Maximising ERP resources using a sequential Bayes factor approach to sample size

Kate Stone[1], Shravan Vasishth[1], Frank Rösler[2], and Bruno Nicenboim[3]
[1]University of Potsdam, [2]Hamburg University, [3]Tilburg University | stone@uni-potsdam.de

Event-related potential (ERP) data is notorious for its low signal-to-noise ratio, requiring large sample sizes to detect evidence for experimental effects [1]. But even relatively large sample sizes—of a size infeasible for many labs—have failed to detect effects that may be genuine but small [2,3,4, 5 Exp. 1; cf. 5 meta-analysis]. How can researchers without the resources to recruit hundreds of ERP participants still contribute to knowledge about an empirical effect? We present a pre-registered sequential Bayes factor design in which resources are maximised by recruiting participants only until either a pre-defined Bayes factor or pre-defined sample size limit is reached [6,7]. 'Peeking' at the data during recruitment is possible under this approach because the Bayes factor quantifies the strength of evidence for competing hypotheses and its interpretation is not affected by how often one peeks [8,9]. Note that such an approach would not be possible with a frequentist analysis, where the probability of falsely rejecting the null increases with every peek that results in further data collection (i.e. if $p > 0.05$), and becomes almost certain if one peeks often enough [10,11].

We illustrate the approach with an experiment concerning the effect of contextual constraint on ERP. Contextual constraint is thought to determine how strong a reader's probabilistic expectations about an event are, but its effects have been difficult to demonstrate in ERP. Manipulating constraint is reported *not* to influence the amplitude of the N400 between two target words, so long as their cloze probability is matched [12,13,14], but *is* reported to affect the amplitude of the anterior post-N400 positivity (PNP) [13,14]. However, several studies have found no or opposing effects of constraint on the PNP [15,16,17,18]. Since compelling evidence for a constraint modulation of the PNP would greatly increase ERP researchers' ability to investigate readers' probabilistic expectations, we aim to replicate the dissociated effect of constraint on the N400 and PNP.

Pre-registered method. The key analysis concerns ERP amplitude at a low predictable noun in strong vs. weak constraint (a vs. b, Table 1). Constraint is operationalised as entropy of the target word (informally, uncertainty about the word's identity according to a cloze test) and predictability as its log cloze probability. Bayesian linear mixed effects models will be fit to the dependent variables mean amplitude in the window 250-500 ms across posterior electrodes (N400) and, separately, in the window 600-1000 ms across anterior electrodes (PNP). Participants will be recruited until either the Bayes factor is at least 10 times in favour of the null or alternative hypothesis for each model, or the limit of our resources is reached (defined as 150 participants). Evidence for no effect of entropy (N400) or a negative effect (PNP) will be assessed using Bayes factors with truncated priors reflecting our hypotheses that amplitude will change with 95% probability by between 0 and -0.4μV as entropy decreases (constraint increases). We will also evaluate the Bayes factor under a range of psycholinguistically plausible prior standard deviations. This protocol has been accepted as a Stage 1 registered report and data collection has begun.

Expected results and conclusions. If we obtain strong evidence for/against an effect of constraint on the N400 and PNP before reaching the 150-participant cap, we will have maximised our resources while contributing knowledge about the utility of the N400/PNP in investigating probabilistic expectations. Alternatively, if we reach the cap without finding strong evidence—that is, a BF that is not strongly in favour of or is even equivocal about an effect of constraint—the results will still be informative: First, the BF provides a continuous scale that can be interpreted even if one does not reach the desired (preregistered) threshold; second, even if the current sample size is unable to yield compelling evidence for or against the effect, then future studies will need to plan for even larger sample sizes, and/or design experiments that elicit larger effect sizes. Crucially, quantifying evidence in this way means that studies with any sample size cap (small or large) can contribute knowledge, including that a particular effect may be difficult to detect with a given sample size and study design [7].

Table 1. Example item showing cloze probability and entropy of the target noun in conditions (a) and (b). Note that log cloze probability and entropy are used as continuous predictors in the analysis. The stimulus in each condition comprises two sentences: the first is read in chunks, and the second, word-by-word (RSVP). For RSVP, each word is presented for 190 ms each plus 20 ms for each letter except the target noun, which is presented for 700 ms. Stimulus onset asynchrony (SOA) is 300 ms.

Context	Target noun	Spillover	Cloze (%)	Entropy (bits)
(a) Strong constraint, low predictable noun				
Auf Annetts Terrasse schien im Sommer zu viel *On Annett's terrace shone in summer too much* Sonne, um noch draußen sitzen zu können. Daher *sun in order outside sit to be able. Therefore* kaufte sie sich einen großen *bought she herself a large* *On Annett's terrace there was too much sun in summer for sitting outside. She therefore bought herself a large **hat** and...*	Hut *hat*	und… *and...*	5 [3, 18]	1 [0, 2]
(b) Weak constraint, low predictable noun				
Annett mag es gerne gemütlich, wenn sie etwas Zeit *Annett likes it really cozy when she some time* für sich findet. Daher kaufte sie sich *for herself finds. Therefore bought she herself* einen großen *a large* *Annett likes to make herself cozy when she finds a moment to herself. She therefore bought herself a large **hat** and…*	Hut *hat*	und… *and...*	5 [4, 15]	2 [1, 3]

References. [1] Luck & Gaspelin (2016) *Psychophysiology*; [2] Nieuwland et al. (2018) *eLife*; [3] Ito, Martin & Nieuwland (2016) *Lang, Cog, Neurosci*; [4] Kochari & Flecken (2019) *Lang, Cog, Neurosci*; [5] Nicenboim, Vasishth & Rösler (2020) *Neuropsychologia*; [6] Schönbrodt et al. (2015) *Psych Methods*; [7] Schönbrodt & Wagenmakers (2018) *Psych Bull Rev*; [8] Edwards, Lindman & Savage (1963) *Psych Rev*; [9] Rouder (2014) *Psychon Bull Rev*; [10] Armitage, McPherson & Row (1969) *J Royal Stat Soc*; [11] Simmons, Nelson & Simonsohn (2011) *Psych Science*; [12] Kutas & Hillyard (1984) *Nature*; [13] Federmeier et al. (2007) *Brain Res*; [14] Kuperberg, Brothers & Wlotko (2020) *J Cog Neurosci*; [15] Federmeier & Kutas (1999) *JML*; [16] Thornill & Van Petten (2012) *Int J Psychophys*; [17] Wlotko & Federmeier (2007) *Neuropsychologia*; [18] Szewczyk & Schriefers (2013) *JML*.