How does administering the semantic fluency task repeatedly within a short time frame affect participants response generation performance?

Joshua Wieler[1], Ivan Nenchev[2], Eva Belke[1] ([1]Ruhr-Universität Bochum, [2]Charité Berlin)
Joshua.Wieler@rub.de

Semantic (e. g. animal) fluency tasks assess the spontaneous retrieval of category exemplars and is frequently used in psychiatric and neuropsychological screening. However, its processing requirements are still unclear (Shao et al., 2014). Zemla et al. (2023) collected typed animal fluency data from a cross-sectional sample of AmE speakers as a basis for modelling individual differences in retrieving information from semantic memory. To obtain sufficient data per person, they had participants complete three runs of fluency of three minutes each within a period of about 35 minutes. Yet, it is unclear how the repeated administration of the fluency task affected participants' responses. Accessing a semantic category repeatedly is known to lead to a decline in retrieval performance (e. g., Howard et al., 2006). Hence, one might expect that participants' performance in second and third runs of a fluency task should decline as compared to the first. However, as participants are allowed to repeat responses, they will benefit from repetition priming across runs. In this pre-registered reanalysis of the data by Zemla et al. (2023), we aimed at characterizing predictors for participants to repeat subcategory clusters (e.g., elephant, rhino, lion) and single exemplars by the age of acquisition (AoA), frequency, valence, dominance, arousal, concreteness of individual exemplars (or *items*; see Appendix A for databases). Based on network science approaches to lexical structure (e. g., Steyvers & Tenenbaum, 2005), we assessed two key hypotheses, using mixed effects models: *H1) Clusters that include as first or only responses[1] an early-acquired, frequent item are repeated more often than clusters with later-acquired, less frequent first or only items.* We found that most clusters from Run 1 are repeated in Run 2 *and* 3 (see Figure 1 and Table 1). In keeping with the evidence from network science, multi-item clusters were repeated more often the more early acquired and valent and the *less* frequent, arousing and dominant their first items were. By contrast, clusters consisting of single items in Run 1 (SICs) were repeated more often the more frequent the items were. *H2) Compared to items retrieved in the second half of a multi-item cluster and those retrieved in single-item clusters, the items retrieved in the first half of a multi-item cluster (= baseline) are acquired earlier and are more frequent, especially those retrieved in the first run.* Our analyses confirmed the predicted main effects only for AoA and frequency but showed no interaction (Figure 2; for concreteness, see Figure caption). In assessing further pre-registered hypotheses, we found that items in single-item clusters were significantly less arousing, valent and dominant than items generated in the first half of multi-item clusters, again showing no interaction with Run (see Figure 2). Our data suggest that the retrieval patterns across runs were comparable and that network science approaches can inform models of lexical retrieval. [1]We refer to both single and a series of subcategory items as clusters.

Figure 1. Predicted probabilities for no, one, and two repetitions of clusters from Run 1, visualising the interaction patterns for the AoA and log10Frequency of the first item per cluster in Run 1 and Cluster Type.

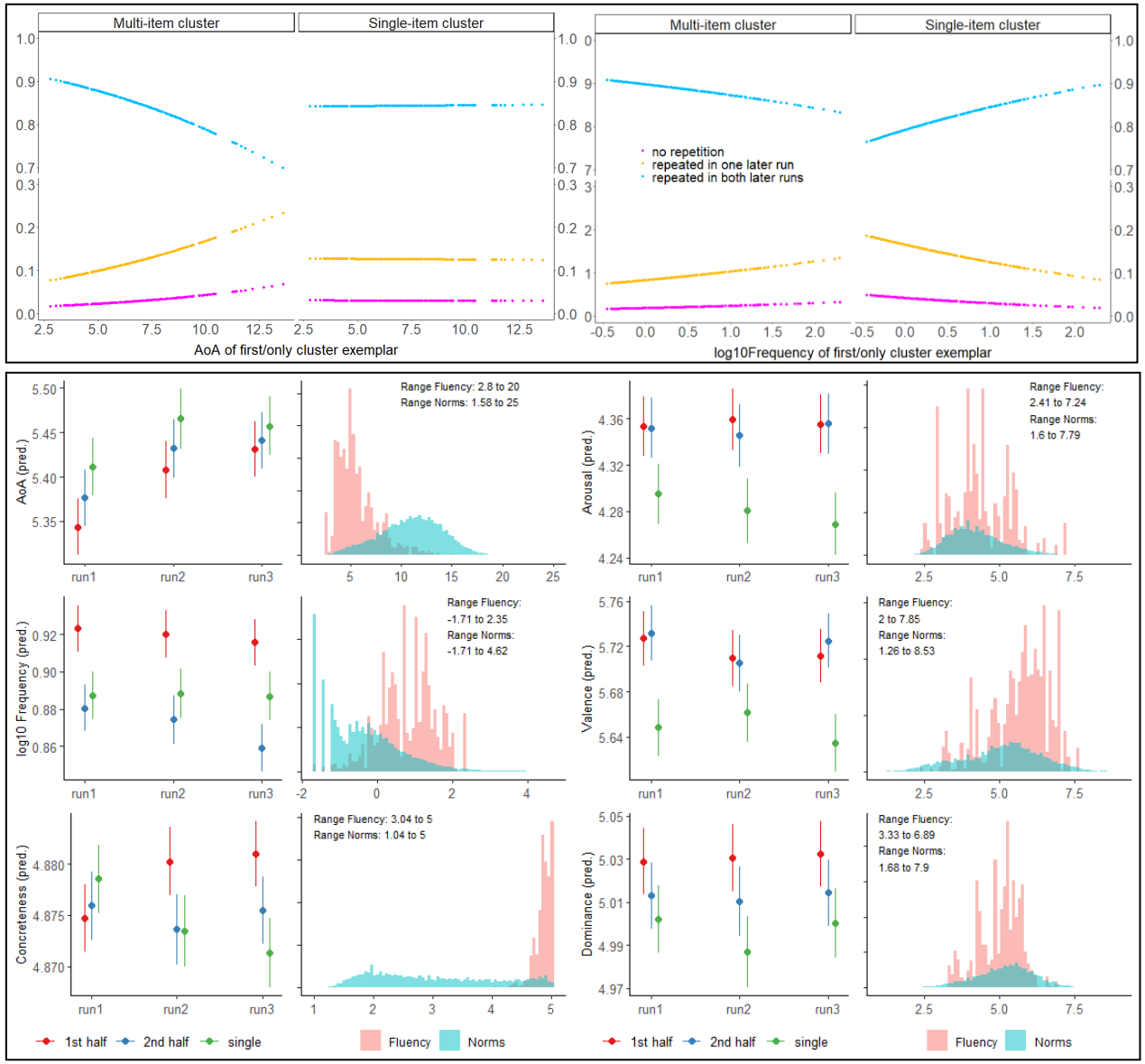

Table 1. Mixed effects model of Cluster Repetition (none, one, two) by Cluster Type in Run 1 (Multi- vs Single-item Cluster, SIC) and the predictors generated from the norms. Effects with confidence intervals [CI] for odds ratios above or below 1 are printed in black (all *p*s < .02).

Fixed Effects	Odds Ratios [CI]	z
- Repetition: no | one	0.02 [0.02-0.03]	-44.2
- Repetition: one | two	0.15 [0.13-0.17]	-28.5
$AoA_{1st\ Item}$	0.78 [0.66-0.92]	-2.99
$log10Freq_{1st\ Item}$	0.83 [0.71-0.97]	-2.33
$Concreteness_{1st\ Item}$	1.07 [0.97-1.17]	1.29
$Valence_{1st\ Item}$	1.18 [1.06-1.31]	2.94
$Arousal_{1st\ Item}$	0.90 [0.82-0.99]	-2.10
$Dominance_{1st\ Item}$	0.81 [0.72-0.91]	-3.41
Single-Item Cluster	0.76 [0.67-0.85]	-4.47
$AoA_{1st\ Item}$ × SIC	1.30 [1.05-1.61]	2.37
$log10Freq_{1st\ Item}$ × SIC	1.56 [1.27-1.92]	4.22
$Concr.ness_{1st\ Item}$ × SIC	1.13 [0.99-1.29]	1.88
$Valence_{1st\ Item}$ × SIC	0.91 [0.79-1.05]	-1.32
$Arousal_{1st\ Item}$ × SIC	1.13 [0.99-1.29]	1.83
$Dominance_{1st\ Item}$ × SIC	1.17 [1.00-1.36]	2.01
Rand. Eff. $\sigma^2 = 3.29$	$N_{id} = 527$, N_{Obs} 8933	
$\tau_{00\ id} = 1.02$	Marginal R^2: = .027	
ICC = 0.24	Conditional R^2: = .256	

Figure 2. Effects of Run and Position of a generated animal (first vs. second half of a multi-item cluster vs. single-item cluster) on its properties (estimated marginal means with all other influences held constant). Histograms show the distribution of all values in the norms and of the values of the items in the fluency data. For concreteness, the effects of run, position and the interaction may be spurious due to the high concreteness values for animals.

Appendix A: Resources and References

The pre-registration is available at https://aspredicted.org/Y2M_R29. In the present research, we focussed on testing hypotheses H1 and H2. As in the pre-registration, we use the terms exemplar and item interchangeably.

In order to establish AoA and frequency of items generated by the participants, we relied on the AoA norms provided by Kuperman et al. (2012) that also include AmE subtitle frequencies (Brysbaert et al., 2012). Valence, arousal, and dominance norms were obtained from Warriner et al. (2013) and concreteness norms from Brysbaert et al. (2014). Note that, as one would expect, concreteness scores ranged from 3 to 5 rather than from 1 to 5 only (see Figure 2), as there are no abstract animals.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*, 991–997. https://doi.org/10.3758/s13428-012-0190-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904-911. https://doi.org/10.3758/s13428-013-0403-5

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100,* 464 - 482.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978-990. https://doi.org/10.3758/s13428-012-0210-4

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology, 22,* 772. https://doi.org/10.3389/fpsyg.2014.00772

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78. https://doi.org/10.1207/s15516709cog2901_3

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191-1207. https://doi.org/10.3758/s13428-012-0314-x

Zemla, J. C., Gooding, D. C., & Austerweil, J. L. (2023). Evidence for optimal semantic search throughout adulthood. *Scientific Reports, 13,* 22528. https://doi.org/10.1038/s41598-023-49858-9