

Cats and Apples: Semantic Fluency Performance for Living Things Identifies Patients with Very Early Alzheimer's Disease

Sabine Krumm^{1,2,*}, Manfred Berres³, Sasa L. Kivisaari⁴, Andreas U. Monsch^{1,2,5}, Julia Reinhardt^{6,7}, Maria Blatow⁶, Reto W. Kressig^{1,2}, Kirsten I. Taylor^{5,8}

¹University Department of Geriatric Medicine FELIX PLATTER, Basel, Switzerland

²Faculty of Medicine, University of Basel, Basel, Switzerland

³Department of Mathematics and Technology, University of Applied Sciences Koblenz, Remagen, Germany

⁴Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Aalto, Finland

⁵Faculty of Psychology, University of Basel, Basel, Switzerland

⁶Department of Neuroradiology, Clinical Neuroscience Center, University Hospital Zurich, University of Zurich, Zurich, Switzerland

⁷Division of Diagnostic and Interventional Neuroradiology, Department of Radiology, University Hospital Basel, University of Basel, Basel, Switzerland

⁸Biomarkers and Translational Technology, Neuroscience and Rare Diseases (NRD), Roche Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland

*Corresponding author at: University Department of Geriatric Medicine FELIX PLATTER, Basel CH-4056, Switzerland. Tel.: +41 61 326 47 67; Fax: +41 61 326 47 77. E-mail address: sabine.krumm@felixplatter.ch (S. Krumm).

Received 26 May 2020; revised 7 September 2020; Accepted 19 October 2020

Abstract

Objective: Reduced semantic memory performance is a known neuropsychological marker of very early Alzheimer's disease (AD), but the task format that best predicts disease status is an open question. The present study aimed to identify the semantic fluency task and measure that best discriminates early-stage AD patients (PATs) from cognitively healthy controls. **Method:** Semantic fluency performance for animals, fruits, tools, and vehicles was assessed in 70 early-stage AD PATs and 67 cognitively healthy participants. Logistic regressions and receiver operating characteristics were calculated for five total score semantic fluency measures. **Results:** Compared with all other measures, living things (i.e., total correct animals + total correct fruits) achieved highest z-statistics, highest area under the curve and smallest difference between the upper and lower 95% confidence intervals. **Conclusion:** Living things total correct is a powerful tool to detect the earliest signs of incipient AD.

Keywords: Alzheimer's disease; Fluency (verbal/nonverbal); Assessment; Dementia; Mild cognitive impairment

Introduction

One chief neuropathological characteristic of Alzheimer's disease (AD) is neurofibrillary tau pathology. Cortical deposition starts in the medial temporal lobe, specifically in the perirhinal cortex (PRC; Braak & Braak, 1991). Several studies suggest that the PRC is involved in semantic object memory (see e.g., Hirni, Kivisaari, Monsch, & Taylor, 2013; Taylor, Moss, Stamatakis, & Tyler, 2006) and particularly important for the processing of living compared with nonliving things (nLT; Kivisaari, Tyler, Monsch, & Taylor, 2012; Tyler et al., 2004). Thus, semantic fluency tests are commonly used for the early diagnosis of AD. However, it is yet unknown which specific measure of semantic fluency performance for which specific category differentiates best between cognitively healthy subjects and patients (PAT) with very early AD. Using four different semantic fluency tasks, we investigated disproportionate decline of semantic fluency performance for living compared with nLT.

In living things, especially animals, PRC is hypothesized to integrate distinct information into a multimodal feature representation (Taylor et al., 2006) that is akin to human semantic object memories (Murray & Richmond, 2001). The PRC is particularly important when discriminating between visually ambiguous objects (i.e., high feature overlap; Tyler et al., 2004). According to the conceptual structure account by Taylor et al. (2007), object features differ with respect to the extent to which they are shared or variable. Although shared features indicate that this object belongs to a certain semantic category, variable features discriminate between different exemplars of the category. Living things (LT) have many shared features but relatively few variable features (Taylor et al., 2007). For example, many animals have four legs but only few are grey. However, this combination of shared and variable features may be valid for many different animals (rabbit, elephant, donkey, etc.). Therefore, LT are suggested to have many commonalities. In contrast, nLT usually have many variable features (e.g., has a blade) that are often strongly correlated among each other (e.g., has a blade and is used for cutting) but only few shared features (e.g., has a handle). Thus, nLT have less feature overlap than LT (Taylor et al., 2007). Consequently, PRC functionality may be particularly important to process LT and performance with these stimuli may be impaired in very early AD subjects. Additionally, highly correlated features (e.g., in nLT) are thought to be less vulnerable to neurodegeneration than weakly correlated features (e.g., in LT; see Taylor et al., 2007). Therefore, AD PATs are expected to show greater deficits for living than for nLT (Taylor et al., 2007). This assumption is bolstered by several studies that found category deficits for LT (i.e., categories with high feature overlap) in PATs with damaged PRCs (Moss, Rodd, Stamatakis, Bright, & Tyler, 2005; Noppeney et al., 2007; Wright, Randall, Clarke, & Tyler, 2015). The findings of Kivisaari et al. (2012) further support this hypothesis: fourteen cognitively healthy participants (NCs), 11 individuals with amnesic mild cognitive impairment (aMCI, a prodromal phase of impending AD), and 15 very early Alzheimer's dementia PATs were asked to name living or nLT on photographs. Cortical thickness of the PRC significantly predicted a quotient of $\frac{(\text{proportion correct living}) - (\text{proportion correct nonliving})}{(\text{proportion correct living}) + (\text{proportion correct nonliving})} \times 100$, suggesting that PRC integrity is important for item-specific discrimination of highly confusable visual objects such as LT. To our knowledge, this hypothesis has not yet been assessed in AD using non-visual tasks.

A simple and commonly used non-visual assessment is the semantic fluency task, that is, generating as many words as possible of a specific semantic category within a given timeframe. Successful performance requires multiple cognitive processes, including accessing and retrieving information from semantic memory, verbal working memory, and articulation (Birn et al., 2010; Henry & Crawford, 2004). In line with the importance of the temporal lobe for semantic memory performance, the semantic fluency task is one of the first tests to be impaired in very early AD PATs (Amieva et al., 2008). Moreover, Hirni et al. (2013) found a significant association between medial PRC signal intensity and animal fluency in a sample of 130 NCs, 32 aMCIs, and 10 PATs with very early Alzheimer's dementia, corroborating Kivisaari et al.'s (2012) results described earlier.

Although "animals" is the most commonly used category in semantic fluency tasks, other categories have been investigated in AD (e.g., colors, fruits, and cities by Amieva et al., 2008). As has recently been shown by Venneri et al. (2019), category fluency for car brands, animals, and fruits is significantly related to left PRC volume. Thus, semantic fluency tasks reflect the integrity of the PRC, which is the first structure affected by incipient AD. However, it remains unclear which category and which fluency outcome measure, best identifies early AD PATs. The present study aimed to address these open questions. Based on the conceptual structure account and the current knowledge about PRC functioning, we hypothesized that either a measure of high feature based commonality (e.g., LT) or a combined measure of high versus low feature similarity (e.g., living vs. nLT) will best discriminate early AD from healthy controls and, thus, best help diagnose very early AD PATs (i.e., PATs that not yet experience severe hippocampus related cognitive deficits in daily living).

Materials and Methods

Participants

A total of 137 native Swiss-German or German speaking adults were assessed at the Memory Clinic FELIX PLATTER, University Department of Geriatric Medicine FELIX PLATTER, Basel, Switzerland. Thorough medical screening ensured the neurologic and psychiatric health of the 67 control participants (NCs; 35 men; mean age = 72.2 years, $SD = 9.2$; mean education = 13.6 years, $SD = 3.8$; exclusion criteria: severe sensory or motor deficits; severe auditory, visual or speech deficits; severe systemic disease; diseases with severe or probable impact on the central nervous system [e.g., neurologic disorders including cerebral-vascular disease, generalized atherosclerosis, and psychiatric problems]; continuous mild-to-intense pain; and intake of potent psychoactive substances except minor tranquilizers). NC participants from two different neuropsychological studies were included. All had been recruited from the Memory Clinic's "Registry of Healthy Individuals Interested to Participate in Research." Depending on the study, marginally different rules to ensure their cognitive health were applied: In the first study,

conducted from 2011 to 2012, forty-two NCs had to receive normal scores on the mini-mental state examination (MMSE; Folstein, Folstein, & McHugh, 1975), Clock Drawing Test (Thalman et al., 2002), 15-items version of the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1978), and a German version of the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987). In the second study, implemented between 2016 and 2017, twenty-five NCs were allowed no more than one out of normal range score in the MMSE (Folstein et al., 1975), CVLT (Delis et al., 1987), Trail Making Test B (Reitan, 1958), or 16-items version of the Informant Questionnaire on Cognitive Decline in the Elderly (Jorm, Scott, & Jacomb, 1989). Seventy PATs (34 men; mean age = 74.7 years, $SD = 6.8$; mean education = 13.0 years, $SD = 3.1$) were recruited from the Memory Clinic FELIX PLATTER, University Department of Geriatric Medicine FELIX PLATTER, Basel, Switzerland, where they had received neuropsychological testing, magnetic resonance imaging (MRI), and medical/neurological examinations including blood analyses. Consensus diagnosis using all available data (i.e., comprehensive neuropsychology and MRI as well as gait analysis, cerebrospinal fluid analysis, positron emission tomography, etc. for some of the cases) was made by an interdisciplinary team of experienced clinicians. Thirty-three PATs were diagnosed with aMCI due to AD (Albert et al., 2011). Thirty-seven PATs were diagnosed with early Alzheimer's dementia according to NINCDS-ADRDA (McKhann et al., 2011) criteria. Groups (i.e., PATs vs. NCs) did not distinctly differ on gender ($\chi^2 = 0.07, p = .796$), age ($t = 1.79, p = .075$), and education ($t = -0.88, p = .383$). As expected, there were distinct differences in the MMSE scores (NCs: mean MMSE = 29.2, $SD = 1.0$; PATs: mean MMSE = 27.2, $SD = 2.0$; $t = -7.34, p = 5.18 \cdot 10^{-11}$). The study was approved by the local ethics committee and performed in compliance with relevant laws and institutional guidelines. Written informed consent was obtained from all participants.

Procedures

Semantic fluency data (total correct [TC] words produced within 1 min) were assessed separately for animals, fruits, tools, and vehicles. Participants were orally instructed to generate as many words as possible within the given, counterbalanced category without repeating any items. Task understanding was checked by asking participants to say a few items for the category "clothes." Perfectly synonymous expressions were counted as repetitions. Labels such as dog breeds or car types were allowed. Individuals were corrected only once per category if they switched category.

Analyses

TC scores were analyzed separately for animals, fruits, tools, and vehicles and for the following combined measures: LT: TC animals + TC fruits; nLT: TC tools + TC vehicles; difference between LT and nLT (LT–nLT): (TC animals + TC fruits) – (TC tools + TC vehicles); LT in relation to nLT (LT/nLT): $\frac{TC\ animals+TC\ fruits}{TC\ vehicles+TC\ tools}$; difference between living and nLT in relation to the sum of living and nLT $(LT - nLT)/(LT + nLT)$: $\frac{(TC\ animals+TC\ fruits)-(TC\ vehicles+TC\ tools)}{(TC\ animals+TC\ fruits)+(TC\ vehicles+TC\ tools)}$. Separate logistic regressions for each of the semantic categories and the combined measures were conducted. We used TC raw scores as independent variables and group as dependent variable. There were no missing values. Because groups were not demographically matched, covariates age, gender, and education were included in all analyses. Additional regression models and likelihood ratio tests were performed using the best measures (i.e., measures with high z -statistics in the logistic regression, high value on the area under the curve [AUC], and small difference between the upper and lower 95% confidence interval). The two best models were compared by DeLong's test for paired samples.

Results

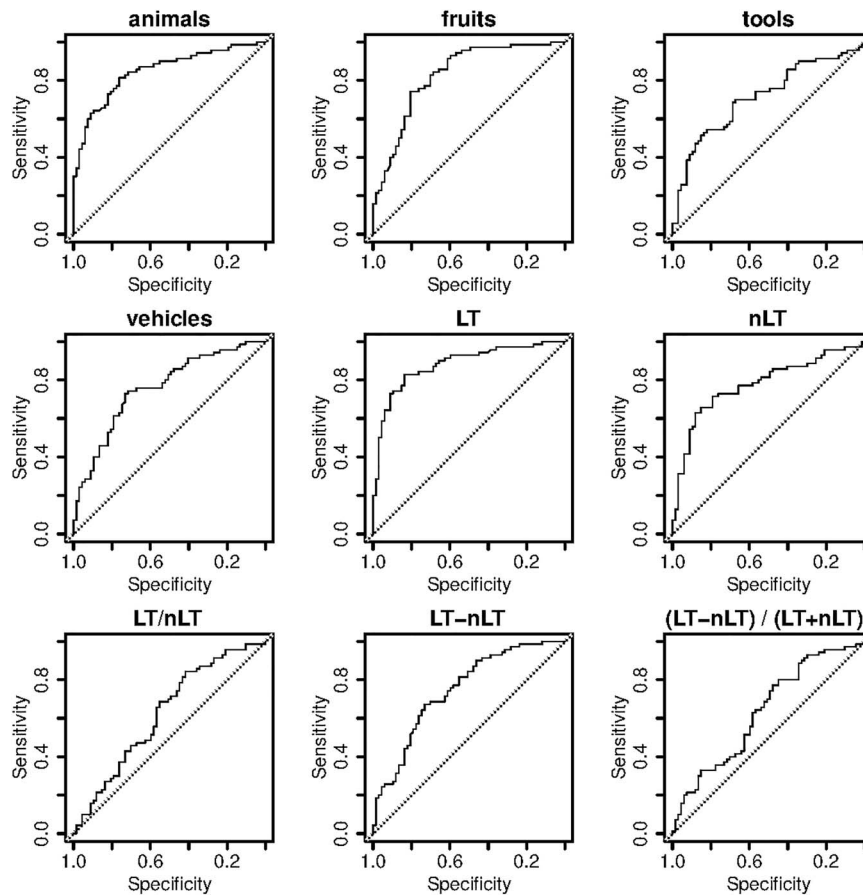
Means of correct items were highest in the animal category (NCs = 23.8, $SD = 4.5$; PATs = 17.2, $SD = 4.9$), second in the fruits category (NCs = 15.8, $SD = 3.7$; PATs = 11.6, $SD = 3.1$), followed by the vehicles category (NCs = 14.6, $SD = 4.0$; PATs = 11.0, $SD = 3.5$). Fewest items were produced in the tools category (NCs = 11.9, $SD = 3.2$; PATs = 9.7, $SD = 3.9$).

Logistic regressions revealed that animals, fruits and LT were the measures with the highest z -statistics and the largest AUC in the receiver operating characteristic (ROC) curves (see Table 1). The AUC of the model with LT was higher than the one for animals ($p = 0.038$ in DeLong's test). ROC curves are displayed in Figure 1. Animals and fruits were further investigated in one logistic regression model. The regression model achieved slopes \pm standard error of -0.25 ± 0.06 for animals and -0.27 ± 0.08 for fruits. The likelihood ratio statistic comparing the model that comprises animals, fruits, and demographic covariates with the model without fruits is 12.21, and comparing it with the model without animals is 18.39. Both are larger than the 99% quantile of the χ^2 distribution.

Table 1. Results of nine logistic regression models as well as receiver operating characteristics sorted by AUC. Results for demographic variables (included in each model) are not shown

Total score measure	Logistic regressions			Receiver operating characteristics	
	β	z	95% CI	AUC	95% CI
LT	−0.26	−5.70	−0.34 – −0.17	0.88	0.82–0.94
Animals	−0.33	−5.56	−0.44 – −0.21	0.85	0.78–0.91
Fruits	−0.42	−5.37	−0.57 – −0.26	0.82	0.76–0.90
nLT	−0.16	−4.14	−0.24 – −0.08	0.78	0.70–0.86
Vehicles	−0.29	−4.28	−0.42 – −0.16	0.76	0.68–0.84
LT–nLT	−0.18	−4.36	−0.26 – −0.10	0.75	0.66–0.83
Tools	−0.17	−2.91	−0.28 – −0.05	0.71	0.63–0.80
(LT – nLT)/(LT + nLT)	−3.43	−1.98	−6.82 – −0.03	0.64	0.54–0.73
LT/nLT	−0.55	−1.23	−1.43 – 0.32	0.62	0.53–0.72

Note: β = log odds ratio; AUC = area under the curve; CI = confidence interval; LT = living things; nonliving things = nLT.

**Figure 1.** Receiver operating characteristic curves of all investigated total scores measures.

Discussion

The present results suggest that TC animals, TC fruits, and TC LT, which is a combined measure of TC animals and TC fruits, best differentiate between cognitively healthy individuals and PATs with very early AD. Animals is the most commonly used category to assess semantic fluency. Our results show that all groups produced most words within this category, followed by fruits and vehicles. Most challenging for all participants was the tool category. These findings are comparable to [Capitani et al. \(1999\)](#), who found identical ranking of these categories with respect to number of produced words. LT achieved the highest z -statistics in the logistic regressions, the highest value on the AUC, as well as the smallest difference between the upper and lower 95% confidence interval. The results for animals and fruits were only slightly inferior to LT, whereas all other measures

achieved notably inferior results. However, the likelihood ratio tests suggested a distinct beneficial effect of adding a second LT category, indicating a substantial benefit when animals and fruits are both included in the model. The logistic regression model with both measures achieved very similar slopes for animals and fruits and suggests that the sum of animals and fruits (i.e., variable LT) is the best model to differentiate between groups. Thus, LT clearly fulfilled our criteria of the most promising measure to identify AD PATs at a very early stage in the disease process.

As predicted by the conceptual structure account, measures using nLT played a subordinate role when differentiating between cognitively NCs and very early AD PATs. One additional reason why nLT categories may be spared in early AD is the relative size of the category. Compared with animals, the clusters for fruits, tools, and vehicles are relatively small. For example, although there are diverse items for the cluster “mammals,” there are only few in the cluster “bicycle.” Cluster switches are very time-consuming (Gruenewald & Lockhead, 1980). Therefore, participants usually generate more items for the category living than nLT in the same amount of time, which leads to more distinct performance differences between PATs and healthy controls for living than nLT.

Our study benefits from the inclusion of well-characterized samples. Normal controls received extensive neuropsychological testing that assured their cognitive health. PATs were diagnosed by an interdisciplinary team of experienced clinicians before participating in this study. Although our results suggest that LT is the most promising measure for the earliest diagnosis of AD, this might not be prominently reflected in the clinical setting, where there are PATs with many different pathologies. Most PATs present themselves to a neuropsychological evaluation when already experiencing substantial cognitive deficits. At this stage, individuals will perform below-average in the LT measure but probably for animals, fruits and eventually other categories as well (see also Kivisaari et al., 2012). However, if individuals do not yet experience cognitive problems in daily living, we suggest that LT is the most auspicious measure to detect those with incipient AD. Hence, it is a valid assessment tool not only in research but also in the clinical setting, especially since acquiring LT needs only marginally more time than assessing animals only. Future studies should evaluate the beneficial effect of LT in a clinical setting, in contrast to assessing animals only. Additionally, comparing LT with phonemic fluency tests might help answer the question, if the difference between semantic and phonemic test performance in AD is just an overstated normal tendency (Laws, Duncan, & Gale, 2010) or if using LT might enlarge this discrepancy. Further, LT should be investigated in an MRI study to relate LT performance with cortical thickness in the PRC.

In conclusion, our findings suggest that the added TC scores of semantic fluency animals and fruits (i.e., a combined measure of LT) is a powerful tool in the diagnostic process of very early AD.

Funding

This work was supported by grants from the Novartis Foundations, Basel, Switzerland to A.U.M., the Swiss National Science Foundation (Ambizione fellowship PZ00P1_126493 to K.I.T. and grant 3200-049107 to A.U.M.), and the Alfred and Anneliese Sutter-Stöttner Foundation.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C. et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7, 270–279.
- Amieva, H., Le Goff, M., Millet, X., Orgogozo, J.-M., Pérès, K., Barberger-Gateau, P. et al. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology*, 64(5), 492–498.
- Birn, R. M., Kenworthy, L., Case, L., Caravella, R., Jones, T. B., Bandettini, P. A., & Martin, A. (2010). Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency. *NeuroImage*, 49(1), 1099–1107.
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82, 239–259.
- Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects word retrieval of certain categories in semantic fluency tasks. *Cortex*, 35(2), 273–278.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California verbal learning test*. San Antonio, TX: Psychological Corporation.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini mental state - a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, 12, 189–198.
- Gruenewald, P., & Lockhead, G. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 225–240.
- Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, 18(2), 284–295.
- Hirni, D. I., Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer's disease. *Neuropsychologia*, 51, 930–937.
- Jorm, A. F., Scott, R., & Jacomb, P. A. (1989). Assessment of cognitive decline in dementia by informant questionnaire. *International Journal of Geriatric Psychiatry*, 4(1), 35–39.
- Kaplan, E. F., Goodglass, H., & Weintraub, S. (1978). *The Boston Naming Test*. Boston: Veterans Administration Medical Center.
- Kivisaari, S. L., Tyler, L. K., Monsch, A. U., & Taylor, K. I. (2012). Medial perirhinal cortex disambiguates confusable objects. *Brain*, 135(12), 3757–3769.
- Laws, K. R., Duncan, A., & Gale, T. M. (2010). 'Normal' semantic-phonemic fluency discrepancy in Alzheimer's disease? A meta-analytic study. *Cortex*, 46(5), 595–601.

- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H. et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 7(3), 263–269.
- Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P., & Tyler, L. K. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*, 15, 616–627.
- Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology*, 11, 188–193.
- Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P. et al. (2007). Temporal lobe lesions and semantic impairment: A comparison of herpes simplex virus encephalitis and semantic dementia. *Brain: A Journal of Neurology*, 130(Pt 4), 1138–1147.
- Reitan, R. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271–276.
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences USA*, 103, 8239–8244.
- Taylor, K. I., Moss, H. E., & Tyler, L. K. (2007). The Conceptual Structure Account: A cognitive model of semantic memory and its neural instantiation. In Hart, J., & Kraut, M. (Eds.), *The neural basis of semantic memory*, (pp. 265–301). Cambridge, UK: Cambridge University Press.
- Thalman, B., Spiegel, R., Stähelin, H. B., Brubacher, D., Ermini-Fünfschilling, D., Bläsi, S., & Monsch, A. U. (2002). Dementia screening in general practice: Optimized scoring for the clock drawing test. *Brain Aging*, 2, 36–43.
- Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, 16, 351–362.
- Veneri, A., Mitolo, M., Beltrachini, L., Varma, S., Della Pietà, C., Jahn-Carta, C. et al. (2019). Beyond episodic memory: Semantic processing as independent predictor of hippocampal/perirhinal volume in aging and mild cognitive impairment due to Alzheimer's disease. *Neuropsychology*, 33(4), 523–533.
- Wright, P., Randall, B., Clarke, A., & Tyler, L. K. (2015). The perirhinal cortex and conceptual processing: Effects of feature-based statistics following damage to the anterior temporal lobes. *Neuropsychologia*, 76, 192–207.