within the two sentences above? If a reader discovers that it was displayed in "memory traces", are we justified to conclude that "ry/tra" was encoded during the initial reading? When only nine words are repeated during 30 minutes, it is likely that participants are able to re-create the part-words from the words, as argued in our reply.

Pauses and the extraction of generalizations

In Peña et al. (2002), we explored whether participants familiarized to AB₁C, AB₂C and AB₃C words could capture the AXC relation accepting as "legal" items like 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 structure AXC over familiar part-words.*

Our critics focus 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 which draw the A and C syllables from different families. Thus, if a word in the stream has structure A₁BC₁ and another A₂BC₂, a scrambled-word would have structure A₁B^{}C₂. After using a 2 min discontinuous stream like Peña et al.'s experiment 5, in*

their second experiment our critics tested three kinds of comparisons: rule-words versus part-words, scrambled-words versus part-words, and scrambled-words versus rule-words. The first comparison is similar to the test conditions of our experiments 3 and 5, although less sensitive due to the scant number of comparisons tested (9 pairs per participant vs. 36 in Peña et al).

Indeed, increasing the number of observations often improves the accuracy of measures. This is generally the case in language research, in which the authors are indisputably competent. However, this is true only if successive observations are independent assessments of the to-be measured phenomenon. In learning and memory research, this condition is generally not fulfilled by the repetition of test items. The idea that it is still possible to capture something interesting about the initial learning of a 2-mn speech flow after being exposed to more than a few pairs of test items is quite naive. Far from being positive (or even neutral), the repetition of test items can have detrimental consequences, due to the possibility of learning during the test (even when no feedback is provided). For instance, in Pena et al.'s tests, rule-words begin with one out of 3 syllables, whereas part-words begin with one out of 6 syllables. This means that among the 72 items played during the test (i.e. 36 couples), there are items the first syllable of which is twice more frequent than the first syllable of other items. Now, selecting those items (on the basis of the familiarity with their first syllable) leads to select the rule-words at the expense of the part-words, hence artificially improving performance (the same reasoning applies for the last syllables). Note that those confounding factors, which are well-known in the learning literature, were controlled in our test.

The only means to increase the accuracy of the measure in learning research consists in increasing the number of participants. It turns out that we used 40 participants, while Pena et al. used 14 participants. This difference is nowhere mentioned in the Bonatti et al.'s comment.

Nevertheless, they replicate our results that with little exposure participants select rule words over part-words. However, Perruchet et al. ignore this and claim that the only "correct comparison" (p.580) is between scrambled-words and part-words.

We wrote in p. 580: "Nine pairs contrasted rule-words and part-words. They provided the correct comparison. Finally, nine pairs contrasted scrambled words and part-words. This comparison was aimed at capturing the overall effect of the irrelevant factors that may have influenced the contrast used by Pena et al." This authors' error in a verbatim quotation is somewhat puzzling.

As participants prefer scrambled-words, they conclude that our data are invalid. This conclusion does not follow, for two reasons. First, Perruchet et al.'s methodology vitiates the results. Perruchet et al. tested the three comparisons within participants as if they 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, s/he might select γ , not because s/he prefers γ , but because s/he knows that s/he had already rejected β . Therefore, preference for scrambled-words over part-words might reflect preference for rule-words against part-words. With the non-independent alternative forced choice design Perruchet et al. use, it is impossible to interpret their results and accept their statistical analyses.

See our reply for a rebuttal of this comment.

Second, the comparison is indeed interesting, but the conclusions it suggests are not the ones Perruchet et al. advance. With methodologically sound tests, Endress (2005) independently

explored this comparison. What our critics call scrambled-words, and Endress calls class-words, are instances of a lawful relation among syllables occurring in 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 they captured this generalization. Many phenomena in natural language are reminiscent of class-words. For example, roots of words remain constant, but both prefixes and suffixes can vary independently when words are derived or inflected. Endress (2005) systematically explored sensitivity to class-word structure, and showed that under certain conditions participants capture it.

Endress 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 and rule-words, Perruchet et al. reach the same conclusion. However, this 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, in order 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.

It is quite ironic that the authors teach us about the validity of the test that we have proposed. However, we are pleased to see that they have understood why this test is sound (although seemingly, this does not lead them to realize that their own test is unsound).

Showing this is tantamount to establishing that participants are sensitive to the abstract structure of distant elements in words in virtue of their ability to generalize. But this was exactly Peña et al.'s conclusion: participants 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.

In contradiction with the authors' allegation, we claimed quite explicitly that "we found reliable evidence for the possibility of learning nonadjacent dependency between syllables" (e.g. p.581), and we devoted all the end of our paper to account for this effect within an associative view of mind.

Having replicated our results, instead of concluding that Peña et al.'s findings are not "invalid", our critics argue that they are at most small-to-medium size effects (Cohen's convention). However, their experiment is not suited to assess effect size. Testing three conditions within participants, and reducing the number of items as compared to Peña et al., it is bound to underestimate effect sizes. Thus Perruchet et al. report a preference for rule-words over class-words with a d of .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 where 20 participants were exclusively tested with 27 rule-word versus 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 large effect size (Cohen's $d = .98$).

It is likely that Endress's test overestimated effect size. Indeed, the number of different scrambled words is superior to the number of different words (when the number of words is > 2). Given that Pena et al. did not controlled for the frequency of words and part-words in their test, it is probable that Endress did not control for the frequency of words and scrambled words in her test. This may lead her to overestimate learning, because the correct items appear more often during the test than

the incorrect items. Responding on the basis of item familiarity, whether intentionally or not, artificially increases the discrimination.

This being said, even if Endress data were unbiased, her results do not undermine our conclusions. Indeed, our concern is not about the absolute value of the effect size, which may vary from one experiment to the other as a function of various factors. Our objective was to compare the effect size in the Pena et al.'s test with the effect size in the sound test of nonadjacent dependency, all things being equal otherwise (number of test items, and so on). We showed that Pena et al.'s test considerably overestimated the genuine learning effect (note that those comparisons were done in conditions that were carefully controlled for the frequency of correct and incorrect items).

To summarize, our critics argued that our data are invalid, yet they replicated them, including the phenomena whose existence they denied. Thus, Perruchet et al. begin by boldly stating that our results are deeply flawed, and end with a failed attempt at reassessing their effect size. At the same time, they claim that a theory based on simple and ubiquitous associative mechanisms can exactly (textual) accommodate them, but offer no hint as to how it would explain the generalization to novel items they also attested. Thus it remains a mystery how Perruchet et al. might explain the results of their own experiment, let alone ours.

The mystery may be easily cleared up by reading our paper. For instance, in p.581-582, we account for "the generalisation to novel items" within an associative theory of learning.

Statistics and speed of learning generalizations

So far we have shown that Perruchet et al.'s experiment 2 (barring its multiple methodological problems) replicates our findings that participants project generalizations after exposure to a short discontinuous speech stream. However, our critics neglect other crucial findings we reported, e.g., 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 Perruchet et al. also attested the existence of fast learning of structural relations, we wonder what alternative explanation they propose. We were expecting to see this issue discussed in a paper that promises to show that there is "no need" for the kind of computations Peña et al. proposed.

The authors are right on this point: Perruchet et al. did not discuss Pena et al.'s Experiment 4. This is now done in our reply, and we thank the authors for offering us this opportunity to complete our initial paper.

However, we only find two passing observations about how structural information can be quickly extracted. In the first attempt, Perruchet et al. argue that Peña et al.'s 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 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 and 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 is due to the forgetting of the middle element, Peña (2002, experiment 16) compared rule-words against actual words after 2 min of exposure to a segmented stream. As rule-words and words differ precisely in their middle element, had this gone unnoticed participants should show no preference. In fact, they preferred words, showing that the middle element is indeed encoded. Hence, our critics are wrong.

This experiment replicates a phenomenon that is well-documented in the implicit learning literature, called the "transfer decrement". This evidence does not contradict our claim that "participants pay more attention to [the] first and last syllables than to [the] middle one". It does not even contradict the claim that the middle syllable goes unnoticed during the familiarisation phase (although this stronger condition is not essential to our interpretation). Indeed, the X syllables used for the rule-words were in fact played during the familiarisation: They were the A and C syllables of other families of words. It is possible that participants have learned that these syllables marked the beginning and/or the end of words. To put the matter simply: Bonatti et al consider that participants selected the words over the rule-words because they knew the intervening syllables of the words. It is also possible that they rejected the rule words, because they knew that their intermediary syllables were rather at the beginning or the end of words.

In their second attempt, Perruchet et al. hint 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 two minutes 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. compare the ability of adults to rapidly extract structural information reported in Peña et al. with Saffran et al.'s (1996) infant experiments on word identification, as if they were the same thing. They seem to believe that since the exposure duration is identical in both studies, participants must be doing the same computations: if Saffran et al.'s infants learn by statistics, so must Peña et al.'s adults do.

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

Perruchet et al. also neglect that in Saffran et al. infants had to learn four words defined by high TPs among adjacent syllables. In the Peña et al.'s relevant experiments, adults had to extract nine words requiring the computation of TPs among non-adjacent syllables. The experience needed to extract 9 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. (2002) ran extensive pilot studies to assess the minimal exposure needed to identify such a lexicon, when defined by both adjacent and non adjacent TPs, and found that under about 10 min of exposure no segmentation occurs. Thus, comparing Peña et al.'s adult experiments with Saffran et al.'s infant experiments is meaningless: with 2 min of exposure to a continuous stream, a 6-9 word lexicon is not extracted.

We have never claimed that Saffran et al. measured the same thing that Pena et al., so the three paragraphs above are directed against a straw man. We just noted that statistical learning may proceed very quickly.

But even disregarding these misrepresentations, a substantive issue remains. We never denied that a theory based on statistical learning might account for fast learning.

Note, however, that quick learning is presented, in Pena et al, as the main argument for assuming algebraic like computations

However, the thesis that all learning can be explained by statistical computations is empty, unless our critics can propose a single mechanism capable of simultaneously explaining segmentation of words after exposure to a long (but not to a short) familiarization with a continuous stream, extraction of structural information after a short familiarization to a discontinuous stream, and failure to extract the same information after familiarization to any continuous stream. Without explaining the phenomena they so sharply criticize, Perruchet et al. claim that statistically based alternatives to Peña et al.'s dual system may exist, yet they provide no credible alternative.

We haven taken care, in our reply, of explaining all the phenomena described here. Of course, in this explanation, we have replaced what Pena et al. intended to measure by what they actually measured.

Generalization, Segmentation and Simulations

Perruchet et al. assert that our conclusions are flawed since their model of segmentation, Parser, simulates both Peña et al.'s experiment 1 and its frequency control (Peña et al., 2002, footnote 16). We leave it to a future paper 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 our experiments, their simulation of experiment 1 generalizes to all our data. However, by construction, Parser can only extract chunks that actually occurred in a stream. Since 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 like Parser fit all the data is groundless.

The point the authors raised is dealt with quite explicitly in the last two pages of our paper.

Second, Perruchet et al. misrepresent the purpose of Peña et al.'s 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 it -- in principle, it trivially can, since a chunking mechanism is computationally more powerful than a mechanism based on TPs. The

point is, 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 draw is the same 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. do, that nonadjacent TPs and chunking are alternatives. In all likelihood, a parsing algorithm will exploit either information to achieve segmentation. However, considerations of simplicity should invite our critics to take a more sober attitude towards 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 and Saffran et al.'s results on segmentation without need of adjusting parameters. In the absence of further evidence, preference should go for the simpler theory. Hence even granting the soundness of Perruchet et al.'s simulations, much remains to be done before concluding 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, they support the phenomenon we documented.

See our reply for a rebuttal of this argument. It is worth adding that the last two arguments of the authors are incompatible. Indeed, they claimed first that Parser is unable to capture nonadjacent dependency (they need to do so to strengthen their argument according to which an associative theory is unable to account for their data), and second that Parser must be able to track long distance relations (they need to do so, because, if Parser is actually unable to capture nonadjacent dependency, this would validate our argument according to which Parser's successful segmentation of their language is a proof for biases in their material).

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 may also play a salient role during language acquisition. Peña et al. proposed a series of empirical theses, backed up 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 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 to hit some unknown rocks. It is all the more important, then, to stay clear of the visible ones. In their critical review, Perruchet et al. 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 deny, and neglected data problematic for their theses. Despite their efforts, we found their attacks to dissolve after cursory examination. While undoubtedly the issues we raised in Peña et al. 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.

After having restated our main arguments in our reply, and addressed the other Bonatti et al.'s remarks in the text above, we let the reader make his/her own judgment about the Bonatti et al.'s conclusion.