Re-analysing the data from Moffatt et al. (2020): What can we learn from an under-powered absence of difference?

Abstract

Moffatt et al. (2020) reported the results of an experiment (N = 26 in the final sample) comparing the facial electromyographic correlates of mental rumination and distraction, following an experimentally induced stressor. Based on the absence of significant difference in the perioral muscular activity between the rumination and distraction conditions, Moffatt et al. (2020) concluded that *self-reported* inner experience was unrelated to peripheral muscular activity as assessed using surface electromyography. We suggest this conclusion is at best hasty. Indeed, concluding on the absence of an effect based on an under-powered non-significant *p*-value is strongly uninformative. In this short commentary, we show that there is limited evidence for the main conclusion put forward by Moffatt et al. (2020). We suggest ways forward, both from a theoretical and methodological perspective. Complete source code, reproducible analyses, and figures are available at https://osf.io/ba3gk/.

Keywords: NHST, Bayes factor, reanalysis, inner speech, rumination, electromyography

Wordcount (excluding abstract, references, tables, and figures): 3168

Introduction

The activity of silently talking to oneself or "inner speech" is a foundational ability, allowing oneself to remember, plan, self-motivate, or self-regulate (for reviews, see Alderson-Day & Fernyhough, 2015; Lœvenbruck et al., 2018; Perrone-Bertolotti et al., 2014). However, whereas the use of inner speech is associated with many adaptive functions in everyday life, inner speech dysfunctions can be identified in multiple psychological disorders. For instance, rumination, broadly defined as unconstructive repetitive thinking about past events and current mood states (Martin & Tesser, 1996), is involved in the onset and maintenance of serious mental disorders such as depression, anxiety, eating disorders, or substance abuse (for review, see Nolen-Hoeksema et al., 2008).

Given the predominantly verbal nature of rumination (e.g., Ehring & Watkins, 2008; Goldwin et al., 2013; Goldwin & Behar, 2012; McLaughlin et al., 2007), we previously proposed to consider rumination as a form of inner speech and to study it using the methods that have historically been used to study other forms of inner speech, namely, by using surface electromyography (EMG) and motor interference protocols (e.g., Nalborczyk, Banjac, et al., 2021; Nalborczyk et al., 2017, 2020; Nalborczyk, 2019). We first showed that induced rumination was accompanied by increased facial (both over a forehead and a perioral site) muscular activity in comparison to a rest period (Nalborczyk et al., 2017). However, because rumination was only compared to a rest period, it remained uncertain whether this perioral activity was specifically related to (inner) speech processes. Therefore, we ran a follow-up study comparing verbal to non-verbal rumination, which suggested that the facial EMG correlates we had previously identified were not specifically related to the verbal content of the ruminative thoughts (Nalborczyk, Banjac, et al., 2021). We discussed these findings in length and proposed several theoretical interpretations that can account for these results in the discussion of Nalborczyk, Banjac, et al. (2021) and more extensively in Nalborczyk (2019).

Moffatt et al. (2020) designed an experiment with the aim of refining our understanding of the involvement of the speech motor system in different varieties of inner speech and clarifying the relation between the peripheral correlates of inner speech and (self-reported) subjective experience. Their main conclusion is that inner experience between induced rumination and distraction differs “without a change in electromyographic correlates of inner speech” (p.1). In other words, they suggest that the subjective experience of inner speech is unrelated (or loosely related) to the electromyographic correlates of inner speech, which are thought to be represented mostly by the EMG amplitude recorded over the orbicularis oris inferior and orbicularis oris superior muscles. However, for this in-sample observation to be of interest in an out-of-sample context (i.e., to be informative for other non-observed individuals, or said otherwise, to bring information about the population), this absence of difference should be substantiated by adequately-powered statistical tests (given the target effect size) as well as reliable measures. This is unlikely to be the case here, for reasons that we will present and discuss in the present article.

Re-analysing the data from Moffatt et al. (2020)

Exploring the data

As typical in studies manipulating induced rumination, Moffatt et al. (2020) designed a two-step protocol. First, they aimed to induce a negative mood by asking participants to solve unsolvable or excessively difficult anagram and subtraction tasks. Second, they prompted participants to either ruminate on these (purportedly induced) negative feelings (by asking them to “think about the causes, consequences, and meaning of their current feelings”) or to distract themselves (by asking them to “think about a village, city or town that you are particularly familiar with”). Rumination and distraction were manipulated within-subject, with all subjects alternating between rumination and distraction, in a counter-balanced order.

Their final sample of participants, after data exclusion, included 26 participants (data available at https://osf.io/hj7tz/). The EMG data is depicted in Figure 1 by condition (where BAS, DIS, and RUM refer to the baseline, distraction, and rumination conditions, respectively) and by muscle (frontalis, FRO; orbicularis oris inferior, OOI; and orbicularis oris superior, OOS). This figure shows that the average natural logarithm of the EMG peak amplitude recorded over the FRO was at similar levels in the baseline and distraction conditions, but was much higher in the rumination condition. However, the average natural logarithm of the EMG peak amplitude recorded over the OOI and OOS muscles was higher than baseline in both the rumination and distraction conditions, with a slight increase from distraction to rumination (both on the mean and median). Having described the data collected by Moffatt et al. (2020), we now turn to a discussion of some problems related to conclusions that can be made from under-powered non-significant results.

Conclusions from under-powered null-hypothesis significance tests

There is an infamous tradition of conducting and interpreting uninformative null-hypothesis significance tests in Psychology (e.g., Meehl, 1990a, 1967, 1978, 1990b, 1997). By "uninformative," we mean that some null-hypothesis significance tests are simply not diagnostic with regards to the substantive effect of interest (e.g., whether there is a difference between conditions A and B).

As highlighted by several authors (e.g., J. Cohen, 1994; Pollard & Richardson, 1987; Rouder et al., 2016), concluding that an effect is probably absent solely based on a non-significant p-value is the continuous (i.e., probabilistic) extension of the modus tollens and is not a valid argument (i.e., the conclusion does not follow from the premises). This fallacious argument is also known as the *fallacy of acceptance*, the *absence of evidence fallacy* or *the argument from ignorance*, and proceeds as follows: "If the null hypothesis is true, then this observation should *rarely* occur. This observation occurred. Therefore, the null hypothesis is false (or has low probability)." In short, this argument is fallacious

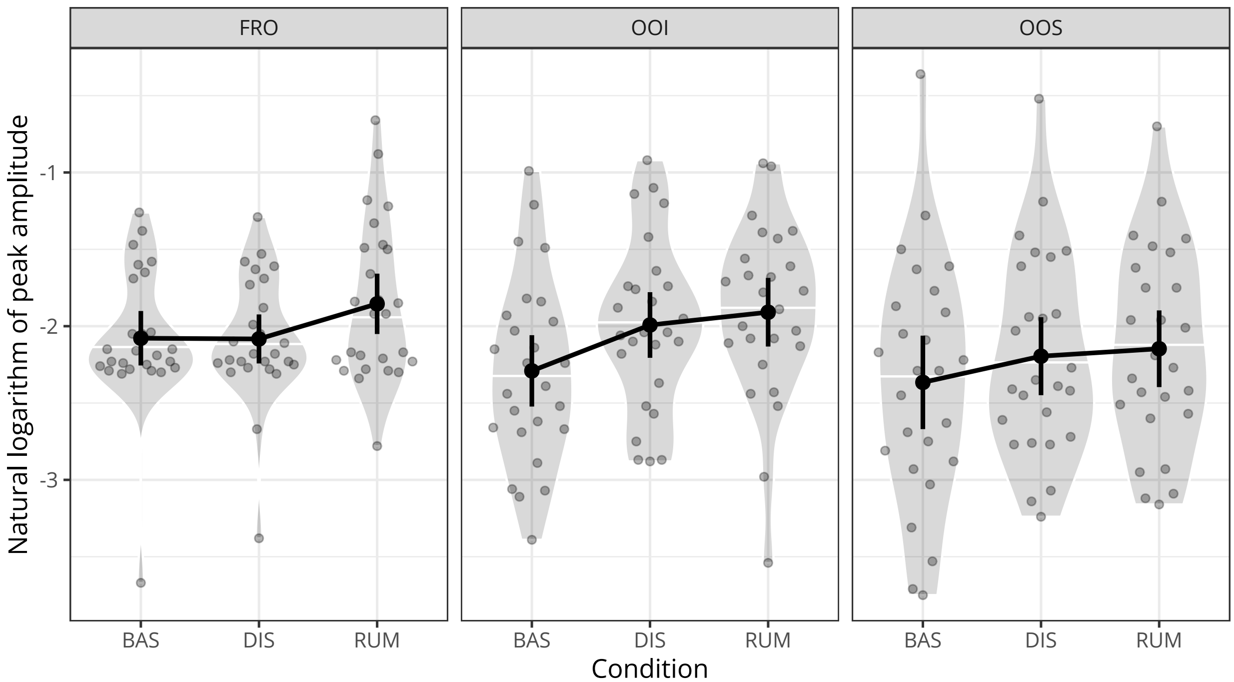

Figure 1. Average natural logarithm of the EMG peak amplitude per muscle and condition. The black dots and intervals represent the by-group average and 95% confidence interval (N = 26). The horizontal white line in the violin plot represents the median. The grey dots represent the individual-level average natural logarithm of the EMG amplitude by muscle and condition.

because it fails to consider the (probability of the data under the) alternative hypothesis.

This problem is tackled in modern usages of null-hypothesis significance tests by ensuring that the claim under scrutiny is submitted to *severe* tests (e.g., Mayo, 2018; Mayo & Spanos, 2006). In general terms, the strong severity principle states that we have evidence for a claim to the extent that it survives a stringent scrutiny, that is, to the extent that it survives *severe* tests. More precisely, some claim (e.g., $\theta = 0$) is said to be *severely tested* if it had great chances of being corroborated/falsified, had the claim been true/false. When a statistical test is under-powered (for detecting a given effect size), the claim under scrutiny is not strongly (severely) tested, hence it not possible to obtain strong or reliable evidence for the claim (bad test, no evidence).

Enthusiastic prior specifications

Anticipating the legitimate critiques on the power of their study, Moffatt et al. (2020) report the results of a power analysis using the effect size reported in Nalborczyk et al. (2017) of $d = 0.72$. This represents a highly optimistic estimate of the substantive effect of interest (i.e., the difference in the natural logarithm of the EMG peak amplitude between the rumination and distraction conditions) as this effect represents the standardised mean difference in EMG amplitude *between a rest and a rumination periods* as estimated in Nalborczyk et al. (2017).

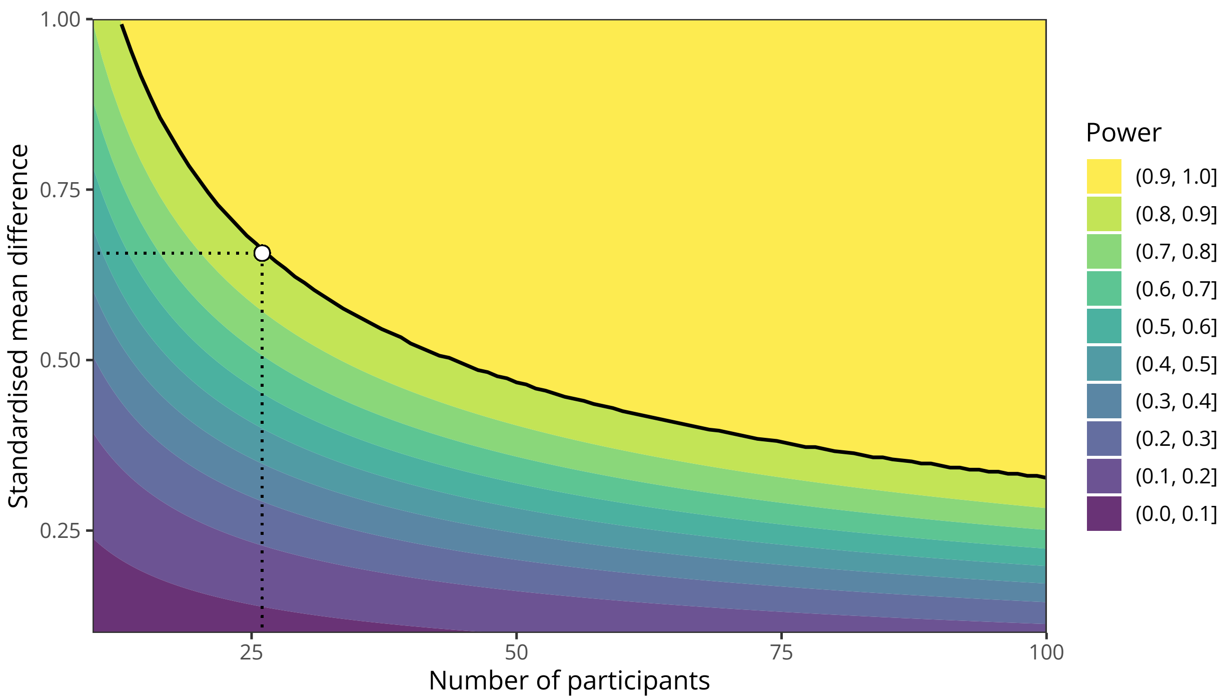

Figure 2. Statistical power as a function of both sample size and effect size, for a one-sample t-test with a significance level of 0.05. The white dot indicates the minimal effect size that can be detected with a probability equal or superior to 0.9 with a sample size of N = 26.

We suggest the (a priori) power of the study ran by Moffatt et al. (2020) was much lower than suggested by the authors. Indeed, we speculate that the standardised mean difference in EMG peak amplitude between the rumination and distraction conditions may be much weaker than the standardised mean difference in EMG amplitude between the rumination and rest conditions. If we assume that the former is half the size of the latter, therefore the a priori power of the main statistical test from Moffatt et al. (2020) was around 0.42, meaning that they had less than 1 chance out of 2 to find a significant effect (given that the population effect size was actually 0.36). Notice that whereas taking half the effect size of Nalborczyk et al. (2017) may seem arbitrary, Figure 2 shows that a one-sample t-test with a sample size of $N = 26$ is under-powered for a vast range of effect sizes.

Frequentist properties of Bayes factors

Once again, anticipating the legitimate critique that the absence of a significant difference is not necessarily "significant" evidence for the absence of an effect, Moffatt et al. (2020) reported the following Bayes factor (BF) analysis (p.12):

> "[...] therefore it is possible that the sample size of the present study lacked sufficient power to detect the effect of rumination on muscle activity. In order to test this, a Bayesian paired samples t-test was conducted for the peak log values of muscle activity between the rumination and distraction conditions. This revealed strong evidence in favour of the alternative hypothesis for the FRO muscle ($B_{10} = 18.79$), and moderate evidence in favour of the null hypothesis for the OOS ($B_{10} = 0.232$) and OOI ($B_{10} = 0.278$) muscles, according to current guidelines for interpreting Bayes factors [43]."

However, the current approach poses new problems. First, contrary to what the authors suggest, whereas computing a BF indeed allows assessing the *relative evidence* for the null, computing a BF (i.e., comparing two models) does not solve the problem of low power. More precisely, the sensitivity (i.e., the ability to attain a certain goal) of an experimental design to detect a given effect is an issue for both frequentist and Bayesian statistical tests. To illustrate this point, we simulated 10.000 datasets (for $N = 26$) under the assumption of either no effect (i.e., the null hypothesis of $d = 0$), an effect size of $d = 0.36$ (i.e., the supposed target effect size in Moffatt et al., 2020), or an effect size of $d = 0.72$ (i.e., the effect size reported in Nalborczyk et al., 2017).

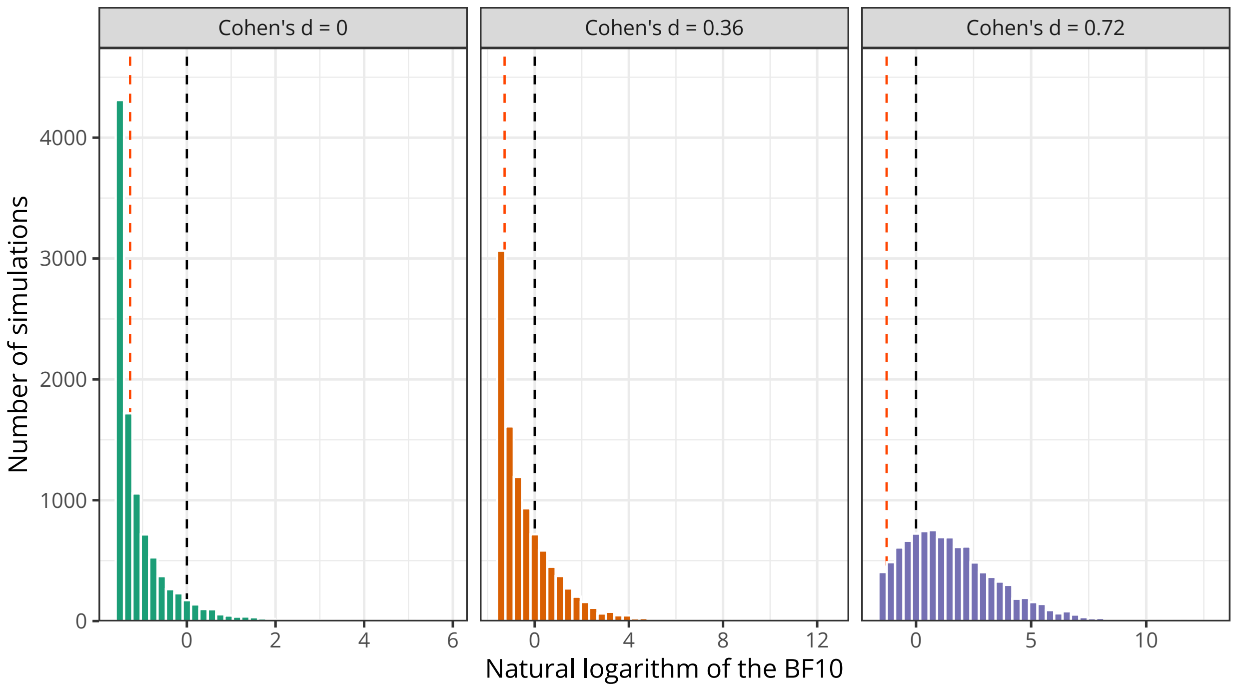

Figure 3. Illustrating the distribution of Bayes factors in favour of the alternative hypothesis for different population effect sizes (N = 26). In the left panel, the effect size is fixed to d = 0 (i.e., the null hypothesis), in the middle panel, it is fixed to d = 0.36 (i.e., the supposed target effect size in Moffatt et al., 2020), and in the right panel, the effect size is fixed to d = 0.72 (i.e., the effect size reported in Nalborczyk et al., 2017). The red vertical dashed line indicates the value of the BF computed for the OOI by Moffatt et al. (2020), on the log scale.

As shown in Figure 3, the distribution of log-BFs computed under each hypothesis reveals important inter-simulation variability. For instance, under the null hypothesis, 7.18% of the computed log-BFs are above 0 and hence support the alternative hypothesis (although the "true" effect size is $d = 0$). When the "true" effect size is of $d = 0.36$, 71.69% of the log-BFs are below 0 and hence support the null hypothesis (although the true effect size is actually non-null). When the "true" effect size is of $d = 0.72$, 25.26% of the log-BFs are still below 0. In other words, for small sample and effect sizes, BFs have high error rates. To assess the extent to which the BF computed for the OOI by Moffatt et al. (2020) (i.e., $\text{BF}_{10} = 0.278$) is "surprising" given or "compatible" with some effect size, we can compute the proportion of simulated BFs that are equal or more extreme than the one reported by the authors for each hypothesis, which is approximately equal to 0.57, 0.28, and 0.04, for the hypothesis of $d = 0$, $d = 0.36$, and $d = 0.72$, respectively. In other words, observing a $\text{BF}_{10} = 0.278$ is more compatible with the hypothesis of $d = 0$ (i.e., the null hypothesis) than with the hypothesis of $d = 0.36$ or $d = 0.72$, for this sample size.

The problems we discussed previously about the interpretation of under-powered non-significant results also apply to the test Moffatt et al. (2020) performed about the effect or the conditions' order. In Nalborczyk, Banjac, et al. (2021), we manipulated the modality of rumination (whether it is verbal or non-verbal) in a between-subject manner to avoid order effects and to avoid dissipating the effects of the negative mood induction. More precisely, we assumed that inducing rumination after a distraction condition in a within-subject manner would dissipate the effects of the mood induction and therefore reduce the impact of the rumination induction. In contrast to this approach, Moffatt et al. (2020) asked participants to ruminate and then distract themselves (or reciprocally), after an induced stressor (an induced failure). Anticipating again that the order of the within-subject conditions may be an issue, Moffatt et al. (2020) say:

> "Unless otherwise reported, the inclusion of order in which the conditions were completed as a between-subjects variable as part of a mixed-design ANOVA produced no significant main effects or interactions involving order." (p.7)

Unfortunately, obtaining a non-significant effect of the conditions' order is very weak evidence that order did not play a role in the results, given the low power of the tests that were performed (the sample size in each group was of N = 12 and N = 14).

Robustness of Bayes factors to prior specifications

Formulated in Bayesian terms, the problem of specifying credible effect sizes in a priori power analyses may be described as a problem of prior specification. However, defining sound prior distributions for the alternative hypothesis is notoriously difficult (for some guidance, see for instance Dienes, 2021, 2019). In Figure 4, we report the result of *sensitivity analyses*, depicting the value of the BF in favour of the alternative hypothesis (relative to the null hypothesis) for the difference between the distraction and rumination conditions, under various prior specifications, for each muscle.

This figure strikingly reveals large variability in the resulting BF with varying prior specification. More precisely, when the scale (width) of the prior put on the standardised effect size is changed (along the x-axis), the BF changes accordingly. For instance, varying the prior scale from 0.1 to 1.0 for the OOI results in BFs from 0.78 to 0.21, respectively.

Discussion and conclusions

With this short paper, we aimed to nuance the strong conclusion made by Moffatt et al. (2020), who asserted that the inner experience of rumination was not related to its peripheral muscular correlates. First, we discussed the statistical and epistemological reasons that cast doubt upon the main conclusion of Moffatt et al. (2020). Because the statistical tests conducted by Moffatt et al. (2020) were heavily under-powered, they provide only weak evidence for an absence of difference between conditions. Second, we highlighted that the frequentist properties of Bayesian tools (e.g., Bayes factors) provide an important piece of information that may help design more informative studies. Third, sensitivity analyses further suggested that various prior specifications may lead to different resulting Bayes factors.

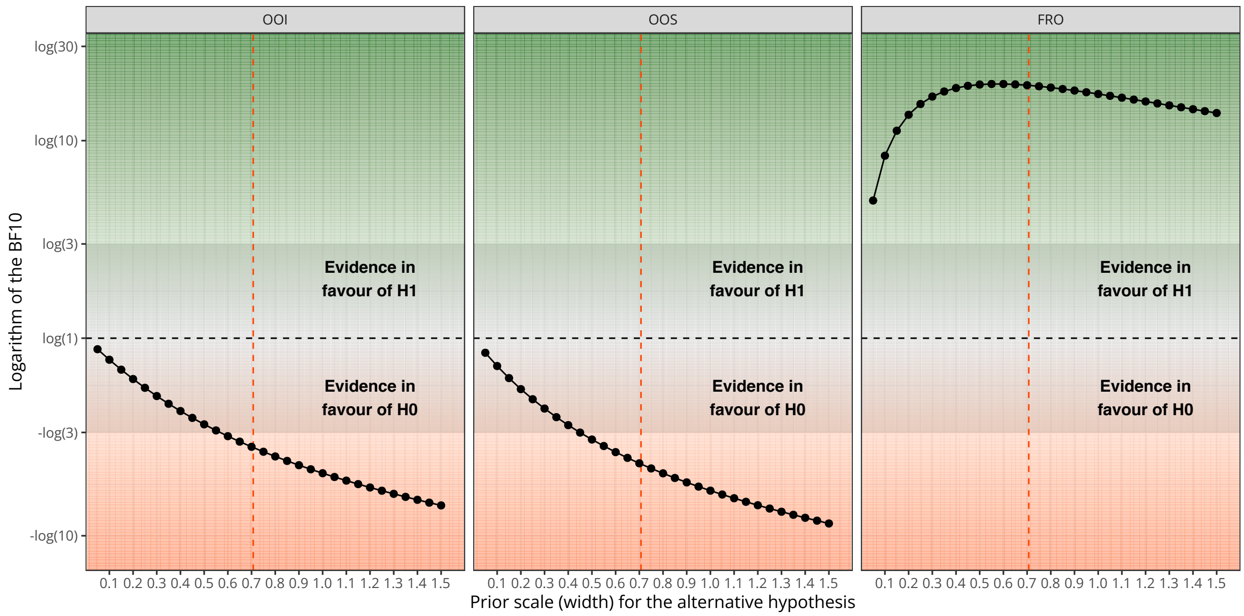

Figure 4. Prior sensitivity analysis for the Bayes factor computed for each muscle (OOI, OOS, and FRO). The x-axis represents the width of the prior put on the standardised effect size (i.e., the prior for the alternative hypothesis). The y-axis represents the logarithm of the Bayes factor in favour of the alternative hypothesis. The horizontal black dashed line represents equal support (evidence) for each hypothesis. The vertical red dashed line depicts the prior width used in Moffatt et al. (2020). The gray shaded area represents the conventional (but questionable) interval in which BFs are usually considered as negligible (i.e., from 1/3 to 3).

In addition to these methodological limitations, we now wish to discuss the theoretical interpretations and implications of these results. As discussed in the introduction section, we previously conducted several studies aiming to assess the role of the speech motor system in rumination. Following our initial study (Nalborczyk et al., 2017), we ran an extension in which we compared verbal to non-verbal rumination. The results suggested that the facial EMG correlates of verbal and non-verbal rumination were similar (Nalborczyk, Banjac, et al., 2021). Given the ample evidence on the EMG correlates of inner speech production (for an overview, see Chapter 1 in Nalborczyk, 2019), we needed to explain why this particular form of inner speech (induced rumination) was not associated with speech-specific peripheral muscular activity.

In Nalborczyk, Banjac, et al. (2021), we suggested that this observation was coherent with the mental-habit view of depressive rumination (Watkins & Nolen-Hoeksema, 2014), which defines rumination as a habitual behaviour, automatically triggered by contextual cues such as negative mood. We know habitual behaviours are more automatic (i.e., they are not intentionally initiated) than non-habitual behaviours. Interestingly, it has been observed that the automaticity with which a verbal thought is evoked may influence the degree to which it is enacted, that is, the degree to which it recruits the speech motor system (e.g., B. H. Cohen, 1986; Sokolov, 1972). According to B. H. Cohen (1986), the presence of peripheral motor activity during inner speech production may be interpreted in terms of attention sharing. For instance, in novel (hence non-automatic) or difficult situations, the vividness of inner speech may be strengthened by increasing the speech motor activity, resulting in more salient auditory percepts. Relating this idea to the motor control framework we previously proposed (e.g., Grandchamp et al., 2019; Lœvenbruck et al., 2018), it may be said that the characteristics of the task or situation (e.g., novelty, difficulty) may influence the amount of inhibition that is applied to motor commands during inner speech production (Nalborczyk, Debarnot, et al., 2021), hence resulting in more or less visible peripheral muscular activity (for a discussion of these ideas in the broader context of motor imagery, see Guillot et al., 2012).

Another possible interpretation is that automatic forms of inner speech may rely more heavily on higher-level (e.g., memory-based) cognitive processes whereas less automatic (i.e., more intentional or deliberate) forms of inner speech may rely more on simulation mechanisms via the use of internal models of the speech motor system (Nalborczyk, Debarnot, et al., 2021; Nalborczyk, 2019). In other words, the production of automatic versus non-automatic inner speech would be underpinned by different processes that would involve the speech motor system to a different extent. This distinction is similar to the distinction between the two routes of prediction-by-association and prediction-by-simulation in speech perception and comprehension (Pickering & Garrod, 2013). The prediction-by-association mechanism would rely more on perceptual sensory experiences and domain-general cognitive abilities whereas the prediction-by-simulation mechanism would rely more on the simulation of the motor action leading to the speech auditory percept. In the former case, no peripheral muscular activity is expected, whereas in the latter case, the speech motor system would be involved in simulating or emulating the corresponding overt action (cf. also the distinction between *motor simulation* and *direct simulation* in Tian & Poeppel, 2012). Whether the physiological correlates of automatic versus non-automatic (deliberate) forms of inner speech differ because of inhibitory constraints or because they rely on different processes (e.g., prediction-by-association or prediction-by-simulation) remains an open empirical question. We previously discussed these issues in more length and suggested ways forward from an experimental perspective in the discussion of Nalborczyk (2019).

To conclude, we wish to bring some nuance to the conclusion of Moffatt et al. (2020), who stated that "In conclusion, induced rumination appeared to involve similar levels of inner speech-related muscle activity to a period of distraction" (p.14). In consideration of the limitations discussed in the present article, this conclusion seems hasty. Indeed, we provided theoretical (epistemological) and empirical (via simulation and sensitivity analyses) reasons to doubt the strength of the evidence in favour of the null hypothesis in this study. This commentary stresses the importance of planning adequately-powered studies of induced rumination, and the need for more thoughtful statistical analyses and data interpretation, as recommended by Wasserstein et al. (2019).

Acknowledgements

We thank Antonio Schettino for suggesting to include the "Dance of the Bayes factors" simulation and for providing helpful comments on a previous version of this manuscript. Thanks to Daniel Lakens and two anonymous reviewers for valuable comments on previous versions of this manuscript.

Funding information

This project did not benefit from any specific funding.

Data accessibility

Reproducible code for the manuscript, analyses, and figures is available at https://osf.io/ba3gk/.

Competing Interests

The author has no competing interests to declare.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Cohen, B. H. (1986). The motor theory of voluntary thinking. In R. J. Davidson, G. E. Schartz, & D. Shapiro (Eds.), *Consciousness and self-regulation*. Springer, Boston, MA.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Dienes, Z. (2021). Obtaining Evidence for No Effect. *Collabra: Psychology*, *7*(1), 28202. https://doi.org/10.1525/collabra.28202

Dienes, Z. (2019). How Do I Know What My Theory Predicts? *Advances in Methods and Practices in Psychological Science*, *2*(4), 364–377. https://doi.org/10.1177/2515245919876960

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy*, *1*(3), 192–205. https://doi.org/10.1680/ijct.2008.1.3.192

Goldwin, M., & Behar, E. (2012). Concreteness of idiographic periods of worry and depressive rumination. *Cognitive Therapy and Research*, *36*(6), 840–846. https://doi.org/10.1007/s10608-011-9428-1

Goldwin, M., Behar, E., & Sibrava, N. J. (2013). Concreteness of depressive rumination and trauma recall in individuals with elevated trait rumination and/or posttraumatic stress symptoms. *Cognitive Therapy and Research*, *37*(4), 680–689. https://doi.org/10.1007/s10608-012-9507-y

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.-P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & Lœvenbruck, H. (2019). The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02019

Guillot, A., Di Rienzo, F., MacIntyre, T., Moran, A., & Collet, C. (2012). Imagining is not doing but involves specific motor commands: A review of experimental data related to motor inhibition. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00247

Lœvenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciu, M., & Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: To predict and to hear, see, feel. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: New voices* (p. 37). Oxford University Press.

Martin, L. L., & Tesser, A. (1996). Some ruminative thoughts. In R. S. Wyer (Ed.), *Advances in social cognition, vol. 9* (Vol. 9, pp. 1–47). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Marwick, B. (2019). *Wordcountaddin: Word counts and readability statistics in r markdown documents*.

Mayo, D. G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press. https://doi.org/10.1017/9781107286184

Mayo, D. G., & Spanos, A. (2006). Severe Testing as a Basic Concept in a Neyman–Pearson Philosophy of Induction. *The British Journal for the Philosophy of Science*, *57*(2), 323–357. https://doi.org/10.1093/bjps/axl003

McLaughlin, K. A., Borkovec, T. D., & Sibrava, N. J. (2007). The effects of worry and rumination on affect states and cognitive activity. *Behavior Therapy*, *38*(1), 23–38. https://doi.org/10.1016/j.beth.2006.03.003

Meehl, P. E. (1990a). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meehl, P. E. (1967). Theory-testing in Psychology and Physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. https://doi.org/10.1037/0022-006X.46.4.806

Meehl, P. E. (1990b). Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports*. https://doi.org/10.2466/pr0.1990.66.1.195

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. *What If There Were No Significance Tests?*, 393–425. http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=72FF987997EFB5F0602B02E1A2E04E40?doi=10.1.1.693.9583

Moffatt, J., Mitrenga, K. J., Alderson-Day, B., Moseley, P., & Fernyhough, C. (2020).

Inner experience differs in rumination and distraction without a change in electromyographical correlates of inner speech. *PLOS ONE*, *15*(9), e0238920. https://doi.org/10.1371/journal.pone.0238920

Müller, K. (2017). *Here: A simpler way to find your files.* https://CRAN.R-project.org/package=here

Nalborczyk, L. (2019). *Understanding rumination as a form of inner speech: Probing the role of motor processes* [PhD thesis, Univ. Grenoble Alpes & Ghent University]. https://thesiscommons.org/p6dct/

Nalborczyk, L., Banjac, S., Baeyens, C., Grandchamp, R., Koster, E. H. W., Perrone-Bertolotti, M., & Lœvenbruck, H. (2021). Dissociating facial electromyographic correlates of visual and verbal induced rumination. *International Journal of Psychophysiology*, *159*, 23–36. https://doi.org/10.1016/j.ijpsycho.2020.10.009

Nalborczyk, L., Debarnot, U., Longcamp, M., Guillot, A., & Alario, F.-X. (2021). *The role of motor inhibition during covert speech production* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/3df57

Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Polosan, M., Spinelli, E., Koster, E. H. W., & Lœvenbruck, H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, *127*, 53–63. https://doi.org/10.1016/j.biopsycho.2017.04.013

Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Spinelli, E., Koster, E. H. W., & Lœvenbruck, H. (2020). *Articulatory suppression effects on induced rumination* [Under review]. Under review. https://psyarxiv.com/5d6h7/

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. https://doi.org/10.1017/S0140525X12001495

Pollard, P., & Richardson, J. T. (1987). On the probability of making Type I errors. *Psychological Bulletin*, *102*(1), 159–163. https://doi.org/10.1037/0033-2909.102.1.159

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rouder, J. N., Morey, R. D., Verhagen, J., Province, J. M., & Wagenmakers, E.-J. (2016). Is There a Free Lunch in Inference? *Topics in Cognitive Science*, *8*(3), 520–547. https://doi.org/10.1111/tops.12214

Sokolov, A. (1972). *Inner speech and thought*. Springer-Verlag.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00314

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond " $p <$ 0.05." *The American Statistician*, *73*, 1–19. https://doi.org/10.1080/00031305.2019.1583913

Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology*, *123*(1), 24–34. https://doi.org/10.1037/a0035540

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. https://CRAN.R-project.org/package=tidyverse

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. https://yihui.org/knitr/

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*.

Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown