Brain network dynamics predict surprise dynamics

Ziwei Zhang (zz112@uchicago.edu)
Department of Psychology, The University of Chicago

Monica D. Rosenberg (mdrosenberg@uchicago.edu)
Department of Psychology, The University of Chicago
Neuroscience Institute, The University of Chicago
Institute for Mind and Biology, The University of Chicago

Abstract:

Surprise is a fundamental human experience. We can be surprised by a plot twist in a movie or an underdog team's victory in a sports match. How much do surprising moments in our life have in common? Is there a generalizable brain signature of surprise? We identified a brain network model, the surprise edge-fluctuation-based predictive model (EFPM), whose regional interaction dynamics measured with functional magnetic resonance imaging (fMRI) predicted surprise in an adaptive learning task. The same model generalized to predict surprise as separate groups of individuals watched suspenseful basketball games and videos that violate psychological expectations. Our results suggest that shared neurocognitive processes underlie surprise across contexts and that distinct experiences can be translated into the common space of brain dynamics.

Keywords: Surprise; learning; uncertainty; network dynamics; fMRI

Introduction

We experience surprise, a transient process supported by distributed brain networks (Mazancieux et al., 2023), when reality conflicts with our expectations. Is surprise in very different situations subserved by similar neurocognitive processes? This is difficult to assess with behavioral measures alone because in some psychological paradigms surprise is measured explicitly (e.g., via button presses) whereas in others it is hidden (e.g., during passive viewing). Although measures such as pupil size (Kloosterman et al., 2015; Liao et al., 2018) and facial expression (Chang et al., 2021) track surprise in some contexts, they may be confounded by low-level visual properties.

Characterizing brain dynamics allows us to discover commonalities between belief-inconsistent surprise in distinct contexts. To this end, we propose an edge-fluctuation-based predictive model (EFPM) trained to identify functional interactions predicting moment-to-moment changes in belief-inconsistent surprise.

Methods

Adaptive learning task dataset

The data[1] associated with the adaptive learning task were shared by Kao et al. (2020) and McGuire et al. (2014).

Participants and task 32 participants (age range=18-30) were recruited. Participants underwent fMRI while performing an adaptive learning task in which they predicted the location of an object. The location of the object was drawn from a hidden distribution, with a mean that remained unchanged most of the time but changed unexpectedly ($p_{change} = 0.125$) when the mean was re-drawn from a uniform distribution.

Model of task surprise McGuire et al. (2014) developed a reduced Bayesian model to operationalize the amount of surprise in the environment. Highly surprising outcomes might signal meaningful changes in the environment that render belief updating and learning important. We operationalized surprise as a composite measure:

$$\text{surprise} = [CPP+RU*(1- CPP)]$$

where change-point probability (CPP) describes the probability of a change point (i.e., a mean shift in the generative distribution) occurring and relative uncertainty (RU) describes the uncertainty of this change occurring relative to noise (sum of the variance of the generative distribution conditional on a change point and no change point divided by this sum plus the variance of the generative distribution).

Naturalistic sports viewing dataset

The data[2] associated with NCAA sports viewing were obtained from Antony et al. (2021).

Participants and task 20 participants (age range=18-35) were recruited. In the MRI scanner, participants

[1] https://openneuro.org/datasets/ds003772/versions/1.0.1

[2] https://openneuro.org/datasets/ds003338/versions/1.1.0

watched the last five minutes of nine NCAA men's basketball tournament.

Model of naturalistic surprise Antony et al. (2021) first created a model to predict the win probability of a given team at each moment. The difference between win probability from one time point to the next was quantified as belief-inconsistent if the change involved the team with the higher win probability at a given moment becoming less likely to win at the next.

Expectation violation dataset

Participants and task We further tested the replicability of our findings in a third dataset[3] collected while participants (age range=18-45; n=29 participants agreed to share their fMRI data) watched 7.75-sec video clips showing expected or unexpected events in the domain of psychology and physics (e.g., agents moving through a solid wall; Liu et al., 2024).

Building the surprise EFPM

We quantified the extent to which activity in a pair of nodes in a whole-brain functional atlas (Shen et al., 2013) co-fluctuated at every moment in time following the edge-time-series method developed by Faskowitz et al. (2020). We used cross-validation to identify edges whose strength varied across trials with surprise. In each training fold, we selected *n*-1 participants and calculated the partial Spearman correlation (*rho*) between their edge time series and surprise time course, controlling for head motion. We selected edges significantly correlated with surprise across the training set. In the held-out individual, we correlated the strength of these edges with the surprise time course. After training the surprise EFPM in the learning task, we tested its generalizability to predict surprise in new contexts. We calculated moment-to-moment surprise EFPM summary scores in the naturalistic sports viewing dataset and ran a linear mixed effects model using this time course to predict belief-inconsistent surprise in the basketball videos, controlling for nuisance regressors (e.g., video motion). In the expectation violation dataset, we compared the strength of edges significantly related to surprise in the learning and basketball datasets during expected vs. unexpected videos.

Results

Model of brain network dynamics (EFPM) predicts surprise in a controlled learning task

EFPM successfully predicted surprise in held-out individuals (Fig. 1). Edges positively correlated with surprise were stronger on trials with more unexpected outcomes (mean within-subject partial *rho*=0.09; *p*=0.001) whereas edges negatively correlated with surprise showed the opposite pattern (mean within-subject partial *rho*=-0.10; *p*=0.001). Thus, moment-to-moment changes in edge strength predict moment-to-moment changes in belief-inconsistent surprise in novel individuals.

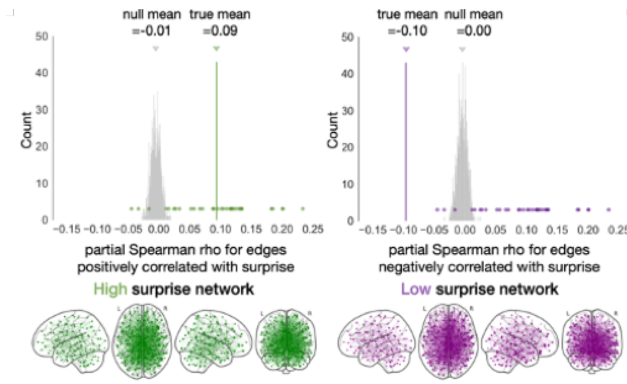

Figure 1: Surprise EFPM. Edges appeared in every cross-validation fold were selected and visualized.

The surprise EFPM generalizes to predict surprise in naturalistic contexts

The surprise EFPM predicted belief-inconsistent surprise in the NCAA basketball videos (β=0.037, t(65136.852)=3.947, *p*=0.047), accounting for nuisance regressors. In other words, the higher the co-fluctuation strength in the surprise EFPM, the more belief-inconsistent surprise in the video. The surprise EFPM thus predicts belief-inconsistent surprise in general.

We next calculated the strength of edges significantly related to surprise during the associative learning task and basketball games. Demonstrating further generalizability, this overlap surprise network was stronger when participants saw unexpected vs. expected psychology videos (t(51.911)=2.052, *p*= 0.045).

Discussion

The surprise EFPM captures common neural underpinnings of surprise in distinct cognitive contexts and individuals. Looking ahead, the surprise EFPM can be validated using complementary measures of surprise, including subjective measures (e.g., human ratings) and model outputs predicting behaviorally relevant outcomes from audiovisual or linguistic input (e.g., large-language-model-generated unsigned prediction errors). Furthermore, EFPM is a general framework that can be applied to predict other cognitive states from neuroimaging data.

[3] https://openneuro.org/datasets/ds004934/versions/1.0.0

Acknowledgments

We thank Joseph T. McGuire, Joseph W. Kable, and Chang-Hao Kao for sharing and addressing questions about the MRI and behavioral data associated with McGuire et al. (2014) and Kao et al. (2020). We thank James W. Antony for sharing and addressing questions about the MRI data, behavioral data, and analysis scripts associated with Antony et al. (2021). We also thank Joshua Faskowitz and colleagues for making the code associated with Faskowitz et al. (2020) publicly available, which we adapted and used in this work. We thank Shari Liu for addressing our questions about the MRI data associated with Liu et al. (2024). We thank Yuan Chang Leong, Kwangsun Yoo, Hayoung Song, and William X. Q. Ngiam for helpful feedback and suggestions on this work. Research was supported by National Science Foundation BCS-2043740 to M.D.R. and resources provided by the University of Chicago Research Computing Center.

References

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, Physiological, and Neural Signatures of Surprise during Naturalistic Sports Viewing. *Neuron,* 109(2), 377-390.e7. https://doi.org/10.1016/j.neuron.2020.10.029

Chang, L. J., Jolly, E., Cheong, J. H., Rapuano, K. M., Greenstein, N., Chen, P.-H. A., & Manning, J. R. (2021). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Science Advances*, 7(17), eabf7129. https://doi.org/10.1126/sciadv.abf7129

Faskowitz, J., Esfahlani, F. Z., Jo, Y., Sporns, O., & Betzel, R. F. (2020). Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nature Neuroscience*, *23*(12), 1644–1654. https://doi.org/10.1038/s41593-020-00719-y

Kao, C.-H., Khambhati, A. N., Bassett, D. S., Nassar, M. R., McGuire, J. T., Gold, J. I., & Kable, J. W. (2020). Functional brain network reconfiguration during learning in a dynamic environment. *Nature Communications*, 11(1), Article 1. https://doi.org/10.1038/s41467-020-15442-2

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068–1078. https://doi.org/10.1111/ejn.12859

Liao, H.-I., Yoneya, M., Kashino, M., & Furukawa, S. (n.d.). Pupillary dilation response reflects surprising moments in music. *Journal of Eye Movement Research*, *11*(2), 10.16910/jemr.11.2.13. https://doi.org/10.16910/jemr.11.2.13

Liu, Shari, Kirsten Lydic, Lingjie Mei, and Rebecca Saxe. 2024. "Violations of Physical and Psychological Expectations in the Human Adult Brain." doi: 10.31234/osf.io/54x6b.

Mazancieux, A., Mauconduit, F., Amadon, A., Willem de Gee, J., Donner, T. H., & Meyniel, F. (2023). Brainstem fMRI signaling of surprise across different types of deviant stimuli. *Cell Reports*, *42*(11), 113405. https://doi.org/10.1016/j.celrep.2023.113405

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, 84(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013