

Measuring behavioural and neural responses to fluctuations in real-world predictability

Buddhika Bellana (bbellana@jhu.edu)

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD USA

Hongmi Lee (hongmi.lee@jhu.edu)

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD USA

Xiaoye Zuo (zoeyzuo@jhu.edu)

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD USA

Janice Chen (janice@jhu.edu)

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD USA

Abstract:

A crucial component of brain function is to predict what will happen next. Although prediction is fundamental to brain function, we often study prediction of low-dimensional abstract tasks, rather than real-world events. We developed a novel behavioural approach to measure the dynamics of “real-world” predictability using audiovisual movies and natural language processing. Participants were shown an 11-minute movie, where viewing was occasionally interrupted by requests to generate sentences predicting what would happen next. These written predictions were converted into sentence embeddings using the Universal Sentence Encoder. Using these embedding vectors, we generated a timecourse of “situation-level” predictability during movie watching, revealing periods associated with homogeneous (high-predictability) and heterogeneous predictions (low-predictability) across-participants. We then regressed this timecourse of predictability on the fMRI data of a separate group of participants who watched the same movie, uninterrupted, in the scanner. During periods of high predictability, we observed higher activity in regions of the default mode network, while during periods of low predictability we observed higher activity in sensory cortices, consistent with internal-external models of cortical organization. Overall, we demonstrate the utility of natural language processing in quantifying fluctuations in real-world predictability.

Keywords: prediction; sentence embeddings; fMRI; default mode network; naturalistic

Background

Generating predictions about the world around us is thought to be a core feature of human brain function (Friston, Kilner, & Harrison, 2006). Our ability to generate predictions also fluctuates over time, with periods of high and low predictability. This fluctuation may have a purpose: during uncertain periods, we may

be biased towards gathering new information for *building internal models* of our environment, while predictable periods may instead be associated with *applying and refining existing models*. Despite the prominence of this dynamic in our theoretical understanding of the mind/brain (e.g., Cohen, McClure, & Yu, 2007; Honey, Newman & Schapiro, 2017), we know little about how the human brain responds to fluctuations in real-world predictability.

Current paradigms for studying predictability are mostly artificial. For example, fluctuations in predictability are often studied using variants of a gambling task (Lowenstein & Cohen, 2007). However, in a normal day, humans do not make a series of nearly identical choices, in rapid succession, based on otherwise meaningless cues and abstract rewards. Instead, we predict based on knowledge structures and schemas about the likely course of events. Moreover, in gambling tasks there is no a priori reason why the statistics of the task should change at a fast or slow timescale. Even in cases where the degree of predictability changes over the course of an experiment (e.g., drift designs), this rate is typically set by the experimenter and does not reflect the rates at which such changes occur in the real world. Thus, overall, the experimental control associated with such trial-based designs requires the imposition of low task dimensionality (e.g., two response options) and lacks the situational qualities and relevance to work knowledge that characterizes naturalistic contexts.

We decided to forgo the strict experimental control of gambling tasks and instead sought to measure the dynamics of predictability using audiovisual movies of the kind seen in theatres. Movies are a microcosm of real life, and thus provide a powerful stimulus for



measuring real-world fluctuations in predictability. Furthermore, a growing body of work in human neuroimaging has demonstrated that situation comprehension is reflected in the degree to which responses in higher-order brain regions (e.g., *default mode network; DMN*) are shared across individuals during movie-viewing (Chen et al., 2016; 2017; for review, see: Hasson, Chen, Honey, 2015). This shared neural response is thought to be driven by viewers building a common “situation model” of the unfolding narrative (e.g., Baldassano et al., 2017; Yeshurun et al., 2017); however, evidence supporting the critical link between ongoing activity in these regions and shared predictions is lacking.

Overall, we hypothesize:

1. Model testing during periods of high-predictability: Activity in higher-order brain regions (e.g., DMN) typically associated with situational comprehension will *increase during periods of high-predictability*, consistent with sampling from a shared internal predictive model.
2. Model building during periods of low-predictability: Activity in low-level sensory regions will *increase during periods of low-predictability*, consistent with a bias towards sampling the external environment.

Characterizing “Predictability” During Movie-Watching

We developed a novel experimental procedure to measure fluctuations in real-world predictability (Figure 1). 160 human participants watched an 11-minute clip of the popular film “Catch Me If You Can” (Spielberg, 2002) online via Amazon’s Mechanical Turk platform. Movie-viewing was periodically interrupted (~1 time per minute) by presenting participants with a prompt asking them “What do you think will happen in the next 30 seconds of this video?”. Participants were instructed to use full sentences to describe their predictions and were given the option to enter up to 5 separate predictions per interruption. Each prediction was followed by a confidence rating on a 1-7 point scale, providing a subjective estimate of how confident each participant

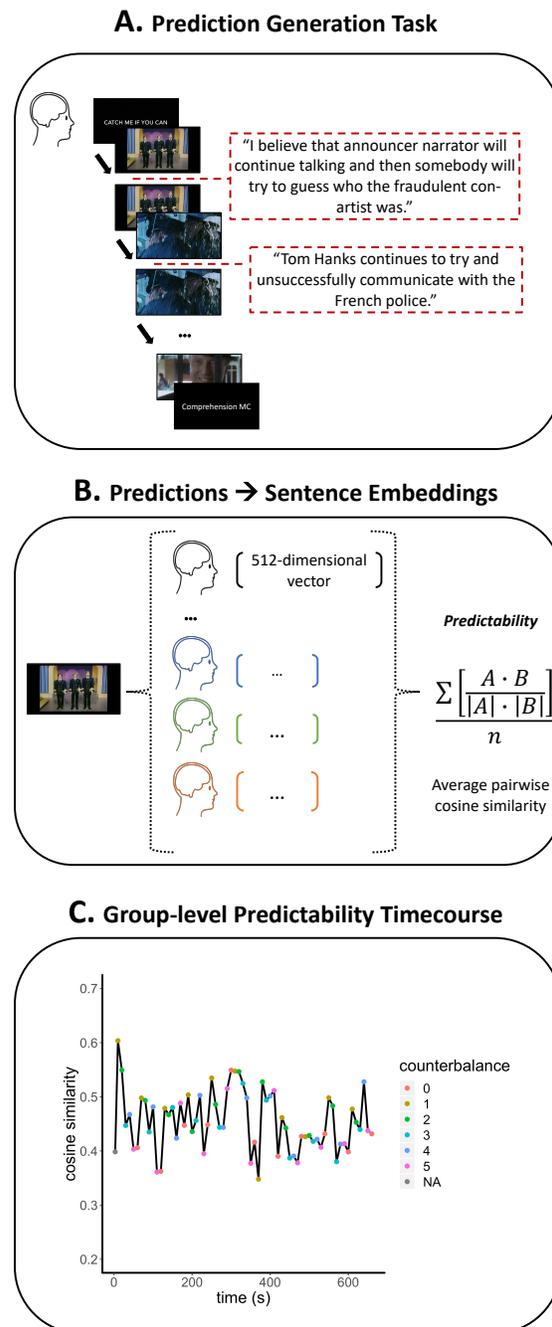


Figure 1. Schematic of overall experimental workflow, beginning with (A) the experimental paradigm where participants generate predictions, then (B) computing the mean cosine similarity across all predictions from all subjects who stopped at a specific timepoint, and (C) combining these estimates across all timepoints to generate a group-level timecourse of predictability for the movie stimulus.

felt about each prediction. The onset of the interruptions was counterbalanced such that 20-25 participants generated predictions for every 10 seconds of the movie. After finishing the 11-minute video, participants additionally completed a multiple-choice comprehension test.

The human-generated written predictions were then transformed into 512-dimensional vectors using the Universal Sentence Encoder (Cer et al., 2018). Sentences whose embedding vectors have high cosine similarity tend to have similar semantic meanings. Therefore, sentence embeddings provided an automated way to quantify the degree to which the predictions at a given timepoint, generated across the 20-25 participants per stop, shared the same meaning. *This measure of how semantically similar all predictions are at a given timepoint is our operational definition of predictability.*

An across-participant timecourse of predictability was calculated by averaging the pairwise estimates of cosine similarity across all predictions generated for each timepoint. Similarly, a confidence timecourse was calculated for confidence ratings by averaging participant-centered confidence ratings across all predictions for each timepoint. Only participants who accurately responded to *all* comprehension questions were included in this analysis. The group-level predictability and confidence timecourses were positively correlated ($r = .34$, $p = .005$), so that participants were more confident at moments when their predictions agreed with those of other participants.

Regressing “Predictability” On The Brain

How do brain dynamics change at moments of high and low predictability in a complex real-world narrative? A separate group of 23 human participants (19-36 years of age; 14 males, 9 females) watched the same movie clip during fMRI scanning. The predictability timecourse was then interpolated to match the sampling rate of the brain data (TR = 1.5s) and regressed onto each voxel of the preprocessed brain using AFNI’s 3dDeconvolve (Cox, 1996). Consistent with our hypotheses, activity in the higher-order regions of the DMN increased during periods of time where across-participant predictions were more consistent (i.e., high-predictability) (Figure 2, voxelwise FDR, $q < 0.05$). Moreover, activity in early auditory cortex and visual cortex was increased for moments at which predictions were more heterogeneous across participants (i.e., at moments of low predictability).

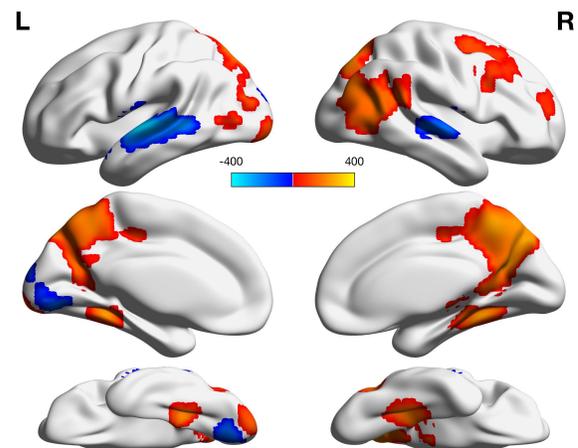


Figure 2. Results of voxel-wise regression of group-level predictability timecourse on the fMRI data of a separate sample ($n = 23$) who watched the same movie in the scanner. Warm-coloured regions show a positive relationship between activity and predictability, and include regions of the DMN. Cool-coloured regions show a negative relationship with predictability and are centered on visual and auditory cortex. False-discovery rate (FDR), $q < 0.05$

Conclusions & Future Directions

Overall, we report the utility of natural language processing techniques, specifically sentence embeddings, in capturing fluctuations in predictability derived during movie-watching. This experimental paradigm is powerful because it allows us to measure predictability at the scale of situation models – reflecting the relationships between entities, actions and outcomes (Zwaan & Radvansky, 1998), and “specifying the gist of the spatial, temporal and causal relationships that apply within a particular context” (Ranganath & Ritchey, 2012). Predictability in this sense is a high-dimensional construct and thus may carry external validity beyond the laboratory.

We observed a bias towards internally-oriented, high order cortical regions of the default mode network during periods of high predictability. This is consistent with previous work suggesting the DMN supports situation models of an unfolding naturalistic narrative (e.g., Chen et al., 2017; Baldassano et al., 2017; Yeshurun et al., 2017), and these data provide a critical link between activity in these regions and explicit participant-generated predictions. Conversely, increased activity in sensory cortices during periods of low-predictability is consistent with models of the brain

that emphasize switching between internal and external modes depending on environmental demands (e.g., Honey, Newman, & Schapiro, 2017).

Beyond these interesting possibilities, many more open questions remain, some of which we hope to address in the near future:

1. How reliable is this pattern across different movies?
2. How does predictability relate to measures of inter-subject correlation (ISC)?
3. How does prediction accuracy compare to predictability as measured here (i.e., reliability across participants)?
4. Does real-world predictability change across vs. within event boundaries?
5. Does information-seeking behavior during movie-viewing differ during periods of high vs. low predictability?

Acknowledgments

We acknowledge the support of the Alfred P. Sloan Foundation (Fellowship to JC).

References

- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017) Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, 95, 709-721.e5.
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y-H., Strobe, B. & Kurzweil, R. (2018) Universal Sentence Encoder. *arXiv* Available at: <http://arxiv.org/abs/1803.11175>.
- Chen, J., Honey, C. J., Simony, E., Arcaro, M.J., Norman, K. A., & Hasson, U. (2016) Accessing Real-Life Episodic Information from Minutes versus Hours Earlier Modulates Hippocampal and High-Order Cortical Dynamics. *Cerebral Cortex*, 26, 3428–3441.
- Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017) Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20, 115–125.
- Cohen, J. D., McClure, S. M., Yu, A. J. (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 933–942.
- Cox, R. W. (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computational Biomedical Research*, 29, 162–173.
- Friston, K., Kilner, J., & Harrison, L. (2006) A free energy principle for the brain. *Journal of Physiology Paris*, 100, 70–87.
- Hasson, U., Chen, J., & Honey, C. J. (2015) Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19, 304–313.
- Honey, C. J., Newman, E. L., & Schapiro, A. C. (2017) Switching between internal and external modes: A multiscale learning principle. *Network Neuroscience*, 1, 339–356.
- Ranganath, C., & Ritchey, M. (2012) Two cortical systems for memory-guided behaviour. *Nature Reviews: Neuroscience*, 13, 713-726.
- Spielberg, S. (Producer). (2012). *Catch Me If You Can* [Motion Picture]. United States: Dreamworks.
- Zwaan, R. A., & Radvansky G. A. (1998) Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.