Modelling Topic Shifts in Spontaneous Thought

William Fisher & Buddhika Bellana
Department of Psychology, York University, Glendon Campus, CAN
Email: wfshr@my.yorku.ca

1. Motivation

Our mental experience is often characterized as a continuous stream of thought[1], in which we move fluidly from one topic to the next.

How we shift from one topic to the next is related to our mental health. For example, **perseverative thought is a hallmark of anxiety and depression**[2].

Current measures of perseverative thought rely on self-report[3] and **we lack an objective measure of this phenomenon**.

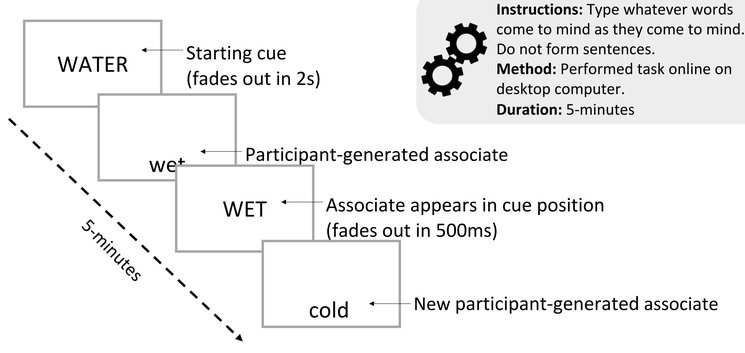
Here, we use a **free association paradigm** along with techniques from natural language processing (NLP) (i.e., word-embeddings) in an attempt to develop an automated way to quantify our movement through spontaneous thought.

2. Experimental design

Principle questions

I. How do we move between topics in our spontaneous thoughts?
II. How can we quantitatively measure this phenomenon?

Paradigm: *Free association*



Instructions: Type whatever words come to mind as they come to mind. Do not form sentences.
Method: Performed task online on desktop computer.
Duration: 5-minutes

Sample free association chains:

Random-generated	*Human-generated*
minute	crispy
tiger	fried
insurance	chicken
body	little
eel	sky
speed	falling
art	goose
solar	gander
gate	bird
ice	duck

Human data has **structure**. Free association is not random – we move between **related topics**. → **How can we measure this?**

3. Manual topic boundaries

body	dog
part	bull
time	cat
clock	hamster
hour	iguana
minute	tiger
rice	entertainer
boil	musician
water	flute
ocean	guitar
spray	cello
bottle	violin
top	piano
bottom	deaf
end	blind

Details: *Four* independent raters read each free association chain and indicated where a topic shift occurred.

Boundaries identified by:
- 100% of raters
- 75% of raters
- 50% of raters
- 25% of raters

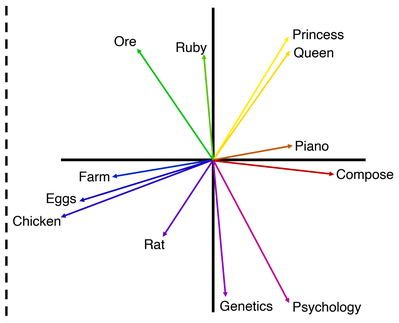
Consistency in boundary placement across raters suggests presence of **reliable topic shifts**. → **Can we detect these shifts automatically?**

4. Word embeddings

Word embeddings

- Technique from **natural language processing** (NLP) that estimates word meaning by looking at their **co-occurrence with other words** in a large corpus (e.g., Wikipedia)
- Words are represented using **300-dimensional vectors** (GLoVe; Pennington et al., 2014)
- Words with **similar vectors** tend to have **similar meanings**
- Similarity between word embedding vectors is calculated using **cosine similarity**
- **Higher** cosine similarity = **more similar** word embedding vectors

2-D schematic of word embeddings



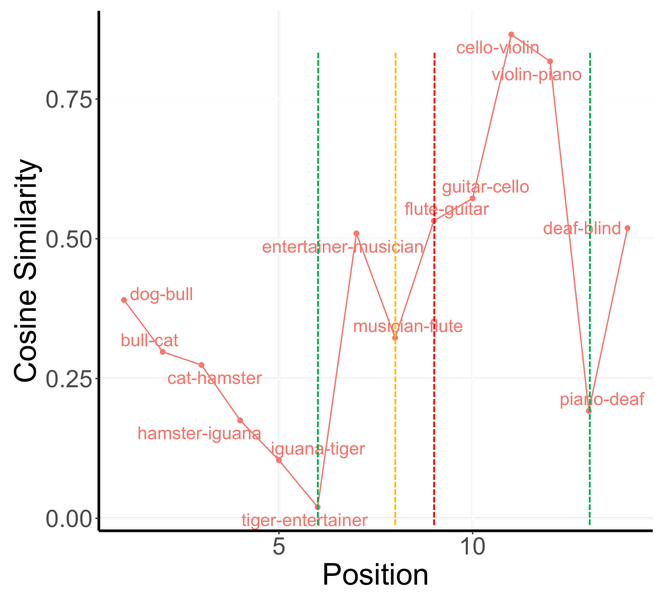
Examples of cosine similarities between adjacent free associates

iguana – tiger	0.10
tiger – entertainer	0.02
entertainer – musician	0.51
musician – flute	0.32
flute – guitar	0.53
guitar – cello	0.57
cello – violin	0.86
violin – piano	0.82
piano – deaf	0.19
deaf – blind	0.52

5. Results

How does our automated method correspond to independent rater boundaries?

Independent Rater Boundaries Vs. Cosine similarities
(An example participant's free association chain)



Data from full sample (n = 10)

Comparing Boundaries Between Automated Model and Independent Raters

Boundaries	Boundaries Identified by Independent Raters: ≥50% of Raters	≥75% of Raters
Total Count	174	91
Aligned with ≥0.1 drop in cosine similarity		
% Exact Match	38.5	42.9
% One-off	33.3	28.6
Combined Total (%)	71.8	71.5
Aligned with ≥0.2 drop in cosine similarity		
% Exact Match	23.6	31.2
% One-off	21.3	20.9
Combined Total (%)	44.9	52.1

Fluctuations in cosine similarity as measured via word embeddings correspond with the boundaries identified by our independent raters **70% of the time** when using a liberal threshold (>0.1).

This drops to 50% of the time with a more conservative threshold (>0.2).

Overall, these data show **initial promise** in using automated techniques (e.g., word embeddings) to detect topic shifts in free association data.

6. Summary

Automated techniques *can* detect shifts in topic in our spontaneous thought. Fluctuations in the similarity between word embeddings were able to correspond with up to 70% of human rated topic shifts.

Our threshold approach *overestimates* boundaries. While most human boundaries were associated with dips in cosine similarity, there were far more dips in cosine similarity than human rated boundaries. (>0.1 boundaries = 235, >0.2 boundaries = 133)

Next steps will be to improve boundary detection using more *sophisticated thresholding*, or by using other measures, like *reaction time*.

7. References

1. James, W., (1890). ***The principles of psychology*** **(Vol. 1). Macmillan.**
2. Ruscio, A. M., Seitchik, A. E., Gentes, E. L., Jones, J. D., & Hallion, L. S. (2011). Perseverative thought: A robust predictor of response to emotional challenge in generalized anxiety disorder and major depressive disorder. ***Behaviour Research and Therapy***, ***49***(12), 867–874.
3. Samtani, S., & Moulds, M. L. (2017). Assessing maladaptive repetitive thought in clinical disorders: A critical review of existing measures. ***Clinical Psychology Review***, ***53***, 14–28.