Available online at www.sciencedirect.com
SCIENCE DIRECT®

Neural Networks 18 (2005) 225–230

Neural Networks

www.elsevier.com/locate/neunet

Information theory, novelty and hippocampal responses: unpredicted or unpredictable?

Bryan A. Strange[a,b,*], Andrew Duggins[a], William Penny[a], Raymond J. Dolan[a], Karl J. Friston[a]

[a]*Wellcome Department of Imaging Neuroscience, Functional Imaging Laboratory, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK*
[b]*Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK*

Received 17 March 2004; accepted 17 December 2004

Abstract

Shannon's information theory provides a principled framework for the quantitative analysis of brain responses during the encoding and representation of event streams. In particular, entropy measures the expected uncertainty of events in a given context. This contextual uncertainty or unpredictability may, itself, be important for balancing [bottom-up] sensory information and [top-down] prior expectations during perceptual synthesis. Using event-related functional magnetic resonance imaging (fMRI), we found that the anterior hippocampus is sensitive to the entropy of a visual stimulus stream. In contrast, activity in an extensive bilateral cortico-thalamic network was dictated by the surprise or information associated with each particular stimulus. In short, we show that the probabilistic structure or context in which events occur is an important predictor of hippocampal activity.

Keywords: Information theory; Entropy; Surprise; Novelty; Predictive coding; Functional magnetic resonancce imaging; Hippocampus

Current notions of representational learning and inference in the brain rest on an interplay between bottom-up sensory information and prior expectations, mediated by lateral and top-down influences (Friston, 2002; Hinton, Dayan, Frey, & Neal, 1995; Kawato, Hayakawa, & Inui, 1993; Mumford, 1992). Irrespective of the precise mechanisms employed by the brain, the relative weight afforded these two sources of information is a generic and important issue. For example, a specific neuronal mechanism has been proposed for balancing the evidence from sensory inputs and prior expectations according to their predictability (Yu & Dayan, 2002). We therefore addressed the hypothesis that the hippocampus is sensitive to the probabilistic context established by event streams. This sensitivity would enable the hippocampus, or related systems, to regulate the balance between top-down and bottom-up effects in sensory cortical hierarchies.

To formulate our hypothesis in a quantitative way we used information to measure the stimulus-bound 'surprise' of a particular event, and 'entropy' to measure the context in terms of the average predictability of a sequence (Jones, 1979; Shannon, 1948). We hoped to show that entropy could explain variations in hippocampal responses, even after accounting for responses induced by event-bound surprise. The distinction between entropy and surprise is critical. Surprise is unique to a particular event and measures its improbability (e.g. a small p-value is informative in classical inference in rejecting the null hypothesis). Conversely, entropy measures the expected or average surprise over all events, reflecting the predictability of an outcome before it occurs.

$$I(x_i) = -\ln p(x_i); \quad H(X) = \sum_i -p(x_i)\ln p(x_i) = \langle I(x_i)\rangle$$

Surprise $I(x_i)$ quantifies the information conveyed by the occurrence of event x_i, whereas entropy $H(X)$ quantifies the expected information of events sampled from X.

Hippocampal damage impairs episodic memory (Scoville & Milner, 1957) and it has been observed that hippocampal

* Corresponding author. Address: Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK. Tel.: +44 207 833 7472; fax +44 207 813 1420.
E-mail address: bstrange@fil.ion.ucl.ac.uk (B.A. Strange).

0893-6080/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2004.12.004

activity reflects regularities embedded in an environment or task (Eichenbaum, Dudchenko, Wood, Shapiro, & Tanila, 1999). Information theory posits that these regularities depend upon the entropy. Using event-related functional MRI and carefully constructed sequences of visual stimuli, we tested the hypothesis that (i) hippocampal responses are sensitive to the entropy established by sequential events whereas (ii) responses in sensory areas reflect the surprise of each event. Areas that respond more vigorously to surprising stimuli can be regarded as exhibiting more prediction error from a predictive coding perspective (Rao & Ballard, 1999).

As entropy is effectively the running average surprise (see below), we were able to manipulate entropy and surprise independently and dissociate their neurophysiological correlates. Briefly, we varied the entropy of the stimuli over blocks of trials. This allowed us to look for brain responses that covaried with entropy over blocks. Because each block comprised likely and unlikely events we were also able to assess correlates of surprise within-block. The analysis proceeded by modeling entropy and surprise-related responses in a subject-specific first-level analysis. The results of this analysis were taken to a second-level analysis to implement random effects inferences about these responses over subjects.

During fMRI, 12 subjects were presented with 12 blocks of 40 trials. Each trial comprised a brief presentation of a colored shape (stimulus duration: 500 ms; stimulus onset asynchrony: 2.2 s). In all trials within a block, two colors and two shapes were combined to form four possible outcomes, with different stimuli presented in the different blocks. The stimuli appeared in miniature for 5 s before the beginning of a block and remained in a row at the bottom of the screen throughout the block. Subjects were required to respond to the sampled item by pressing a key to identify the target's position in the row. A schematic of a trial is shown in Fig. 1a.

Each trial used an independent sample from a distribution that remained constant within a block, but that varied over blocks. Note that there was no underlying sequence governing stimulus presentation, only the relative proportions of stimuli were varied from block to block. Subjects were asked to consider each block as a 'hat' containing a large number of objects of four distinct types with each hat containing a different set. They were also told that each trial would be equivalent to sampling an object and then returning it to the hat. Subjects were informed that the proportion of objects in a particular hat was completely unpredictable, and independent of the other hats sampled.

From the point of view of the subject, the suprise of each trial depended on the history of previous trials within a block. In these circumstances allocation of neuronal resources is based on the subjective probabilities inherent in the modeling process, not on the objective frequencies of the events (Sinkkonnen, 2000). The information or 'surprise' inherent in an event is based on the probability of that event. Clearly for a system to encode events efficiently it must know these probabilities or infer them on the basis of experience (i.e. how frequently they occur). The brain may embody these probabilistic regularities through plastic changes in development (e.g. language) or in a context-specific form (e.g. the 'oddball' paradigm). Our study represents an example of the latter. To compute surprise and entropy we treated each subject as an ideal observer and used the Bayesian posterior probability of an event, given the history of trials within a block. Entropy and surprise were calculated from posterior or conditional probabilities on a trial-by-trial basis and used to predict neuronal responses. These metrics are 'inferred' on the basis

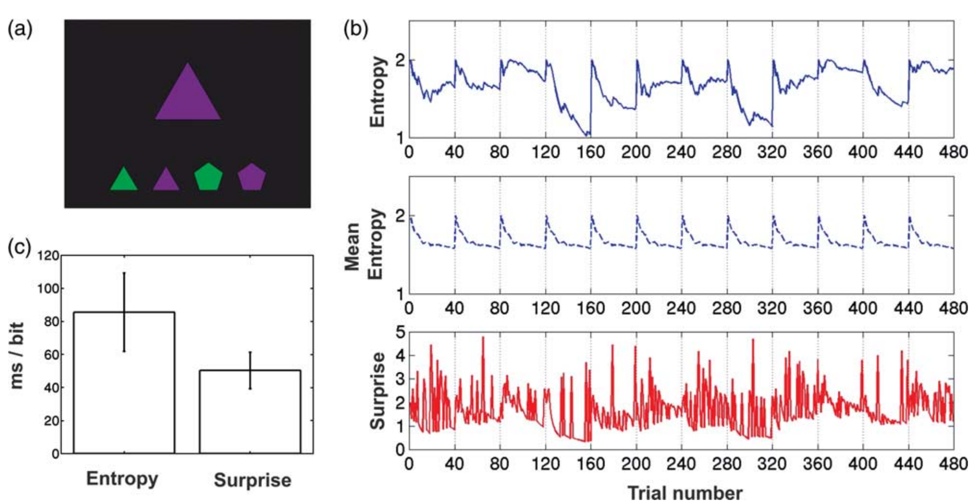

Fig. 1. Task design. (a) The choice reaction time paradigm. Subjects were required to respond to the sampled item (in this case a purple triangle) by pressing the key to indicate the position of that item in the row of alternative coloured shapes (below). (b) Information theoretic quantities for a typical scanning session. Entropy, mean entropy (averaged across all blocks) and surprise are plotted (units: bits). Dotted vertical lines divide successive blocks. A fixation cross was presented for 20 s between blocks. (c) Behavioral data. Increase in reaction time per bit of entropy and surprise (±SE of the mean of 12 subjects). Mean RT across all trials: 550.2 (±15.0) ms.