Journal of Experimental Psychology:
Human Perception and Performance
1999, Vol. 25, No. 5, 1472–1475

0096-1523/99/$3.00

OBSERVATION

Baseline Conditions in Structural Induction: Comment on Pitt, Smith, and Klein (1998)

Athanassios Protopapas
Scientific Learning Corporation

Steven A. Finney and Peter D. Eimas
Brown University

M. A. Pitt, K. L. Smith, and J. M. Klein (1998) claimed that a baseline condition is necessary in experiments using the structural induction paradigm (S. A. Finney, A. Protopapas, & P. D. Eimas, 1996; C. Pallier, N. Sebastián-Gallés, T. Felguera, A. Christophe, & J. Mehler, 1993) to correctly interpret interaction effects. In this article it is shown that this constitutes a misunderstanding of the interpretation of statistical interaction and that a baseline condition offers little unless comparisons between cells of different target types within induction conditions are necessary.

In their recent article, Pitt, Smith, and Klein (1998) used the attentional allocation paradigm of Pallier, Sebastián-Gallés, Felguera, Christophe, and Mehler (1993) to explore the role of syllabic structure in the online perception of spoken English, as did Finney, Protopapas, and Eimas (1996). This paradigm is a variant of the phoneme monitoring task, in which an expected syllabic position for the phoneme target is created by manipulating the relative frequencies of items with different syllabic structures. That is, when the majority of target phonemes in an experimental session occur in syllabic codas, participants respond relatively faster to phonemes in syllabic codas than to phonemes in syllabic onsets, and the converse is true when the majority of target phonemes occur in syllabic onsets. Results using this paradigm (Pallier et al., 1993, for French and Spanish, and Finney et al., 1996, for English) have shown that listeners are sensitive to the syllabic structure of the experimental stimuli even in the absence of specific instructions mentioning syllables or obvious target properties that might turn their attention to syllable-sized units.

In their research, Pitt et al. (1998) used this paradigm to make a further contribution to the understanding of structural effects in on-line speech processing. They provided evidence consistent with the hypotheses that (a) medial consonants following stressed syllables (phonologically "ambisyllabic," according to Kahn, 1980) belong to both syllables, as opposed to being "unclearly syllabified" (as suggested by Cutler, Mehler, Norris, & Seguí, 1986), and (b) the effects of syllabic congruence, such as those reported earlier by Finney et al. (1996), were due to a match in syllabic structure rather than to a match in surface acoustics and phonetics produced by coarticulation. To be able to address these issues, Pitt et al. had to compare response times (RTs) across different items (words); for example, issues of ambisyllabicity could be examined only if RTs to words with ambisyllabic segments were contrasted with RTs to words of unique syllabification. This between-items comparison was possible only after baseline differences between items, determined by a neutral induction condition, had been removed.

However, Pitt et al. (1998) extended their claims further, stating that use of a baseline is necessary to determine the "true effects" (p. 1598) of induction using this paradigm. We argue against this claim on two bases, first by showing that it constitutes a misunderstanding of the statistical meaning of interaction and second by arguing that the usefulness of a baseline depends both on the precise question being asked and on the reliability of the baseline itself.

Interpretation of Interaction

Pitt et al. (1998) claimed that unless mean scores are normalized by neutral baseline data, the "statistically reliable interaction and means . . . inaccurately reflect the effect of the induction manipulation," and that "the interaction changes" (p. 1598) if existing baseline differences are eliminated. However, as has been repeatedly pointed out by Rosnow and Rosenthal (1989, 1991, 1995), the linear model that is the basis of the analyses of variance defines the interaction effect in a 2 × 2 analysis of variance as the residuals after the global mean and all main effects have been removed. Therefore, a graphical illustration of the values of group means (even "normalized" means) is not a plot of an interaction effect, nor can one judge the size or significance of an interaction effect by simply examining the slopes of lines joining group means within conditions across levels of another condition. Consider the top left panel in Figure 1, in which the results of Experiment 1 of Finney et

Athanassios Protopapas, Scientific Learning Corporation, Berkeley, California; Steven A. Finney and Peter D. Eimas, Department of Cognitive and Linguistic Sciences, Brown University.

Correspondence concerning this article should be addressed to Athanassios Protopapas, who is now at the Department of Psychology, Panteion University of Athens, Leoforos Syngrou 136, GR-176 71, Athens, Greece. Electronic mail may be sent to protopap@panteion.gr.

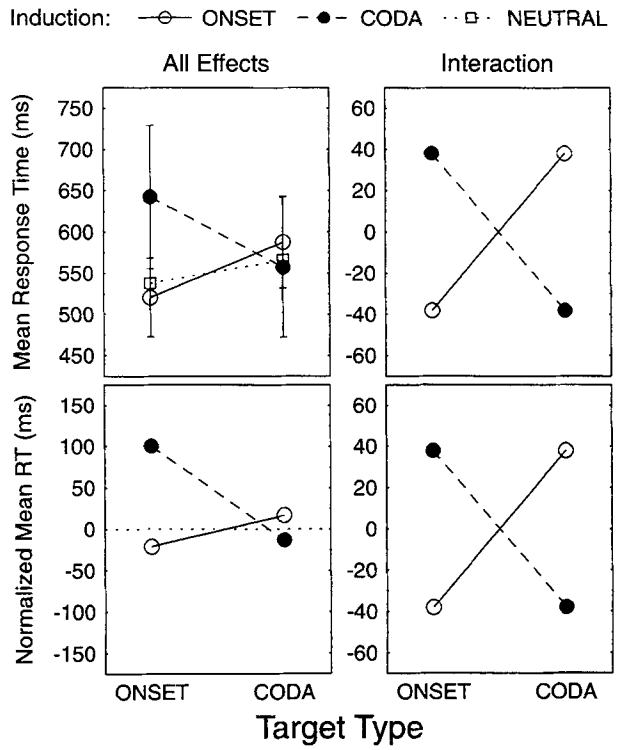

Figure 1. Top left: Response times (RTs) to phonemes in two target types (onset and coda) under three (between-subjects) induction conditions (onset, coda, and neutral). Top right: RTs in the onset and coda induction conditions after subtracting the global mean and main effects of each of the two factors. Bottom left: RTs in the onset and coda induction condition after subtracting RTs in the control condition as a method of normalization. Bottom right: Normalized RTs in the onset and coda induction conditions after subtracting the global mean and main effects of each of the two factors (from the bottom left panel).

al. (1996) are plotted (including a neutral baseline mentioned in their Footnote 1) as a function of two factors: induction condition (onset, coda, and neutral) and target type (onset and coda). An effect of structural congruence manifests itself in the relative differences between RT to each type of target as a function of induction condition, that is, an interaction between target type and induction type. If one considers RTs in the two main induction conditions only (onset and coda), a statistically significant interaction is obtained with both subjects and items as random factors, $F_1(1, 18) = 17.11, p = .001$; $F_2(1, 30) = 19.29, p < .0005$. Subtracting the global mean and main effects of each factor from the mean RTs in these two induction conditions (not including the neutral induction condition) results in the top right panel, in which only the interaction effect remains. Note that the interaction effect itself is perfectly symmetrical, as is necessarily the result in the 2×2 case with equal N (Rosnow & Rosenthal, 1995).

Now consider the bottom left panel of Figure 1, in which the mean RT to each item in the neutral induction condition has been subtracted from the mean RT to the same item in the onset and coda induction conditions; we refer to the resulting values as "normalized" RTs. This, Pitt et al. (1998) would argue, illustrates the "true effects of induction." Consider, however, what becomes of the statistical interaction in this process of normalization: In the bottom right panel, the normalized RT values have been stripped of the (normalized) global mean and main effects of the two factors (induction type and target type), and one is left with a pure graphical representation of the interaction. The fact that the two panels on the right side of Figure 1 are indistinguishable is hardly coincidental, for all we did by the baseline subtraction was to shift each pair of target-type group RTs vertically by the mean RT for that target type in the neutral condition. The relative differences between group means, and hence the interaction, were unchanged.[1] Thus, the use or nonuse of a baseline has no effect on the interaction itself at all.

Baseline Conditions

In the preceding discussion we have demonstrated that a baseline has no effect on the interpretation of the interaction that is the hallmark of a syllabic effect (and Pitt et al., 1998, in fact acknowledged that no more than a statistically significant interaction is necessary to claim that listeners are sensitive to syllable structure). What, then, are the possible uses of a baseline condition within this paradigm?

The classic use of a baseline (Pitt & Samuel, 1990; Posner & Snyder, 1975) is to serve as a standard against which to conduct analyses of facilitation and inhibition (alternatively, benefit and cost). In fact, Finney et al. (1996) ran a baseline condition for this purpose (see their Footnote 1), although the results did not appear in the final published version of their article. Those data are now shown in Figure 2. In the top row, the mean RTs of the five experiments from Finney et al. are plotted, along with the RTs in a neutral baseline condition similar to the baseline condition of Pitt et al. (1998). The bottom row contains normalized data (with the baseline value for each target type subtracted from the data points for that target type), where a negative value would indicate facilitation relative to baseline and a positive value inhibition. For Experiments 1–4, in which second-syllable stress items were used, it appears difficult to draw a consistent facilitation story from these data, even though the interactions themselves were consistent.

Specifically, the mean condition RT in the neutral condition is sometimes longer and sometimes shorter than mean RTs in the onset and coda induction conditions, although the main effect of induction condition was not statistically significant by subjects in any of these five experiments. One possible source for this baseline inconsistency is that it is a result of theoretically uninteresting between-groups differences. A second possibility is that the neutral baseline

[1]There was in fact a slight difference in the F value of the interaction in the subjects analysis between the unnormalized and the normalized case, normalized $F_1(1, 18) = 16.48$, $p = .001$, because missing data points contributed in different ways to averaging by subjects and by items given the item-by-item normalization that was used.

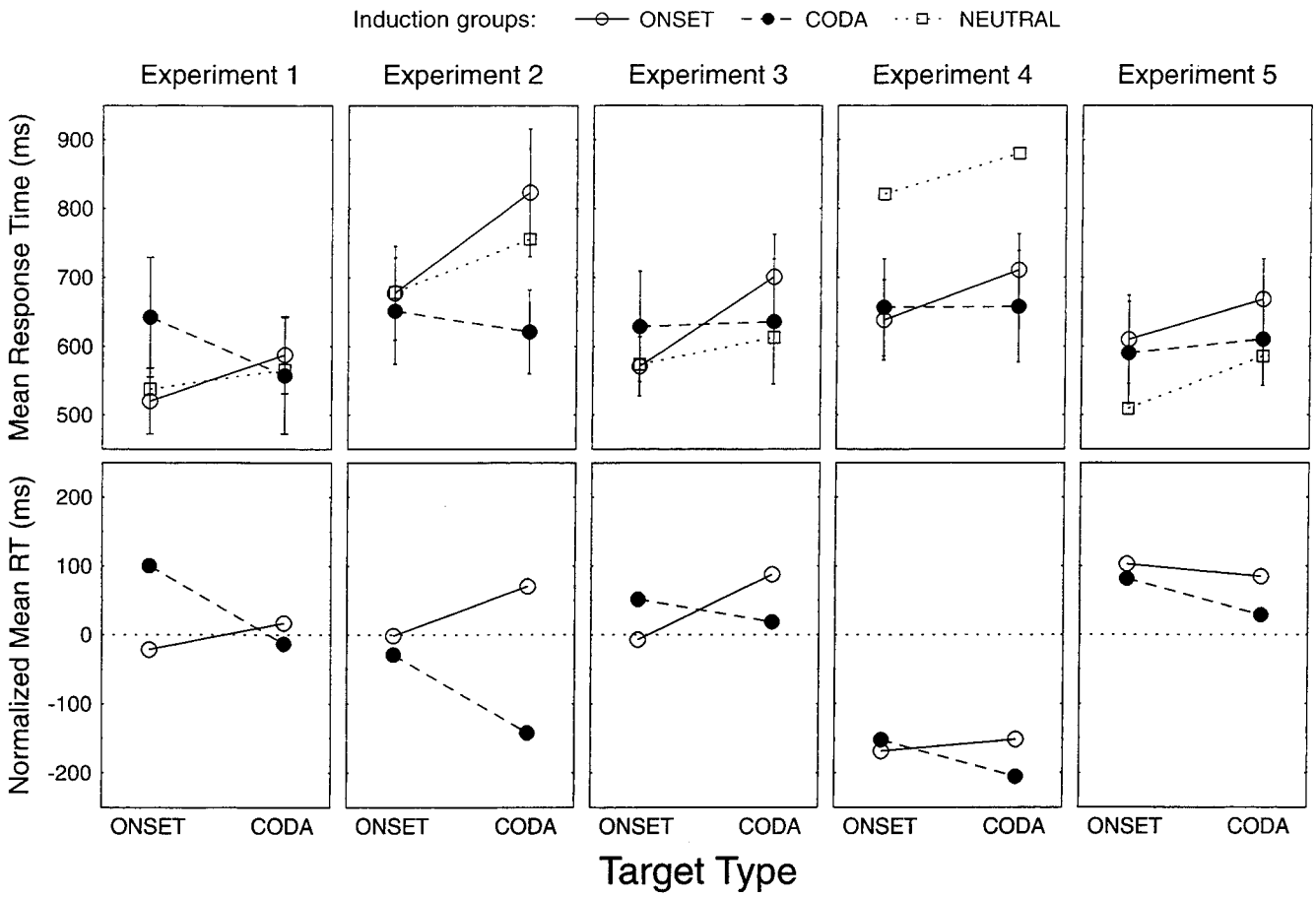

Figure 2. Mean response times (RTs; top row) and mean normalized RTs (by subtracting RTs in the neutral condition, bottom row) in the experiments of Finney, Protopapas, and Eimas (1996). Experiment 1: Replication of Pallier, Sebastián-Gallés, Felguera, Christophe, and Mehler (1993) using English materials. Experiment 2: Use of spliced items to equate the time from word onset to target. Experiment 3: Use of target words with intervocalic medial consonants while keeping inductor words with consonant clusters. Experiment 4: Use of pseudoword targets. Experiment 5: Use of first-syllable stress words (in Experiments 1–4 second-syllable stress words were used).

condition is not directly comparable to the induction conditions. As noted by Jonides and Mack (1984), care must be taken when baseline conditions are used to ensure that the "neutral" condition involves processing similar to that in the other experimental conditions. In this attentional allocation paradigm, the onset and coda induction conditions serve to induce expectations about the structural (syllabic) position of target phonemes, whereas the neutral condition includes equal proportions of the two types of inductor items (assumed not to lead to any consistent syllabic expectations). However, it is possible, for instance, that a preponderance of phoneme targets at a specific syllabic position might force attention to a syllabic level of processing in the induction conditions, whereas in the baseline condition participants might use either a phonemic or a syllabic strategy, leading to noise in the baseline data. Regardless of the source of the inconsistencies with the neutral condition of Finney et al. (1996), it makes results using such inconsistent baselines difficult to interpret.

A second reason to use a baseline is when it is desirable to perform statistical comparisons across different items that may have some important underlying differences, as Pitt et al. (1998) did in their analysis of the membership of allophonic consonants; a baseline is necessary in this case. As another example, a baseline is also necessary to interpret simple effects across different groups of items (e.g., in Experiment 1 of Finney et al., 1996, the difference between onset and coda items in the onset induction condition). Here Pitt et al. were correct in pointing out a weakness in Finney et al., as Finney et al. did provide statistical tests for such simple effects. These simple effects are meaningless without some compensation for confounding differences (acoustic or otherwise) between items that might lead to differences in RT in the absence of any induction. For example, in the baseline conditions for Experiments 1–5, phonemes in syllabic onsets were consistently identified faster than the same phonemes in syllabic codas, and this effect was statistically significant in Experiments 2, $F_1(1, 9) = 14.23$,

$p = .004$, and $F_2(1, 30) = 3.85$, $p = .059$, and 5, $F_1(1, 9) = 5.14$, $p = .050$, and $F_2(1, 30) = 9.13$, $p = .005$. Any interpretation of the effect of onset versus coda target type within an induction condition should take such differences into account. Crucially, however, none of the conclusions of Finney et al. was based on these simple effects.

Moreover, it is unclear what is gained by using a baseline in the 2 × 2 case for testing these simple effects. After normalization of RTs in th eexperiments illustrated in Figure 2 (bottom row), there was a significant effect of target type only within the onset induction condition in Experiment 3, $F_1(1, 9) = 6.95$, $p = .027$, and $F_2(1, 30) = 9.53$, $p = .004$, and within the coda induction condition in Experiment 2, $F_1(1, 9) = 29.54$, $p < .0005$, and $F_2(1, 30) = 13.50$, $p = .001$. Considering only the onset and coda target types and induction conditions, the results of Pitt et al. (1998) were also inconclusive (in terms of statistical reliability): The simple effect of target type was significant by both subjects and items only in the onset induction condition of their Experiments 1 and 2b and in the coda induction condition of Experiment 3. Again, the use of a baseline in and of itself does not guarantee consistent and interpretable results.

Conclusion

In conclusion, we disagree with Pitt et al. (1998) about the absolute "necessity" of a baseline condition in the structural induction paradigm, although certain specific questions (such as the ambisyllabicity and allophonic variation issues addressed by Pitt et al.) may require such a baseline. Furthermore, we wish to strongly reiterate that the interaction effect itself, when properly interpreted, is not affected by the use of a baseline.

References

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385–400.

Finney, S. A., Protopapas, A., & Eimas, P. D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language, 35,* 893–909.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44.

Kahn, D. (1980). *Syllable-based generalizations in English phonology.* New York: Garland.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32,* 373–389.

Pitt, M. A., & Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language, 29,* 611–632.

Pitt, M. A., Smith, K. L., & Klein, J. M. (1998). Syllabic effects in word processing: Evidence from the structural induction paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1596–1611.

Posner, M., & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. Rabbit & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). New York: Academic Press.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105,* 143–146.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110,* 574–576.

Rosnow, R. L., & Rosenthal, R. (1995). "Some things you learn aren't so": Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science, 6,* 3–9.

Received April 27, 1998
Revision received August 18, 1998
Accepted August 18, 1998 ■