Forgetful Inference in a Sophisticated World Model

Sanjeevan Ahilan (ahilan@gatsby.ucl.ac.uk)
Gatsby Computational Neuroscience Unit, 25 Howland Street
London, United Kingdom, W1T 4JG

Rebecca B. Solomon, **Kent Conover**, **Ritwik K. Niyogi**, **Peter Shizgal**, **Peter Dayan**

Abstract

Humans and other animals are able to discover underlying statistical structure in their environments and exploit it to achieve efficient and effective performance. However, the largest scale structures such as 'world models' are often difficult to learn and use because they are obscure, involving long-range temporal dependencies. We analyzed behavioral data from an extended experiment with rats, showing that the subjects discovered and exploited a world model, albeit suffering at times from immediate inferential imperfections as to their current state within it. To describe this process, we built a hidden Markov model (HMM) of the subjects' models of the experiment, describing overall behavior as integrating recent observations with the recollections of an imperfect memory. Over the course of training, we found that subjects came to track their progress through the task more accurately, indicating improved inference of the partially-observable state. Model fits attributed this improvement to decreased forgetting of the previous state. This 'learning to remember' decreased reliance on more recent observations, which can be misleading, in favor of a more dependable memory.

Keywords: latent state model; hidden Markov model; brain stimulation reward; learning to learn; partial observability

The natural world is replete with statistical structure which may be extracted by animals in the form of world models (Daw et al., 2006) supporting predictions of future states and demands. When this structure involves long term regularities, immediate observations are insufficient for determining the current state and thus accurate prediction depends on memory (Zilli & Hasselmo, 2008). We investigate the ability of rats to build world models and use memory effectively to support inference in such environments.

Results

We analysed data from a cumulative handling time task in which rats hold down a lever for an experimenter-defined time period, called the price, in return for rewarding electrical stimulation of the medial forebrain bundle at a given frequency. In this paradigm, a trial consists of a duration of fixed price and frequency in which subjects may balance working for reward with leisure breaks as they see fit. Trials last a duration of 25 times the price (except for a minority of trials with price less than 1 second lasting 25 seconds).

Trials come in a predictable cyclic triad consisting of 'lead', 'test' and 'trail' trial types (Fig. 1A). These are associated with different values of frequency and price (Fig. 1B) and thus variable amounts of work over the duration of a trial (Fig. 1C). Lead trials involve fixed, high frequency, stimulation with a short price of 1 second. They are sufficiently rewarding that subjects typically work the entire duration of the trial. Trail trials involve fixed, low frequency, stimulation with the same short price of 1 second. These are negligibly rewarding and so rats barely work. Test trials involve a range of frequencies and prices which change from trial to trial (but are fixed across a particular trial). Work on test trials varies depending on the particular values of the frequency and price.

Each trial starts identically, with a prime (a high frequency stimulation pulse) signalling its beginning. The subjects are then free to choose whether and when to engage with the lever. We analyze the engagement probability (EP) and, given this, the initial response times (IRTs) which quantify how long this takes. These are a pure measure of the subjects' beliefs about trial type prior to obtaining any within-trial information. They thus reflect subjects understanding of the triadic structure and their place within it.

We studied six rats, each of which had experienced approximately 1500 triads of trials. EPs and IRTs showed their knowledge of the triadic task structure (Fig. 1D, upper histograms) – to a first approximation, the larger the expected value of a trial, the greater the chance of engagement and the shorter the IRT. Subjects responded on almost all lead and test trials, with shorter IRTs for the highly rewarding lead trials than for the less, and less certainly, rewarding test trials. Subjects responded at all on only a fraction of the relatively worthless trail trials; when they did, their IRT distributions were bimodal, taking on both short and long values.

Since the trail trials offer nugatory reward, short IRTs might seem surprising. We argue that these cases are examples of confusion, in which the subject has misidentified the trail trial as either a lead or a test trial. By sorting the trail trial IRTs by the frequency and price of the previous test trial (Fig 1D, lower histograms), we found that this misidentification was most likely to occur when that test trial might plausibly have been seen as a lead or trail trial itself, implying by the triad structure that the trail trial would be a test or lead trial respectively, and so meriting a short IRT. For test trial frequency-price combinations dissimilar to those of either lead or trail trials, subjects were rarely confused, and so short IRTs occurred much more rarely.

The predominant regularities suggest that the subjects have learned the triadic structure. We describe the remaining confusion by building and performing inference on an HMM of the rat's model of the experiment (Fig. 1E). Given that lead trials are unambiguous, we assumed that subjects were initially certain when they were in a test trial. This is consistent with their consistently short IRTs. Then, over the duration of the test trial, the subject could forget the trial type, even as it experienced the current trial. By the end of the test trial, imperfect memory is combined with recent observations to generate the rat's posterior belief about the trial type. Applying the task transition matrix to subject's belief state leads to a belief for the trail trial. We then simulated or calculated the probabilities of the

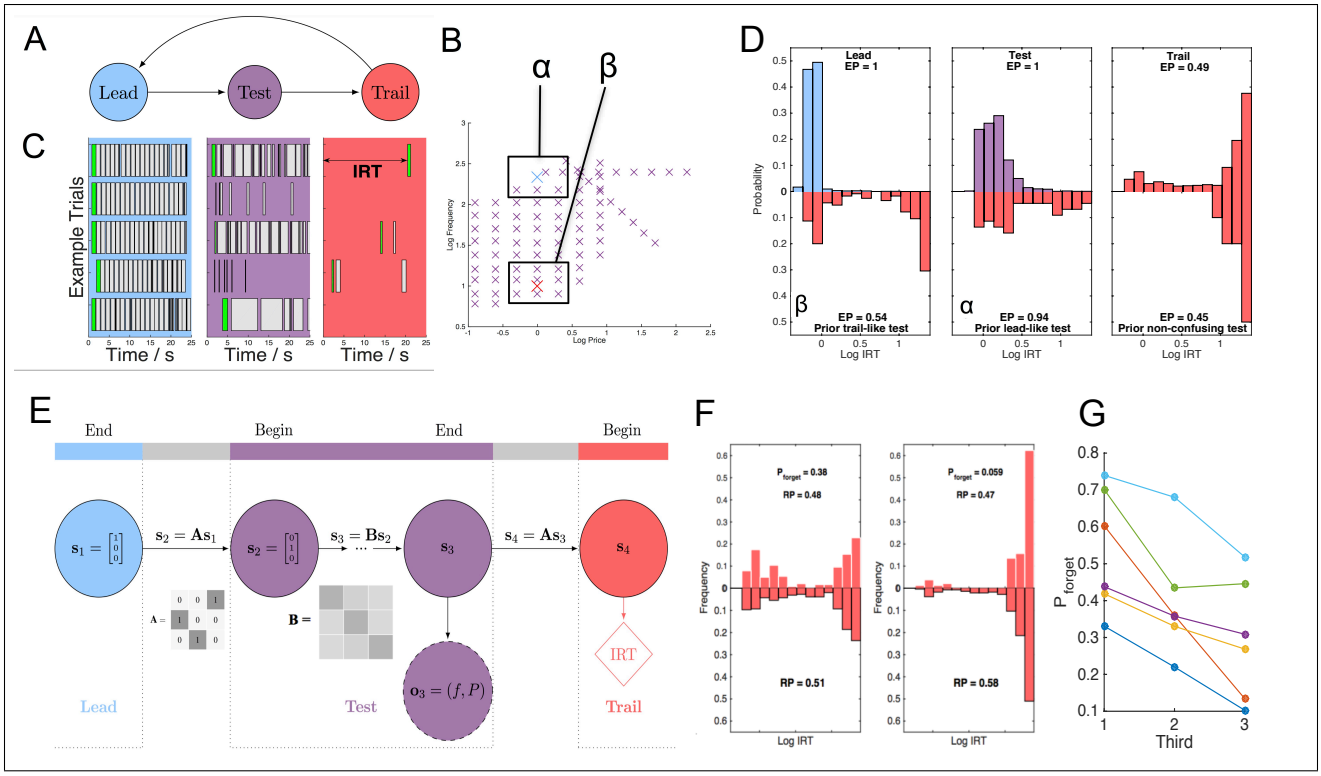

Figure 1: **(A)** Trials comes in a predictable cyclic triad. **(B)** Frequency and price for lead, test and trail trials. For test trials this varies from trial to trial. **(C)** Five example trials per type (initial responses in green). Pressing is almost continuous (lead), variable (test; only first 25s shown) and rare (trail). **(D)** IRT distributions from one subject. *Upper histograms*: Response probabilities (EPs) are 1 for lead (left) and test (middle) trials; IRTs are generally short. Trail trials (right) have a lower EP; and the IRT distribution is bimodal. *Lower histograms*: Trail trial IRTs sorted by previous test price and frequency. Left: lead-like trail IRTs follow test trials similar to trail trials (region β, Fig. 1B). Middle: test-like trail IRTs follow test trials similar to lead trials (region α, Fig. 1B). Right: when trail trials follow non-confusing test trials, short IRTs are much less frequent **(E)** The Model. At the beginning of a test trial, the subject is certain of the trial type; but over the duration of the trial can forget, as captured by transition matrix B (with off-diagonal elements $P_{\text{forget}}/3$). Evidence from more recent observations of frequency and price/duration are integrated with this memory to form a posterior at s_3. The A matrix describes the perfect transition to the subsequent trial and an associated IRT. **(F)** Example subject. *Lower histograms*: Real data. The large fraction of short IRTs on trail trials in the first third of data (left) greatly decreased for the final third (right), reflecting improved inference. *Upper histograms*: Simulated data using fitted parameters closely match the data. **(G)** Across all 6 rats fitted values of P_{forget} decreased from first to last thirds of trials, suggesting that their improved inference is a result of improved usage of working memory.

observed responses, using non-parametric fitting of lead, test and non-confusing trail trial IRTs.

The model uses two parameters. One is a variance parameter which specifies a kernel density estimate for the distribution of an observation in frequency-price space given the trial type (we can also use trial duration instead of price due to the high degree of correlation). The second parameter specifies a 'probability of forgetting', which describes the fidelity of a rat's memory of the previous trial type.

To analyse the data we divided it sequentially into thirds for each subject. In the final third, the fraction of short IRTs was greatly decreased, indicating an improved ability of the rats to track their progress through the task (Fig. 1F, lower histograms). By fitting model parameters independently to each third using maximum likelihood estimation we closely matched the observed distribution of IRTs as well as the response probabilities (Fig. 1F, upper histograms) for all subjects. To account for the change in the distribution of IRTs, the parameter fits consistently identified a lower probability of forgetting in the last third relative to the first third (Fig. 1G). This suggests that over time the rats learn to remember the previous trial type better and so improve their identification of the current trial type.

Acknowledgments

This work was supported by the Gatsby Charitable Foundation and the Medical Research Council (MRC). P.S. received funding from the Natural Sciences and Engineering Research Council of Canada grant and Concordia University Research Chair (Tier I).

References

Daw, N. D., Courville, A. C., & Touretzky, D. S. (2006). Representation and timing in theories of the dopamine system. *Neural computation*, *18*(7), 1637–1677.

Zilli, E. A., & Hasselmo, M. E. (2008). The influence of markov decision process structure on the possible strategic use of working memory and episodic memory. *PloS one*, *3*(7), e2756.