Computing the Bayes Factor for the Ruhnau et al. (2012) data

Wiens, Szychowska, Eklund, & van Berlekom (2018)

The main results in the Ruhnau et al. (2012) study were as follows (see their Figure 3): For frontal-central electrodes, the mean amplitudes in the MMN-relevant interval (between 140 and 190 ms) were 0.30 μV to the control tone in the cascade condition, 0.17 μV to the control tone in the no-repetition condition, and −0.68 μV to the deviant in the oddball condition. The difference score between deviant and the control tone (i.e., corrected MMN) was numerically more negative for the cascade (−0.68 minus 0.30 = −0.98) than for the no-repetition condition (−0.68 minus 0.17 = −0.85). However, this difference (0.13 μV) between cascade and no-repetition condition was reported as not significant.

The mean N1 amplitudes (between 90 and 120 ms) were 0 μV in the cascade condition and −0.71 μV in the no-repetition condition. The N1 was significantly smaller (i.e., amplitudes were less negative) in the cascade than in the no-repetition condition; cascade minus random = 0.71 μV.

In their study, Ruhnau et al. (2012) relied on null hypothesis significance testing. Although null hypothesis testing may provide evidence against the null (with a small *p*), a large *p* cannot provide evidence for the null hypothesis because non-significance could simply result from low statistical power. Furthermore, the *p* value does not provide a relative measure of how much support there is for the alternative hypothesis versus the null hypothesis (Dienes, 2008, 2016). Accordingly, the non-significant difference of the mean amplitudes in the MMN-relevant interval between the cascade and no-repetition conditions is uninformative about whether the results suggest a difference between conditions, no difference between conditions, or that the results are inconclusive.

In contrast, computing the Bayes Factor (BF) provides exactly that information (Dienes, 2016; Wagenmakers, Marsman, et al., 2017; Wiens & Nilsson, 2017). It expresses the likelihood of the data given the alternative hypothesis (i.e., theoretical predictions) relative to the likelihood of the data given the null hypothesis (Dienes, 2016; Wagenmakers, Marsman, et al., 2017; Wiens & Nilsson, 2017). Simply put, it captures how much better the data are explained by the alternative hypothesis versus the null hypothesis. The better the data are explained by one rather than the other hypothesis, the more evidence there is in support of this hypothesis. For example, a $BF_{10} = 3$ means that there is three times more evidence for the alternative than the null hypothesis, whereas a $BF_{01} = 3$ means that there is three times more evidence for the null than the alternative hypothesis. Note that if one knows either BF_{10} or BF_{01}, then the other can be derived with $BF_{01} = 1 / BF_{10}$.

Because Ruhnau and his colleagues provided us with their preprocessed data (i.e., mean amplitudes for each condition and subject), we could conduct secondary analyses to perform Bayesian hypothesis testing. The BF was calculated with *Aladins Bayes Factor in R* (Wiens, 2017). These scripts compute and plot the BF for mean differences in raw units if the alternative hypothesis is modelled as a normal, *t*, or uniform distribution, and the likelihood is modelled as a normal or *t* distribution (Dienes & Mclatchie, 2017).

As described below, results provided strong evidence ($BF_{10} > 40$) that the cascade condition reduces the N1 relative to the no-repetition condition but provided only anecdotal evidence for no differences between conditions ($1 < BF_{01} < 3$) in the MMN-relevant interval.

N1

To compute the BF for the N1, mean amplitudes were extracted to the deviant in the oddball condition (0.12 μV) and the control tones in the cascade condition (0 μV) and the no-repetition condition (−0.71 μV). The null hypothesis was that the amplitude difference of cascade minus no-repetition condition would be zero. In contrast, a positive difference score would support the hypothesis that the N1 was smaller (i.e., less negative) in the cascade than no-repetition condition. The actual mean difference was 0.71 μV (0 minus −0.71). Further, the

N1 should be smaller to the deviant than to the control tones. Because in the oddball condition, deviant and standard tones are tonotopically close, the N1 to the deviant should be decreased because of neural adaptation. In contrast, because in the cascade and no-repetition conditions, the tones within each condition are tonotopically further away from each other, the N1 should be less decreased because of neural adaptation. As a consequence, the N1 difference between cascade and no-repetition should not be larger than the difference between deviant and no-repetition. Because a small N1 to the deviant minus a large N1 to the no-repetition control should give a large positive value, the alternative hypothesis was modeled as a uniform distribution with the lower limit defined as zero and the upper limit defined as the N1 difference of deviant minus no-repetition control (0.12 minus −0.71 = 0.83 µV). As shown in Figure 2, the BF_{10} = 40.26 indicated that the N1 difference between cascade and random conditions is 40 times more likely under the alternative hypothesis than under the null hypothesis. According to a classification scheme (Wagenmakers, Love, et al., 2017), the BF is very strong evidence that the N1 is smaller in the cascade than random condition.

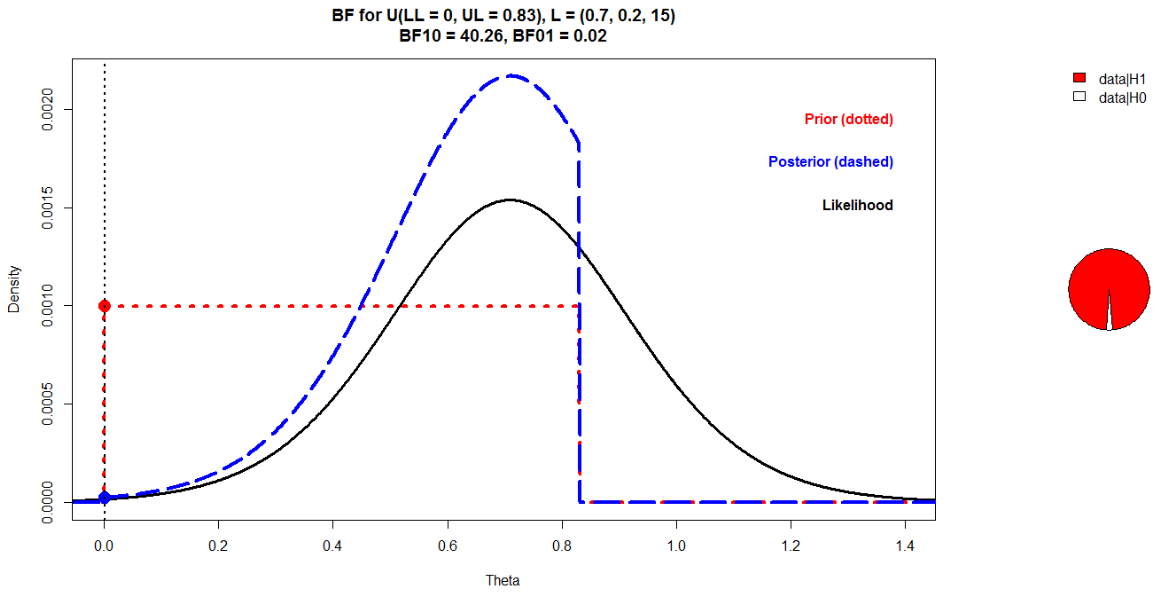

Figure 2. Plot of the Bayes Factor (BF) for N1.

Because in principle, the BF compares two models with each other, we conducted additional analyses with other alternative hypotheses to assess the robustness of the findings. First, although the deviant should elicit a small N1 due to neural adaptation, it may be biased towards a larger N1 by an early effect from a true MMN. If so, the N1 to the deviant would be larger (i.e., more negative) than it should be. Although the size of this bias on the deviant cannot be estimated easily, the standard is completely unbiased by a true MMN. Therefore, we defined the upper limit as the N1 difference of standard minus no-repetition control (0.71 minus −0.71 = 1.42 μV). The BF_{10} = 33.00 provides very strong evidence for a difference in N1. Second, the frequency steps between consecutive tones are more variable in the no-repetition than cascade condition. Because effects of this frequency variability on the N1 are unclear, we performed a two-tailed analysis in which the upper limit was defined as standard minus no-repetition control (1.42 μV) and the lower limit as its inverse (−1.42 μV). The BF_{10} = 16.55 provides strong evidence for N1 differences between cascade and no-repetition conditions.

MMN

To compute the BF for the MMN, the MMN-relevant amplitudes were extracted to the deviant (mean = −0.68 μV) and standard (0.32 μV) in the oddball condition, and to the control tones in the cascade (0.30 μV) and no-repetition (0.17 μV) conditions (see Figure 3 in Ruhnau et al., 2012). The null hypothesis was that the cascade-corrected MMN (deviant minus cascade) would be same as the no-repetition-corrected MMN (deviant minus no-repetition). However, if the cascade condition improves the measurement of the corrected MMN (deviant minus control) and thus, the cascade-corrected MMN is larger (i.e., more negative) than the no-repetition-corrected MMN, computing the difference of no-repetition-corrected MMN minus cascade-corrected MMN should result in a positive score. That is, the less negative value for the no-repetition-corrected MMN minus the more negative value for the cascade-corrected MMN should result in a positive score. This can be expressed as follows: (deviant minus no-repetition) minus (deviant minus cascade). However, this computation can be simplified to cascade minus no-repetition. In terms of actual numbers, (−0.68 minus 0.17) minus (−0.68 minus 0.30) = −0.85 minus −0.98 = 0.13; or simply 0.30 minus 0.17 = 0.13. Accordingly, the more positive the difference score, the more it supports the idea that the cascade-corrected MMN is larger (i.e., more negative) than the no-repetition-corrected MMN. However, this difference between conditions should be no larger than the oddball MMN (deviant minus standard, i.e., −0.68 minus 0.32 = −1.00). Therefore, the alternative hypothesis was modeled as a uniform distribution with the lower limit defined as zero and the upper limit defined as the absolute size (i.e., +1.00 µV) of the oddball MMN (to obtain a positive upper limit).

Figure 1 shows a plot of the BF for MMN. It illustrates the prior (i.e., the alternative hypothesis), likelihood, and posterior. A positive theta (in µV) reflects a larger corrected MMN in the cascade than no-repetition conditions. The BF_{01} = 2.17 provides only anecdotal evidence for no differences between the conditions. This is also shown in the pie chart, as the size of the white area is only slightly larger than that of the red area.

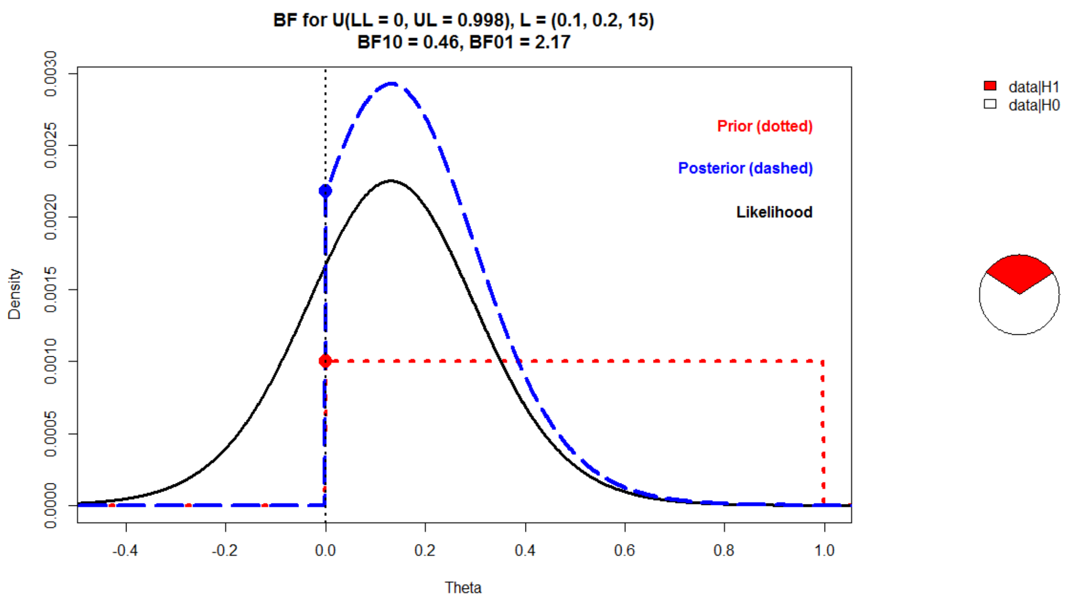

Figure 1. Plot of the Bayes Factor (BF) for MMN.

To assess the robustness of these findings, we conducted additional analyses with other alternative hypotheses. First, one could argue that the oddball MMN may overestimate the upper limit of the true MMN, as the standard is biased towards positivity due to neural adaptation. Unfortunately, the true MMN is unknown, and the cascade-corrected MMN probably underestimates the true MMN. Nonetheless, if the cascade-corrected MMN is used instead to define the upper limit, the BF was unaffected, $BF_{01} = 2.13$. Second, one may argue that there is no strong theoretical reason to think that the corrected MMN should be stronger for the cascade than the no-repetition rule, and effects may actually be reversed. To conduct a two-tailed analyses, the alternative hypothesis was modeled to range from −1.00 to + 1.00, which is the size of the oddball MMN. With this wide alternative hypothesis, the $BF_{01} = 3.33$ provides moderate evidence for similar effects of the cascade and no-repetition rule on the corrected MMN.

References

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York: Palgrave Macmillan.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78-89. doi: 10.1016/j.jmp.2015.10.003

Dienes, Z., & Mclatchie, N. (2017). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, 1-12. doi: 10.3758/s13423-017-1266-z

Ruhnau, P., Herrmann, B., & Schröger, E. (2012). Finding the right control: The mismatch negativity under investigation. *Clinical Neurophysiology, 123*(3), 507-512. doi: 10.1016/j.clinph.2011.07.035

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., . . . Gronau, Q. F. (2017). Bayesian Inference for Psychology. Part II: Example Applications with JASP. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-017-1323-7

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., Love, J., . . . more, a. (2017). Bayesian Inference for Psychology. Part I: Theoretical Advantages and Practical Ramifications. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-017-1343-3

Wiens, S. (2017). Aladins Bayes Factor in R. figshare. Retrieved from 10.17045/sthlmuni.4981154.v1

Wiens, S., & Nilsson, M. E. (2017). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement, 77*(4), 690-715. doi: 10.1177/0013164416668950