Verbal Fluency and Its Semantic Basis: A Latent Semantic Analysis

Kyongje Sung[1], Kerry Ledoux[1], Erin J. Pickett[1], Tracy D. Vannorsdall[2], Shaina C. Fieldstone[2], David J. Schretlen[2,3], and Barry Gordon[1,4]

[1]Cognitive Neurology/Neuropsychology, Department of Neurology, [2]Psychiatry & Behavioral Sciences, [3]H. Russell Morgan Department of Radiology & Radiological Science, The Johns Hopkins University School of Medicine, Baltimore MD; [4]Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD

INTRODUCTION

- Semantic memory and semantic processing are important components of human cognition, and are affected by many disease conditions. One widely accepted model of semantic functions posits a long-term semantic memory store that includes language skills and is accessed by the executive system. The long-term semantic memory store is thought to depend on temporoparietal structures, while the executive system is thought to depend on prefrontal systems.
- Deficits in semantic performance can arise in this model from disruptions in the search for semantic information, problems retrieving semantic information, or degradation of the semantic store itself. Alzheimer's disease, semantic dementia, and Wernicke's aphasia are strongly associated with semantic impairments. Schizophrenia is also thought to involve semantic impairments, manifested by disorganized thoughts and speech.
- Category-cued word fluency tasks are widely used to test the semantic system. One measure derived from category fluency tests involves the clustering of semantically related items that are produced. Multi-dimensional scaling (MDS) has been used by several investigators to identify and characterize such semantic clusters. In one such study, Aloia et al. (1996) concluded that individuals with schizophrenia showed less coherent semantic clustering than healthy controls. This was interpreted as evidence of a disorganized semantic store in schizophrenia.
- However, applying MDS techniques to category fluency tasks is inherently limited: Only a small number of high frequency category exemplars (< 20) can be analyzed due to unstable proximity measures for low frequency exemplars (which might even be impossible to calculate). As a result, some conclusions drawn from MDS analyses might be suspect.
- We examined category fluency protocols from treated patients with schizophrenia (SZ) and healthy adults controls (NC), using a different clustering method called singular value decomposition (SVD; as used in latent semantic analysis [Landauer et al, 1997]). Our analyses confirm some findings suggested by MDS, but extend these in two ways: First, SVD techniques show that group differences can arise from the frequency of category exemplars produced, the dimensionality of semantic cluster representation, and the specific semantic category assessed. Second, our SVD results suggest that, in schizophrenia, impaired category fluency results from impaired retrieval of (or access to) semantic knowledge rather than an impoverishment of the semantic network itself.
- SVD is a matrix factorizing method and the factor analysis is a special case. **C** is a m-word by n-protocol binary matrix to be analyzed. The goal is to find the matrix **U** that gives meaningful clusters of m-word vectors in *'r'* optimal dimensional space based on the co-occurrence pattern of m-words in **C**.

$$\underset{m\times n}{\mathbf{C}} = \underset{m\times r}{\mathbf{U}} \cdot \underset{r\times r}{\mathbf{\Sigma}} \cdot \underset{r\times n}{\mathbf{V}}$$

METHOD

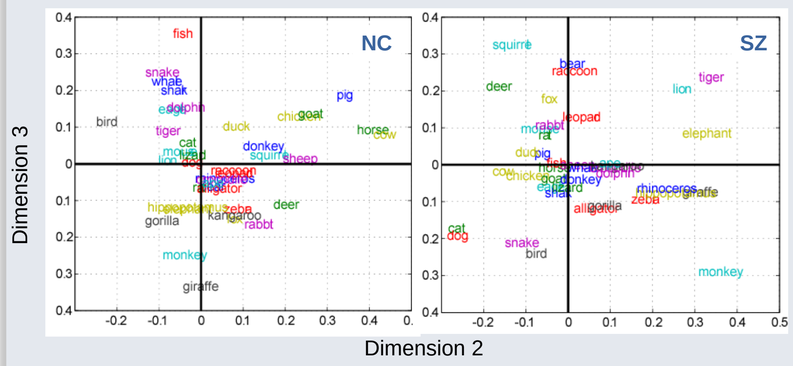
- **SZ:** 102 treated adult outpatients; Age = 39.7 ± 11.1 yrs; 70 males
- **NC:** 102 healthy adults; Age = 39.8 ± 11.0 yrs; 59 males
- Categories: Animals & Supermarket items (each for 1 min)

No. of examples produced

Group	Category	Mean (SD)
NC	Animal	20.3 (5.0)
	Supermkt	24.9 (6.5)
SZ	Animal	15.6 (5.3)
	Supermkt	18.8 (6.7)

RESULTS

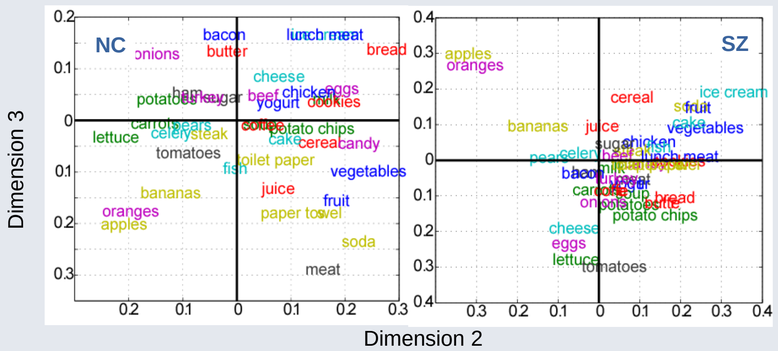
- 2-D cluster representations of top 40 animals show two common clusters for the top 20 animals in SZ and NC: 'domestic/farm' and 'wild/African'. However, only NC show a 'sea animal/avian' cluster for low frequency animals (rank 21-40). There was no clustering of the four highest ranked animals (dog, cat, lion, tiger) in NC, but SZ show clear clusters of them.
- Only 2-D representations are shown for illustration purpose.



- Correlations of clustering patterns (cosines of vector angles) between SZ and NC show that at least a 4-D representation is needed to accurately compare SZ and NC. Also, excluding the four top-ranked animals (dog, cat, lion, tiger) increases the similarity with which NC and SZ groups semantically represent the next highest ranked 20 animals.

Animal ranks	2-D space	3-D space	4-D space	5-D space
1 through 20	+.17 (.47)	+.29 (.39)	+.41 (.39)	+.39 (.26)
5 through 20	+.19 (.42)	+.38 (.32)	+.51 (.32)	+.44 (.21)
21 through 40	+.14 (.34)	+.31 (.27)	+.44 (.23)	+.47 (.22)

- 2-D cluster representations of supermarket items (1 to 40) show that there are clear clusters of 'dairy products', 'fruits & vegetables', 'meats', 'food categories' (e.g., fruits, fish), and 'non-food products' in NC. In contrast, SZ patients showed no clear clusters except for some low rank items on dimension 2.



- Correlations of clustering patterns in supermarket condition show that 2- or 3-D space representation is enough for comparison of two groups. SZ only show simple clusters along dimension 2 ('vegetable/fruit' & 'others').

Supermarket item ranks	2-D space	3-D space	4-D space	5-D space
1 through 20	+.38 (.41)	+.37 (.28)	+.41 (.28)	+.44 (.26)
21 through 40	+.52 (.39)	+.37 (.29)	+.36 (.29)	+.35 (.27)

CONCLUSIONS

- Our results confirm the superiority of the SVD technique over the MDS approach to extract semantic clustering information. In our data, at least 4 dimensions were required to accurately compare SZ patients and healthy controls on the animal category task. Analyzing fewer than 4 dimensions (as in MDS studies) would distort differences in semantic patterns shown by the two groups. Also, the disassociation of two category fluency tests among patients with SZ underscores the value of using multiple semantic fluency categories to reveal semantic impairments.
- Overall, the patients with SZ, even though treated, differed from healthy controls in their patterns of semantic clustering. Their clustering of low frequency animals was less coherent than it was for healthy controls, while their clustering of high frequency exemplars was comparable to that of the NC group. In contrast, on supermarket items, they showed less coherent clustering than NCs across nearly the entire range of item frequency, except for simple clusters of 'vegetable/fruit' and 'others'.
- Also noteworthy was the clear clustering of the four top ranked animals by patients with SZ, but not by NCs. A simple frequency explanation (i.e., high frequency animals co-occur with all other animals) does not explain why animals that are named a similar number of times by patients with SZ make clear clusters. An alternative explanation is that lexical retrieval processes for these top ranked animals could differ from those used to retrieve other animal names. Thus, the cue 'animal' might elicit such a strong automatic activation that no sub-category context is needed to guide retrieval. If this is correct, it implies that impaired automatic semantic priming could be a key deficit in SZ.
- The current findings are better explained by impaired access to or retrieval of semantic information in the patients with SZ than by a degradation of semantic knowledge. The reason is that, if the items successfully produced are the concepts that survive the degradation of information, this explanation would require an additional mechanism to explain why the successfully produced items are well clustered in one condition, yet not in the other. Especially since high frequency examples are thought to be less vulnerable to storage disruption (e.g., Rossell et al., 2006), the disassociation in performance between the two category tests cannot be adequately explained by simple degradation of information. The search/retrieval explanation may also need to assume a differential disruption in activation process between concepts, but it does not seem to require degradation of semantic information.
- We conclude that these data from clustering on the semantic category tasks are further evidence that schizophrenia is associated with a disruption in semantic processing. More specifically, this disruption appears to be in the access to or retrieval from semantic knowledge stores, rather than in the semantic long-term memory store itself.

REFERENCES

- Aloia et al. (1996). *J. Int. Neuropsych. Soc*. 2, 267-273.
- Landauer et al. (1997). *Psych. Rev.* 104, 211-240.
- Rossell et al. (2006). *Schizophr. Res.* 82, 121-134.

ACKNOWLEDGEMENTS

- This research was supported by the Therapeutic Cognitive Neuroscience Gift Fund and by the NIMH grants MH60504 and MH43775.