Computing uncertainty

Uncertainty arises from having more than one option, and that the motivation to opt for one of those options is somewhat distributed, and there is no one option that is always prefered. Considering that the probability of choosing any given option has a uniform distribution, then uncertainty increases proportionally with the number of options. Shannon entropy (Shannon 1948) formalizes this intuition

$$H = -\sum_{i=1}^{n} p_i log_2 p_i$$

So maximum entropy (one bit) is achieved when all the alternatives have the same probability, such as a coin flip. However, if the coin happens to have two heads, then Shannon entropy is 0.

If we consider a simple environment with only one state s, one action a initiate a food-seeking bout, and only two possible outcomes food is found (p) or not found ($q = 1 - p$), then $H = -(plog_2 p + qlog2q)$. If an animals performs multiple food-seekign bouts and non of them are succesful $H = 0$ the same is true if all are successful. However, if the probability of a successful food-seeking bout is 0.5, then entropy is maximized $H = 1$. Neural representation of entropy has been found in the middle cingulate cortex (MCC) for the particular implementation of encoding outcome entropy (Goñi et al. 2011; Gloy, Herrmann, and Fehr 2020) so this computation seems to be biologically plausible. However, entropy is not available as sensory input it must derive from actions and outcomes, which are dependent on environment state. Previously, through Thompson sampling, we provided a way in that entropy could be enconded as variance in the posterior distribution, nevertheless, a more direct way to compute entropy is possible through the prediction error.

The classical model of Rescorla-Wagner (Rescorla et al. 1972) modeled how animals could predict the reinforcing value of a given stimuli

$$y_t = y_{t-1} + \alpha \delta_n$$

where the value representation of the stimuli y is obtained by considering the previously estimated value $y_{t-1}, but weighted by a learning rate α and a prediction error δ. *delta* is the simple difference between the expected reward and the actual reward $\delta_n = r_t - y_t$ where r_t is the obtained reward, an extension to this has proposed by (Sutton and Barto 2018) where the prediction error consider an estimate of the rewards that give more weight to current rewards while still considering past rewards

$$\delta_n = Rt + \gamma \hat{V_{n+1}} - \hat{V}_n$$

γ is a discount factor $0 <= \gamma < 1$ for all the history of rewards, and $\hat{V}$ is a proxy for the true value of the reward. Finally, $\alpha : [0, 1]$ is the learning rate which effectively weights the reward prediction error δ so to make small updates $\alpha \approx 0$ or rather large ones $\alpha \approx 1$.

The simple model presented allows to derive a prediction error based on experience, and the learning rate can be set lower to simulate unexpected uncertainty or higher to simulate expected uncertainty. However, α in such models is a hyperparameter, thus not derived from experience. Pearce and Hall (1980) model proposes that α can be controlled by the prediction error magnitude $|\delta$ so

$$\alpha = \gamma|\delta_{n-q}| + (1-\gamma)\alpha_{n-1}$$

Higher entropy on reward outcomes increases the minimal error probability (Feder and Merhav 1994), thus increasing $|\delta|$, and consequently α. The behavioral result is that the animal should increase learning for options with uncertain outcome by directing its attention (Diederen and Fletcher 2021).

Dopamine encodes prediction error (Nasser et al. 2017; C. D. Fiorillo 2003; C. D. Fiorillo 2011; Lak, Stauffer, and Schultz 2014; Glimcher 2011; Khaw, Glimcher, and Louie 2017; Gershman and Uchida 2019)

Dynamically adjust learning rate in face of uncertainty.

the uncertainty bonus.

uncertainty increases food-seeking behavior.

dopamin, ach, NE. Hint orexin.

Diederen, Kelly M. J., and Paul C. Fletcher. 2021. "Dopamine, Prediction Error and Beyond." *The Neuroscientist* 27 (1): 30–46. https://doi.org/10.1177/1073858420907591.

Feder, M., and N. Merhav. 1994. "Relations Between Entropy and Error Probability." *IEEE Transactions on Information Theory* 40 (1): 259–66. https://doi.org/10.1109/18.272494.

Fiorillo, C. D. 2003. "Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons." *Science* 299 (5614): 1898–1902. https://doi.org/10.1126/science.1077349.

Fiorillo, C.D. 2011. "Transient Activation of Midbrain Dopamine Neurons by Reward Risk." *Neuroscience* 197 (December): 162–71. https://doi.org/10.1016/j.neuroscience.2011.09.037.

Gershman, Samuel J., and Naoshige Uchida. 2019. "Believing in Dopamine." *Nature Reviews Neuroscience* 20 (11): 703–14. https://doi.org/10.1038/s41583-019-0220-7.

Glimcher, P. W. 2011. "Understanding Dopamine and Reinforcement Learning: The Dopamine Reward Prediction Error Hypothesis." *Proceedings of the National Academy of Sciences* 108 (Supplement_3): 15647–54. https://doi.org/10.1073/pnas.1014269108.

Gloy, K., M. Herrmann, and T. Fehr. 2020. "Decision Making Under Uncertainty in a Quasi Realistic Binary Decision Task – An fMRI Study." *Brain and Cognition* 140 (April): 105549. https://doi.org/10.1016/j.bandc.2020.105549.

Goñi, Joaquín, Maite Aznárez-Sanado, Gonzalo Arrondo, María Fernández-Seara, Francis R. Loayza, Franz H. Heukamp, and María A. Pastor. 2011. "The Neural Substrate and Functional Integration of Uncertainty in Decision Making: An Information Theory Approach." Edited by Matjaz Perc. *PLoS ONE* 6 (3): e17408. https://doi.org/10.1371/journal.pone.0017408.

Khaw, Mel W., Paul W. Glimcher, and Kenway Louie. 2017. "Normalized Value Coding Explains Dynamic Adaptation in the Human Valuation Process." *Proceedings of the National Academy of Sciences* 114 (48): 12696–701. https://doi.org/10.1073/pnas.1715293114.

Lak, Armin, William R. Stauffer, and Wolfram Schultz. 2014. "Dopamine Prediction Error Responses Integrate Subjective Value from Different Reward Dimensions." *Proceedings of the National Academy of Sciences* 111 (6): 2343–48. https://doi.org/10.1073/pnas.1321596111.

Nasser, Helen M., Donna J. Calu, Geoffrey Schoenbaum, and Melissa J. Sharpe. 2017. "The Dopamine Prediction Error: Contributions to Associative Models of Reward Learning." *Frontiers in Psychology* 8 (February). https://doi.org/10.3389/fpsyg.2017.00244.

Pearce, John M., and Geoffrey Hall. 1980. "A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned but Not of Unconditioned Stimuli." *Psychological Review* 87 (6): 532–52. https://doi.org/10.1037/0033-295X.87.6.532.

Rescorla, Robert A, Allan R Wagner, Abraham H Black, and William Frederick Prokasy. 1972. "Classical Conditioning II: Current Research and Theory."

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction.* Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.