Reward probability encoding in human neurons

T Alexander Price (alexander.price@neuro.utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Rhiannon L Cowan (rhiannon.cowan@utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Tyler S Davis (tyler.davis@utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Shervin Rahimpour (Shervin.rahimpour@hsc.utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Ben Shofty (ben.shofty@hsc.utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Matthew M Botvinick (botvinick@google.com)
Deepmind, 5 New Street Square
London, WC1N 3BG, UK

Timothy H Muller (timothymuller127@gmail.com)
Institute of Neurology, Department of Clinical and Movement Neurosciences, University College London,
London WC1N 3BG, UK

Steven W Kennerley (steven.kennerley@psy.ox.ac.uk)
Department of Experimental Psychology, Mansfield Road
Oxford, OX1 3SR, UK

John D Rolston (jrolston@bwh.harvard.edu)
Department of Neurosurgery, 60 Fenwood Road,
Boston, MA 02115 USA

Elliot H Smith (e.h.smith@utah.edu)
Neurosurgery Department, 175 N Medical Dr.,
Salt Lake City, UT 8410 USA

Abstract:

People routinely make decisions with uncertain outcomes. Economists have defined this uncertainty about a possible outcome as risk. Previous studies have found subcortical neural computations related to reward and risk. The prefrontal cortex has also been shown to play a role in these computations. Here we present evidence of computations underlying reward and risk in human prefrontal and temporal cortices. These representations may be signatures of Distributional Reinforcement Learning, a framework by which the brain makes value-oriented predictions and updates those predictions to make decisions.

Keywords: distributional reinforcement learning; value-based decisions; risk processing; human neuronal computation.

Introduction

Economists have long-since tried quantifying Risk, Value, and Utility and to develop mathematical models that might explain or predict human behavior (Glimcher, 2008). These models iterated from Pascal's ideas about Expected Value to the development of Prospect Theory (Kahneman and Tversky, 1979). This work led to the definition of risk as a measurement of uncertainty around a reward (Platt & Huettel, 2008). Previous work has shown that subcortical dopaminergic areas in the brain are the areas responsible for computations concerning reward probability (Cohen et al., 2012; Starkweather & Uchida, 2021). In an effort to elucidate the algorithms responsible for reward learning under uncertainty, Dabney et al. found a modification to Reinforcement Learning that would allow neurons to represent a distribution of rewards (Dabney et al., 2020). This Distributional Reinforcement Learning (DistRL) was first identified in the mouse ventral tegmental area (VTA). More recent work has shown that cortical neurons in non-human primates (NHPs) also encode a distribution of predictions about reward and the associated error of those predictions ((Muller et al., 2024). Additionally, neuroimaging studies have highlighted the anterior insula, dorsal striatum, and the noradrenergic system, as areas where reward is encoded quadratically (Preuschoff et al., 2006, 2008, 2011). This neural encoding is often considered an encoding of risk and is distinct from linear encoding of reward (Presuchoff et al., 2006, 2008, 2011). Here we sought to discover how neurons in these areas of humans' brains represented reward probability and uncertainty during risky choices. To do so, we measured the activity of single units from human neurosurgical patients while they carried out the Balloon Analog Risk Task (BART; Lejuez et al., 2002). We discovered units that encode reward (linearly), and risk (quadratically) as a function of reward probability with a majority of those neurons reversing predictive encoding of reward probability upon reward outcome.

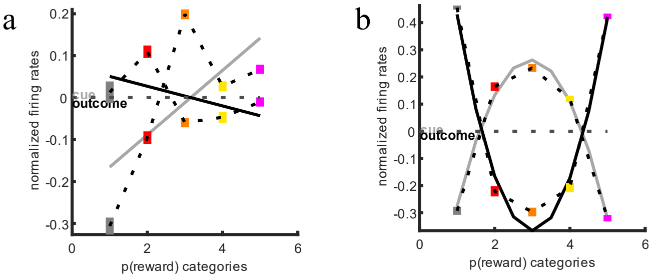

Figure 1: Reward and risk encoding examples. a, mean firing rates across the full two-second window and linear fit for cue (gray) and outcome (black) -aligned firing for an example neuron recorded in MTL. **b**, same as a, but for quadratic fit. An example neuron from MTL.

Methods

Single neurons were recorded from patients who were undergoing neuromonitoring for treatment of drug-resistant epilepsy using Behnke-Fried microwires extending from the distal tip of two to three of the patients' clinical macroelectrodes (Misra et al., 2014). During each BART trial, patients press a button to begin inflating a computerized balloon. They must push the same button to stop the balloon's inflation. If the inflation is stopped prior to the balloon popping, patients receive points linearly related to the size of the balloon. If the balloon pops, the patients neither receive nor lose any points. BART contains both active, wherein the subject is responsible for stopping the balloon to get points, and passive trials, wherein the balloon automatically inflates to its maximum size. There are five reward categories of balloons: gray (unrewarded passive trials), yellow, orange, red (rewarded active trials) and what are represented as pink (rewarded passive trials). The color of balloon, and the presence/absence of an indicator of a passive trial cued patients to the potential for reward on each trial. Single units were isolated by bandpass filtering the signal from the microwires between 0.25 and 7.5 kHz and sorting waveforms that crossed -3.5 times the mean root squared of the filtered signal using Offline Sorter (Plexon, Inc.; Dallas, TX). Firing rates were examined in two-second windows, 0.25 seconds following the appearance of the balloon, and following reward outcome. These firing rates were modeled as linear monotonic and quadratic functions of reward probability categories using generalized linear models and significance was assessed using ANOVAs on each model term (Fig 1). Best-order polynomials were fit to mean firing rates for each reward probability category to visualize reward probability tuning curves across local neuronal populations. Finally, the reversal point (RP) for each significant unit was calculated at cue and outcome. For units that had significant linear encoding, the RP was calculated by finding the value of the abscissa where the best-fit line crossed the ordinate axis, which represented the reward probability when the unit had a normalized firing rate of zero. If a linear model was the best fit but the normalized firing rate crossed the threshold of zero normalized firing rate more than once, an average of the values at which the ordinate axis was crossed was used to calculate the RP. For units with significant quadratic encoding, the RP was considered the value of the abscissa that corresponded to the extrema of the best-fit quadratic. Next, the asymmetric scaling for units with significant linear encoding of reward probability was calculated. This was done in a manner similar to the calculations done in mouse VTA as well as NHP cortex where the scaling value that was considered was a ratio of positive and negative RPEs (Dabney et al., 2020; Muller et al., 2024).

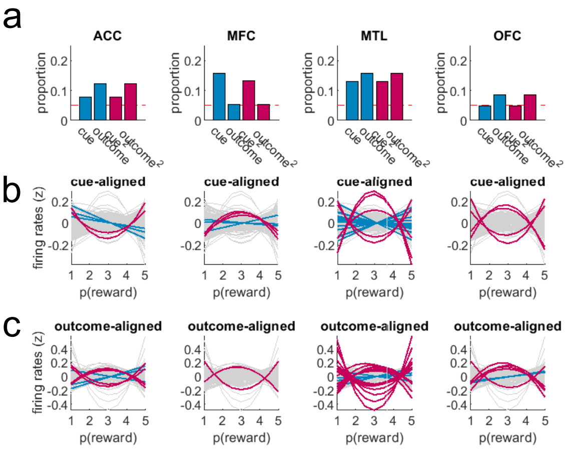

Figure 2: Risk encoding examples. a, proportions of neurons that significant encode reward probability linearly (blue) and quadratically (red) or each of the four brain areas. **b**, fitted cue-aligned mean firing rates across reward probability categories for all units in each brain area. **c**, fitted outcome-aligned mean firing rates across reward probability categories for all units in each brain area. Gray lines in **b** and **c** show insignificant best-fits.

Results

Thirty-two human participants (18 female) carried out a mean±s.d. of 234.6±28.9 trials of BART with a mean±s.d accuracy of 83.2±6.1.

We recorded from 334 well-isolated units during the BART task. These recordings were grouped into four anatomical areas: The Orbitofrontal Cortex (OFC; 83 units), The Medial Frontal Cortex (MFC; 39 units), the Anterior Cingulate Cortex (ACC; 66 units), and the Mesial Temporal Lobe (MTL; 146 units). By fitting generalized linear models to the firing rates of these areas, we found significant proportions of units that monotonically and quadratically encoded reward probability in each of these areas. Across units, we found significant encoding of reward probability in most brain areas. OFC was the only brain area in which significant proportions of neurons were not found to encode reward probability at cue (binomial test p > .05) (Fig 2). The majority of neurons that significantly predicted the reward probability categories in response to the cue, exhibited a reversal of their reward probability encoding in response to the outcome (60% of units in ACC, 50% in MFC, 74% in MTL, and 50% in OFC). Each of the brain areas had significant quadratic encoding of reward probability at cue and outcome, indicating a significant encoding of risk.

There are two signatures of DistRL tested in these analyses. Firstly, reward-encoding units are expected to have a distribution of reversal points, which allows for a range of predictions and updated expectations of a reward. The second prediction is that these reward-encoding neurons will scale reward prediction errors asymmetrically. We found correlates of DistRL in both linear and quadratic encoding of reward probability, as these reversal points were distributed and were not all a single value. For units with significant linear encoding of reward probability, we also found diverse scaling of reward prediction errors. The ratio of betas for positive reward prediction errors and negative prediction errors was also widely spread over a non-normal distribution.

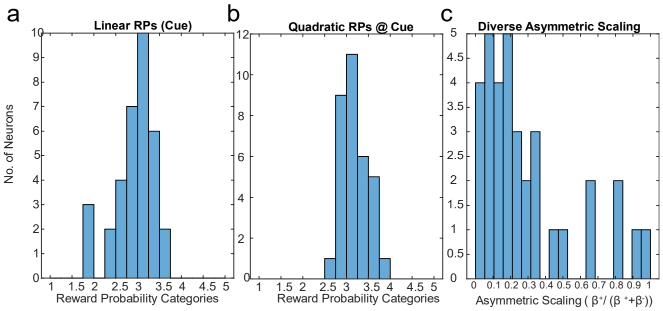

Figure 3 Signatures of DistRL: a, Distribution of reversal points for units when the best fit polynomial of the normalized firing rates was linear. **b**, Distribution of reversal points for units the best fit polynomial was quadratic. **c**, Distribution of asymmetric scaling for units that had significant linear encoding.

Discussion

We studied encoding of predicted reward and economic risk in single neuron activity recorded from human prefrontal and temporal lobes. We found signatures of computations that have previously been identified in subcortical structures in the human cortex. The majority of these neurons changed the direction of encoding following presentation of reward, which supports the idea that these neurons also encode prediction error, a fundamental variable in Reinforcement Learning. Future work will continue to identify additional DistRL signatures like the optimism and pessimism of individual units as well as asymmetric learning rates.

Acknowledgments

We gratefully acknowledge support from a grant from the National Institute of Mental Health: R01MH128187.

References

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B.B>, & Uchida, N. (2012). Neuron-type-specific signals for reward and punishmen tin the ventral tegmental area. *Nature,* 487(7383), 85-88. https://doi.org/10.1038/nature10754

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*, *577*(7792), Article 7792. https://doi.org/10.1038/s41586-019-1924-6

Glimcher, P. W. (2008) "Understanding Risk: A Guide for the Perplexed." *Cognitive, Affective, & Behavioral Neuroscience* 8, no. 4 348–54. https://doi.org/10.3758/CABN.8.4.348.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. JSTOR. https://doi.org/10.2307/1914185

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. https://doi.org/10.1037//1076-898X.8.2.75

Misra, A., Burke, J., Ramayya, A., Jacobs, J., Sperling, M., Moxon, K., Kahana, M., Evans, J., & Sharan, A. (2014). Methods for implantation of micro-wire bundles and optimization of single/multiunit recordings from human mesial temporal lobe. *Journal of Neural Engineering*, *11*(2), 026013. https://doi.org/10.1088/1741-2560/11/2/026013

Muller, T. H., Butler, J. L., Veselic, S., Miranda, B., Wallis, J. D., Dayan, P., Behrens, T. E. J., Kurth-Nelson, Z., & Kennerley, S. W. (2024). Distributional reinforcement learning in prefrontal cortex. *Nature Neuroscience*. https://doi.org/10.1038/s41593-023-01535-w

Platt, M. L., & Huettel, S. A. (2008). Risky business: The neuroeconomics of decision making under Uncertainty. Nature Neuroscience, 11(4), 398–403. https://doi.org/10.1038/nn2062