Response to Reviewers:

The relationship between the late posterior positivity/P600 in language comprehension and task demands

Gina Kuperberg and Trevor Brothers

There has been much discussion about the *late posterior positivity/P600* in relation to task demands and its relevance to 'natural' language comprehension. Here we discuss some of these issues. We first consider the question of whether an overt judgment task is necessary or sufficient to produce a *late posterior positivity/P600* effect. We then ask whether the study of ERP responses to anomalies, with or without a coherence judgment task, is relevant to understanding neurocognitive mechanisms engaged in 'natural' language comprehension.

We note that both these questions relate to a more general issue concerning the relationship between the *late posterior positivity/P600* evoked during language comprehension and the posteriorly distributed P300 component. We are currently writing a review about these relationships and plan to post this review on BioRxiv when available. This review will also include some of the points discussed below.

1. Is a task-relevant decision necessary or sufficient to produce *a late posterior positivity/P600* effect?

In many of the studies reporting a *late posterior positivity/P600* to semantic incongruities, comprehenders are asked to judge the coherence or acceptability of each scenario, typically after a short delay. As discussed by Kuperberg (2007), the presence of a coherence judgment task is one of several factors that can increase the likelihood observing a semantic P600. Additional factors include: 1) whether or not the semantic incongruity is actually anomalous, producing an impossible interpretation

(as opposed to simply an odd or unlikely interpretation), 2) whether the semantically incongruous word is semantically attracted or associated with the prior context, and 3) whether the semantically incongruous word appears in an extended linguistic context, particularly one that is highly lexically constraining. Importantly, as emphasized by Kuperberg, 2007, *none of these factors alone are necessary or sufficient to produce a late posterior positivity/P600 effect. "Rather, the P600…appears to be triggered at a particular threshold that can be influenced by some or all of these factors acting in consort."*

For a given stimulus (i.e. holding other factors constant), there is certainly evidence that the *late posterior positivity/P600* is enhanced when comprehenders are engaged in a coherence judgment task, relative to more passive reading tasks (e.g., Kolk, Chwilla, van Herten, & Oor, 2003; Xiang & Kuperberg, 2015). In addition, the *late posterior positivity/P600* is thought to be triggered when new bottom-up input conflicts with the comprehender's high-level mental model, and the detection of conflict may well be linked to decision-making mechanisms (for further discussion, see Kuperberg & Brothers, in preparation; Kuperberg, Brothers, & Wlotko, in press, and Brothers, Wlotko, Warnke & Kuperberg, under review). This, however, does not imply that the effect is dependent on participants being *explicitly* instructed to carry out a coherence judgment; nor does it mean that a *late posterior positivity/P600* is triggered by all types of binary decisions about sentences. Indeed, a *late posterior positivity/P600* can be produced without explicit judgment tasks, and not all task-relevant decisions to semantic incongruities are associated with a *late posterior positivity/P600*. Consider the following pieces of evidence:

a) A *late posterior positivity/P600* can be produced by semantic anomalies, even when comprehenders do not carry out a coherence judgment task (e.g. Nieuwland & Van Berkum, 2005; Nakano, Saron &

Swaab, 2010; van de Meerendonk, Kolk, Vissers & Chwilla, 2010; DeLong, Quante & Kutas, 2014; Szewczyk & Schriefers, 2011; Munte, Heinze, Matzke, Wieringa & Johannes, 1998; Quante, Bolte & Zwitserlood, 2018; Pijnaker, et al., 2010). This is also true of the original stimulus set that reported a *late posterior positivity/P600* to semantic anomalies (e.g. "Every morning at breakfast the eggs would *eat…*unpublished findings*).

b) A *late posterior positivity/P600* is *not* typically produced by words that are implausible-but-not-impossible, *even when* participants carry out a coherence judgment task (e.g. Kuperberg et al., 2003; Kuperberg et al., 2007; Paczynski & Kuperberg, 2012). For example, whereas a robust *late posterior positivity/P600* is produced by "plant" in "Every morning at breakfast the eggs would plant…", no such effect is produced by "plant" in "Every morning at breakfast the boys would plant…", even though participants correctly judge these sentences to be implausible (e.g. Kuperberg et al., 2007).

c) A *late posterior positivity/P600* effect is *not* produced by outright semantic anomalies when the prior context is minimal, even when participants carry out a judgment task and correctly categorize these continuations as anomalous. For example, no *late posterior positivity/P600* is produced by semantic anomalies in very short sentences like "James unlocked the *gardener…", relative to "James unlocked the door" (Brothers, Wlotko, Warnke & Kuperberg, under review; see also Gunter, Friederici, & Hahne, 1999; Osterhout & Nicol, 1999).

d) A *late posterior positivity/P600* is *not* always produced when comprehenders are asked to make other types of binary task-relevant categorizations, such as judging whether or not they correctly predicted the final word of a sentence (Brothers, Swaab & Traxler, 2015; Dave et al., 2018).

Taken together, these findings suggest that carrying out an explicit coherence judgment task is neither necessary nor sufficient to generate a *late posterior positivity/P600* during language comprehension. We discuss the reasons for this in detail in Kuperberg & Brothers (in preparation), and in Brothers, Wlotko, Warnke & Kuperberg (under review).

2) Is the study of *late posterior positivity/P600* to semantic anomalies relevant to understanding neurocognitive mechanisms engaged in natural language comprehension?

Some have argued that studying ERP the study of responses to semantic incongruities and anomalies, particularly in combination with a coherence judgment task, can tell use little about the processes underlying 'natural' language comprehension. This criticism subsumes two sub-arguments, which we consider below.

a) *Semantic anomalies do not appear in natural language and so studying the neurocognitive mechanisms that they evoke is irrelevant to natural language comprehension.*

A response that is often given to this critique is that, by studying responses to anomalies, we are pushing the language processing system to its limits, enabling us to isolate mechanisms that are engaged during 'normal' language comprehension. This is certainly a valid argument. For example, ERP responses to syntactic anomalies can be similar to responses to syntactically dispreferred continuations in garden path sentences (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; 1993). However, perhaps an even more relevant argument is that comprehenders do, in fact, encounter anomalies during everyday communication. These include speech errors generated during language production (Dell, 1995), errors made by non-native speakers, and errors in visual or auditory perception on the part of the comprehender (Levy, Bicknell, Slattery & Rayner, 2009). There is also clear evidence

that readers are able to notice these incongruities and anomalies during online comprehension, even when they are not explicitly instructed to monitor for errors, resulting in longer reading times (e.g. Albrecht & O'Brien, 1993; Levy et al., 2009), and, as noted above, to an enhanced *late posterior positivity/P600*. Obviously, these sorts of errors are encountered less frequently in day-to-day comprehension than in psycholinguistic experiments that are specifically designed to examine their consequences. Nonetheless, we still believe that these experiments can tell us something relevant about how the brain identifies, recovers and learns from these errors "in the wild".

b) *Measuring responses to anomalies when comprehenders engage in a coherence judgment task is irrelevant to understanding responses to anomalies during natural language comprehension.*

If one accepts that studying neural responses to anomalies is relevant to natural language comprehension, the question then arises of whether, in an experimental setting, it is appropriate to study these responses while asking comprehenders to explicitly monitor for coherence, i.e. make acceptability judgments after each sentence. To our mind, this depends on the goal of a particular study. If the goal is to understand the neurocognitive mechanisms engaged in response to semantic anomalies, then a coherence judgment task can provide an important control — it ensures that participants are motivated, engaged in deep comprehension, and that they are capable of detecting these anomalies on the majority of trials. Without such a task, an absence of an anomaly response could be attributed to shallow processing. Indeed, there is evidence that no *late posterior positivity/P600* is produced when comprehenders fail to detect semantic anomalies (Sanford, Leuthold, Bohan, & Sanford, 2011).

It is also worth considering the alternative to asking participants to make coherence judgments. One task that that is often used in ERP studies, and that is often assumed to be more 'naturalistic' than coherence monitoring, is to ask participants to simply comprehend each sentence, and to answer

intermittent comprehension questions. This approach is certainly appropriate with some designs. However, when a participant is asked to read hundreds of short unrelated sentences with 50% of them containing anomalies, with this type of task, they may quickly learn to ignore anomalies altogether, particularly if the comprehension questions have nothing to do with their content. Another approach, which might be more analogous to our real-world experience with anomalies, would be to use a more implicit comprehension task, but to introduce anomalies in only a small proportion of stimuli. However, this would be subject to other criticisms, such as an imbalanced experimental design.

Finally, we note that implicit coherence monitoring is thought to be an important component of successful reading comprehension (see Cain, 2016; Garner, 1980; Wagoner, 1983), and that tasks requiring readers to monitor for errors or inconsistencies are often used to assess deep comprehension. Poor readers tend to be worse at detecting inconsistencies or anomalies than good readers (e.g. Rubman & Salatas Waters, 2000; see also Albrecht & O'Brien, 1993), probably because, without monitoring for understanding, they are less aware of any failures to achieve coherence, and, even if they are aware, they may be less able to compensate for such failures (Hacker, 1998).

The bottom line is that there is no 'perfect' experimental task that mirrors 'natural' language processing. Our view is that language processing is *always* constrained by our communicative goals, and that these goals will always influence the mechanisms engaged in comprehension — often qualitatively. If we are interested in isolating particular mechanisms of comprehension in the lab, then it is important to give comprehenders tasks that are most likely to capture these mechanisms and interpret our data in this light. For this reason, if we are interested in understanding the precise neurocognitive mechanisms engaged in response to semantic anomalies during deep comprehension, we think that it is sometimes appropriate to ensure that participants are comprehending deeply, and actually detecting these anomalies, by asking them to monitor for coherence.

Bibliography

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1061-1070.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition, 136*, 135-149.

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (under review). Going the extra mile: Effects of discourse context on two late positivities during language comprehension.

Cain, K. (2016). Reading comprehension development and difficulties: an overview. *Perspectives on Language and Literacy, 42*(2), 9.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia, 117*, 135-147.

Dell, G. (1995). Speaking and misspeaking. In L. R. Gleitman & M. Liberman (Eds.), *Language: An Invitation to Cognitive Science* (Vol. 1, pp. 183-208). Cambridge, MA: MIT Press.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61C*, 150-162.

Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior, 12*(1), 55-63.

Gunter, T. C., Friederici, A. D., & Hahne, A. (1999). Brain responses during sentence reading: visual input affects central processes. *Neuroreport, 10*(15), 3175-3178.

Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 165-191). Mahwah, NJ: Lawrence Erlbaum Associates.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang Cogn Process, 8*(4), 439-483.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang, 85*(1), 1-36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res, 1146*, 23-49.

Kuperberg, G. R., & Brothers, T. (in preparation). What can the P300 tell us about the P600? Understanding language comprehension within a decision theoretic framework.

Kuperberg, G. R., Brothers, T., & Wlotko, E. (in press). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain Lang, 100*(3), 223-237.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Brain Res Cogn Brain Res, 17*(1), 117-129.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, USA, 106*(50), 21086-21090.

Munte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22*(12), 2886-2898.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research, 24*(3), 691-701.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J Mem Lang, 31*(6), 785-806.

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. In S. M. Garnsey (Ed.), *Language and Cognitive Processes. Special Issue: Event-related brain potentials in the study of language* (Vol. 8 (4), pp. 413-437). Hove: Lawrence Erlbaum Associates.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence and time course of the brain responses to syntactic and semantic anomalies. *Lang Cogn Process, 14*, 283-317.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *J Mem Lang, 67*(4), 426-448.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: an ERP study on context sensitivity in autism. *Neuropsychologia, 48*(10), 2940-2951.

Quante, L., Bolte, J., & Zwitserlood, P. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: evidence from late-positivity ERPs. *PeerJ, 6*, e5717.

Rubman, C. N., & Salatas Waters, H. (2000). A, B seeing: The role of constructive processes in children's comprehension monitoring. *J Educ Psychol, 92*(3), 503-514.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: an ERP study. *J Cogn Neurosci, 23*, 514-523.

Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Res, 1368*, 208-221.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring language perception: mild and strong conflicts elicit different ERP patterns. *J Cogn Neurosci, 22*(1), 67-82.

Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Read Res Q*, 328-346.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Lang Cogn Neurosci, 30*(6), 648-672.