

REJOINDER

Still No Phonological Typicality Effect on Word Reading Time (and No Good Explanation of One, Either): A Rejoinder to Farmer, Monaghan, Misyak, and Christiansen

Adrian Staub, Margaret Grant, and
Charles Clifton, Jr.
University of Massachusetts Amherst

Keith Rayner
University of California, San Diego

In this brief rejoinder, we respond to Farmer, Monaghan, Misyak, and Christiansen (2011). We argue that the data still do not support the claim that reading time is affected by the phonological typicality of a word for its part of speech. We also question Farmer et al.'s claim that interleaving syntactic structures in an experiment modifies grammatically based syntactic expectations.

In their reply to our article (Staub, Grant, Clifton, & Rayner, 2009), Farmer, Monaghan, Misyak, and Christiansen (2011) argued that prediction is important in sentence processing, that the phonological typicality of nouns and verbs can affect judgments about sentences, and, most centrally, that the phonological typicality of a word with a predicted part of speech has an immediate and substantial effect on online reading. The present debate is entirely about the last of these claims.¹ The initial report by Farmer, Christiansen, and Monaghan (2006) suggested that phonological atypicality slowed self-paced reading of nouns and verbs by nearly 50 ms. Because we considered an effect of this magnitude to be surprising, we conducted and reported two experiments (Staub et al., 2009), one using eye tracking and one using self-paced reading, in which we attempted to replicate the effect that Farmer et al. (2006) reported. In neither study did any hint of this effect appear.

In their article in the current issue of this journal, Farmer et al. (2011) argued that we did not obtain a phonological typicality effect in our experiments because we intermixed the critical noun and verb stimuli, in contrast to Farmer et al. (2006), who conducted separate experiments with nouns and with verbs. In our intermixed design, they suggested that “it is likely that subjects implicitly learn to recognize the structure shared between the noun and verb items . . . and that when such a word order is used, the main verb can be followed by either a noun or a verb structure” (p. XX). They then presented data from a new self-paced reading

experiment with an intermixed design. As in our experiments, they did not find evidence of a phonological typicality effect overall. However, they argued that there is evidence of an effect at the very beginning of the experiment, which decreases over time, consistent with their proposal that structure-based expectations for a noun or verb diminish over time when the two sentence types are intermixed.

We have two objections to this account. The first is simply that there is very little support for Farmer et al.'s main empirical claim, that is, that there is a phonological typicality effect at the beginning of their experiment with an intermixed design. The second is that even if the data were stronger, their hypothesis about the effect of an intermixed experimental design is puzzling. We take these points in order.

Lack of Empirical Support

As we pointed out in Staub et al. (2009), the original Farmer et al. (2006) article did not include any analyses with items as a random factor. In our (2009) article, we found no significant, or even marginal, effects of phonological typicality. In their new experiment, Farmer et al. (2011) found no overall effect of typicality.

Farmer et al. (2011) did report a significant three-way interaction among order, part of speech, and phonological classification. (We note that reanalysis of our 2009 experiments provided no hint of an interaction with order or of a typicality effect at the beginning of the experiments.) However, this was not due to a significant phonological typicality effect at the beginning of the experiment, where the effect is putatively the strongest. The only reported test that approached significance was a one-tailed *t* test

Adrian Staub and Charles Clifton, Jr., Department of Psychology, University of Massachusetts Amherst; Margaret Grant, Department of Linguistics, University of Massachusetts Amherst; Keith Rayner, Department of Psychology, University of California, San Diego.

Preparation of this article was supported by National Institute of Health Grants HD26765, HD065829, and HD18708.

Correspondence concerning this article should be addressed to Adrian Staub, 430 Tobin Hall, Department of Psychology, University of Massachusetts Amherst, Amherst, MA. 01003. E-mail: astaub@psych.umass.edu

¹ In particular, we do not dispute the other evidence presented and cited by Farmer and his colleagues showing that phonological typicality can bias judgments about a word and possibly even have early effects on its processing. Our objections are limited to the claims made about the time taken to read a word in the course of reading text.

($p = .07$) that assessed the typicality effect for the first three nouns. Note that with the standard two-tailed test, the p value would have been .14. We assume, by its omission, that the t test for the numerically smaller effect comparing the first three verbs was even further from significance. Thus, the interaction with order must, in fact, have been due in large part to the unexplained reversal at the end of the experiment (a reversal which, we submit, begins at the fourth item²).

In sum, there has yet to be a demonstration, in any experiment or in any part thereof, of an effect of the phonological typicality of a word on how fast the word is read that reaches, or even approaches, conventional standards of statistical significance. As usual, the burden of proof is carried by the argument against the null hypothesis. That burden is, in this case, made heavier by repeated null effects in well-controlled experiments.

Logic of the Argument

We think that the lack of compelling positive evidence for a phonological typicality effect on reading times, together with several null findings, should effectively end the debate. However, we also question Farmer et al.'s (2011) hypothesis that an intermixed design could weaken the phonological typicality effect over time.

In terms of the requirements of the grammar, a transitive verb followed by a determiner is as predictive of a noun in an intermixed design as in a separated design, and an intransitive verb followed by the word *to* is as predictive of a verb. All the experimental materials that Farmer et al. (2006) and Staub et al. (2009) used had these predictive qualities. Though Farmer et al. referred to the expectation of a noun or a verb at the critical point in each sentence as a "bias," in fact this is a categorical requirement of the grammar: A verb is an ungrammatical continuation of "The curious young boy saved the . . .", and a noun is an ungrammatical continuation of "The very old man attempted to . . ."

What Farmer et al. (2011) argued is that in an intermixed design, subjects' predictions of a noun or a verb are gradually less influenced by the grammatical requirements of the individual sentences they are reading, as they become sensitive to the fact that sometimes noun phrase-verb sequences in the experiment are continued with a noun and sometimes with a verb. As the experiment progresses, subjects base their predictions, at least in part, on the higher order statistics of the experiment itself. In other words, they are learning the word order contingencies that hold within the experiment, and this learning is weakening the role of the contingencies embodied in the grammar, though in fact these grammatical contingencies do hold throughout the experiment. They suggested that "at the onset of the experiment, the participant may be using the entire sentence frame to predict the category of the target word. By the end of the experiment, the participant has learned to disregard most of the frame as predictive of category" (p. XX). And because subjects are no longer clearly predicting nouns or verbs at the relevant points, the phonological typicality effect disappears over time.

We see two related difficulties with this account. The first pertains to ecological validity. In normal discourse, and normal reading, sentences that begin with a subject and a verb sometimes continue with a direct object, sometimes with a nonfinite clause, and sometimes with other structures in an interleaved (if not

altogether random) distribution. Thus, an environment in which, for example, a subject-verb sequence always continues with a nonfinite clause is a decidedly unnatural one. We would be hesitant to conclude that an effect that emerges only under these conditions is informative about normal language processing. For example, an apparent implication of the claim that this effect emerges only in "pure" designs is that researchers attempting to model word reading times in corpora do not need to consider phonological typicality as a predictor variable.

The second difficulty is a logical one. The (statistically unsupported) claim that readers predict part of speech at the beginning of an experimental session but not thereafter (by our estimate, not after the first three experimental sentences they see; see Footnote 2), assumes that they come into the experiment with strong grammatically based expectations about how each subject-verb sequence will continue, but when the experiment presents varying syntactic structures, these expectations quickly disappear. However, there is no reason to think that readers' pre-experimental exposure to the language differs substantially in syntactic variability from that found in the intermixed experiment. Farmer et al. remarked that "during the course of the interleaved experiment, the contextual expectations from natural language appear to be weakened" (p. XX). But if the intermixing of syntactic structures in the experiment causes subjects to reduce their reliance on the grammatical requirements of those structures (signaled by verb subcategorization and the following function word), why wouldn't the intermixing found in normal language use have the same effect, well before subjects enter into the experimental situation?

We submit that readers do not ignore the grammatical requirements of the sentences they read, either at the beginning of an experiment or at the end, and that they do use these grammatical requirements online to anticipate upcoming structures (see, e.g., Staub & Clifton, 2006; Staub, Clifton, & Frazier, 2006). However, we also submit that there is no good evidence that the time they take to read a word of a predicted part of speech is sensitive to this word's phonological typicality.

Conclusions

In sum, there is still no compelling evidence that there is an online phonological typicality effect in reading. Moreover, if there were evidence for such an effect restricted to experimental situations in which the part of speech of the content word following a NP-V sequence is completely predictable, this would not have meaningful implications for our understanding of normal language processing.

² Examination of Farmer et al.'s (2011) Figures 2 and 3 indicates that the size of the phonological typicality effect is substantially smaller when averaged over the first five experimental items that a subject saw than when averaged over just the first three. The decrease is so large that the phonological typicality effect for Items 4 and 5 has to be negative to counteract the (nonsignificant) positive effect for Items 1-3.

References

- Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of*

- the National Academy of Sciences of the United States of America*, 103, 12203–12208. doi:10.1073/pnas.0602173103
- Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either . . . or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 425–436. doi:10.1037/0278-7393.32.2.425
- Staub, A., Clifton, C., Jr., & Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language*, 54, 389–406. doi:10.1016/j.jml.2005.12.002
- Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 806–814. doi:10.1037/a0015123

Received January 24, 2011

Revision received April 6, 2011

Accepted April 13, 2011 ■