A Perceptual Confirmation Bias from Approximate Online Inference

Richard D. Lange[1], Ankani Chattoraj[1], Matthew Hochberg[1], Jeffry M. Beck[2], Jacob Yates[1], Ralf M. Haefner[1]

1. Brain and Cognitive Sciences, University of Rochester, 2. Department of Neurobiology, Duke University

Introduction

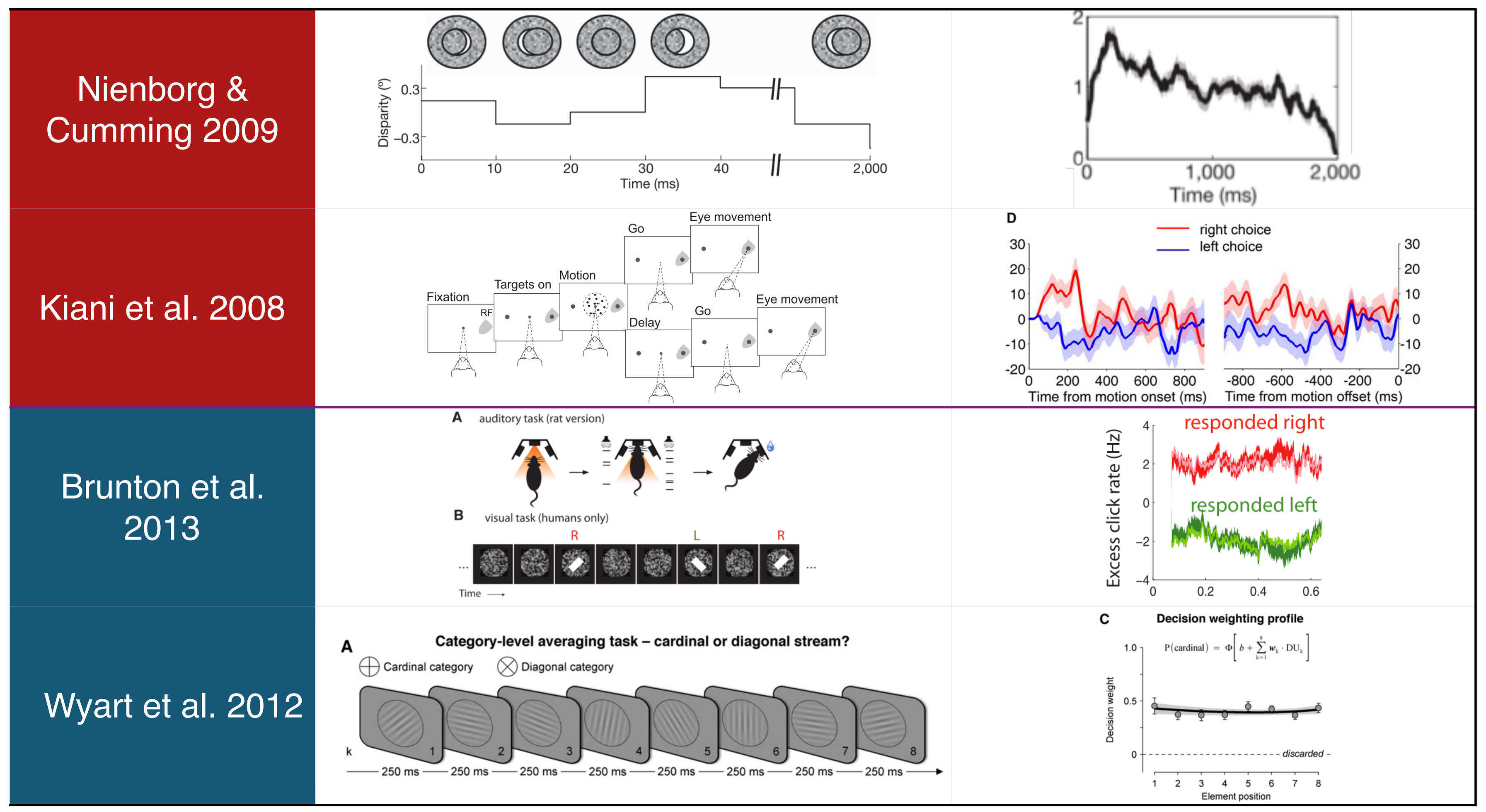
In **evidence integration** tasks, subjects make a categorical decision from a sequence of (typically i.i.d.) sensory information.

A **psychophysical kernel (PK)** quantifies the 'weight' subjects give to evidence in space or in time.

A **confirmation bias (CB)** occurs when people upweight information confirming existing beliefs, thus strengthening those beliefs.

A **Perceptual CB** implies a PK that decreases over time.

Different studies have reported different temporal PK shapes, typically flat or decreasing.



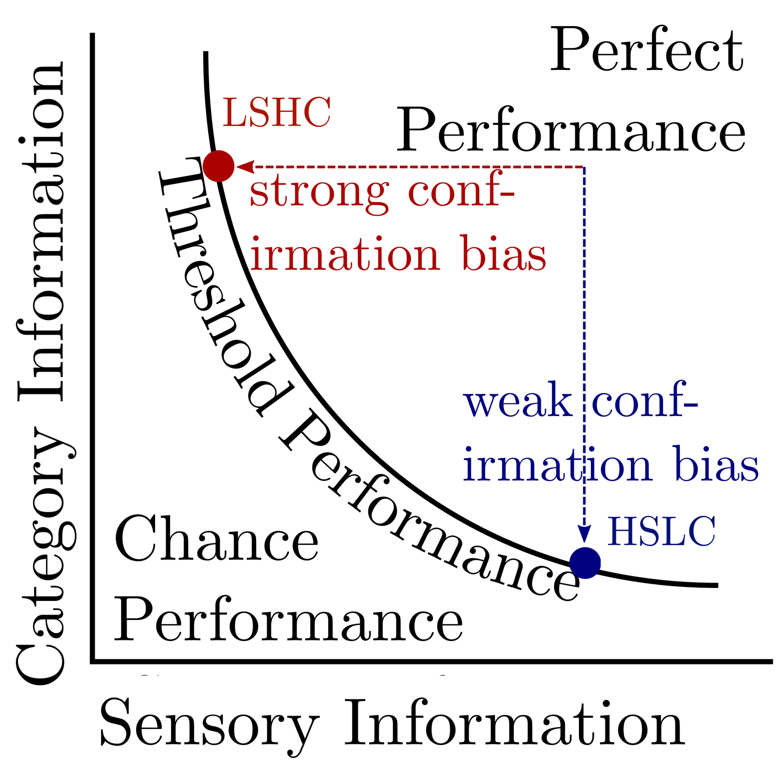
Our Framework

2 Sources of uncertainty:

With **high-contrast stimuli** that are each **weakly predictive** of the correct choice, recency (or flat weights) observed

With **low-contrast stimuli** that are each **highly predictive** of the correct choice, primacy is observed.

Threshold performance is achieved at a balance between these.



Category information: probability of an ideal observer guessing the category of a single 'frame' $\mathbf{x}_t$ given **C**

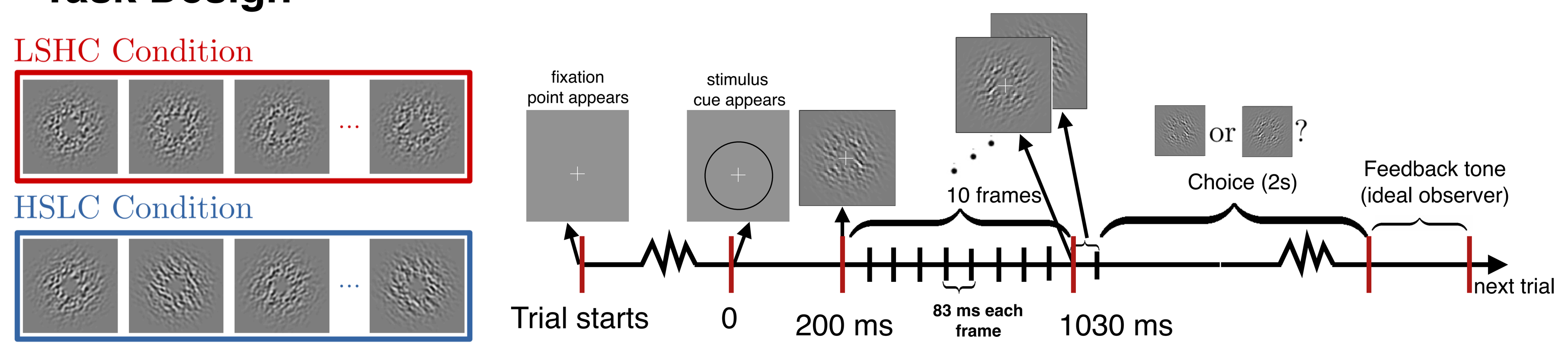
We define **sensory** or **likelihood information** as the probability of guessing $\mathbf{x}_t$ given $\mathbf{e}_t$

Experiments

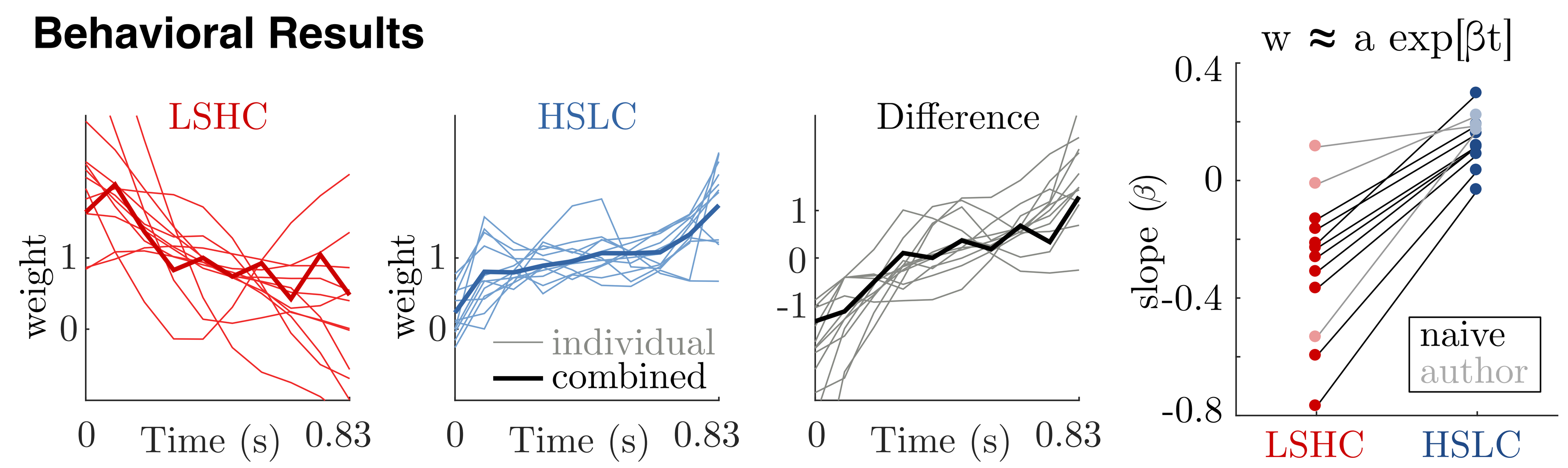
We used an orientation-discrimination paradigm with orientation-band-pass stimuli.

- In the **LSHC context** staircase on the **noise** is run to reach 70% performance
- In the **HSLC context**, a staircase is run on $\mathbf{p}_{\text{prior}}$
- Both tasks' staircases begin at the same set of parameters.

Task Design



Behavioral Results

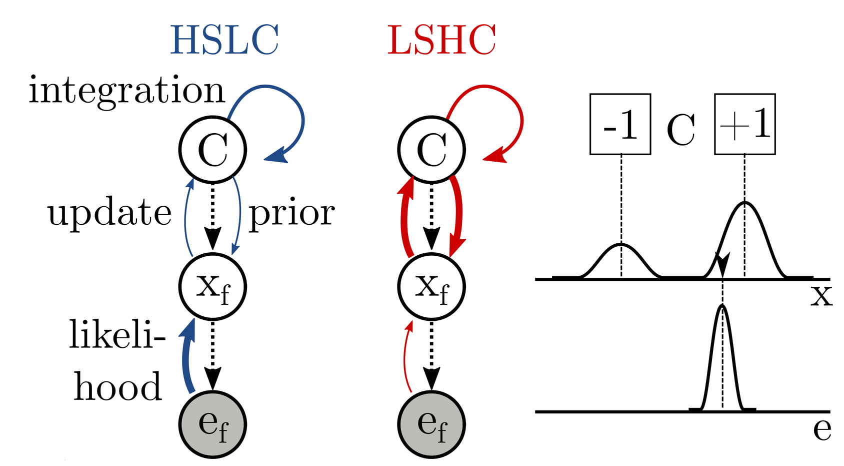


Change in PK slopes consistent with our framework's predictions, but significant variability between subjects (possibly explained by different γ?).

Sampling Model[5,6]

Generative model:

- **C** = category / decision-area
- **x** = sensory representation
- **e** = evidence



Goal: compute posterior over **C** given **e**

$$p(C|e_1,\ldots,e_T) \propto p(C)\prod_{t=1}^{T} p(e_t|C)$$

...using **online updates**

$$\log\frac{p_t(C=+1)}{p_t(C=-1)} \equiv \log\frac{p(C=+1|e_1,\ldots,e_t)}{p(C=-1|e_1,\ldots,e_t)} \qquad p_0(C)\equiv p(C)$$

$$= \log\frac{p_{t-1}(C=+1)}{p_{t-1}(C=-1)} + \log\frac{p(e_t|C=+1)}{p(e_t|C=-1)}$$

update to log posterior odds each frame

...using **importance sampling** from the **full posterior** to marginalize over the sensory variable **x**

$$p(e_t|C=c) = \int p(e_t|x_t)p(x_t|C=c) \approx \frac{1}{S}\sum_{x^{(i)}\sim Q} p(e_t|x_t^{(i)})p(x_t^{(i)}|C=c)/Q(x_t^{(i)})$$

$$\log\frac{p(e_t|C=+1)}{p(e_t|C=-1)} \approx \log\frac{\sum p(x_t^{(i)}|C=+1)w_i}{\sum p(x_t^{(i)}|C=-1)w_i}$$

$$Q(x) \propto \sum_c p(e_t|x)p(x|C=c)p_{t-1}(C=c)$$

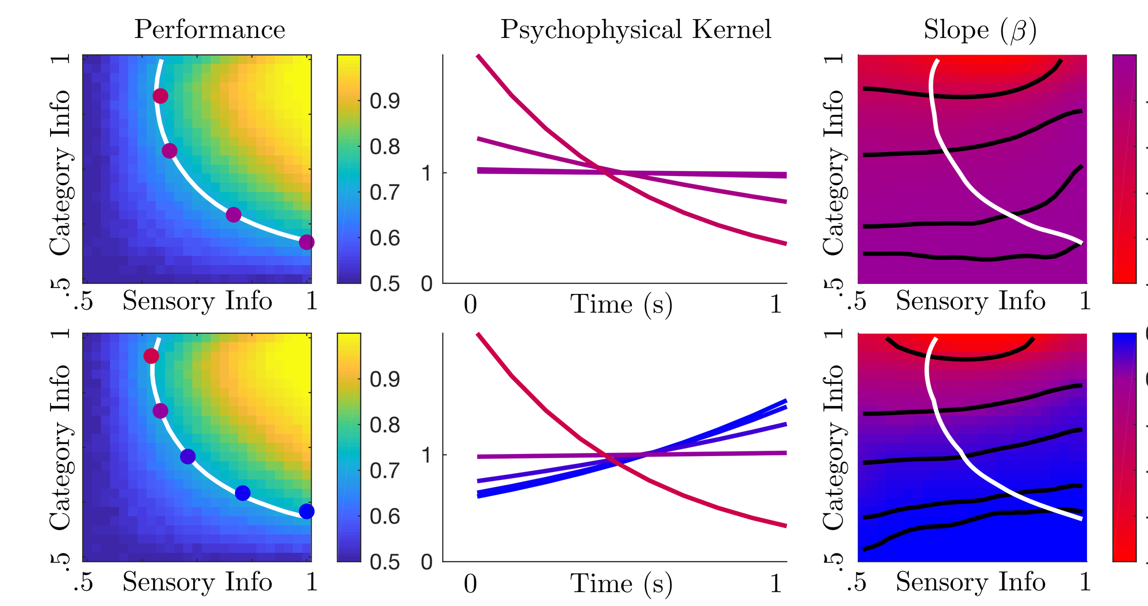
$$w_i = \left(\sum_c p(x_t^{(i)}|C=c)p_{t-1}(C=c)\right)^{-1}$$

Final update rule:

$$\log\frac{p_t(C=+1)}{p_t(C=-1)} \approx \log\frac{p_{t-1}(C=+1)}{p_{t-1}(C=-1)} + \log\frac{\sum_{i=1}^{S} p(x_t^{(i)}|C=+1)w_i}{\sum_{i=1}^{S} p(x_t^{(i)}|C=-1)w_i} - \underbrace{\gamma\log\frac{p_{t-1}(C=+1)}{p_{t-1}(C=-1)}}_{\text{bias correction (small } S)}$$

Sampling Model gamma = 0

Sampling Model gamma = 0.1



Variational (Parametric) Model[7,8]

Same setup and objective as sampling model, except approximation is due to **mean field** assumption:

$$p(C,x|e) \approx q(C)q(x)$$

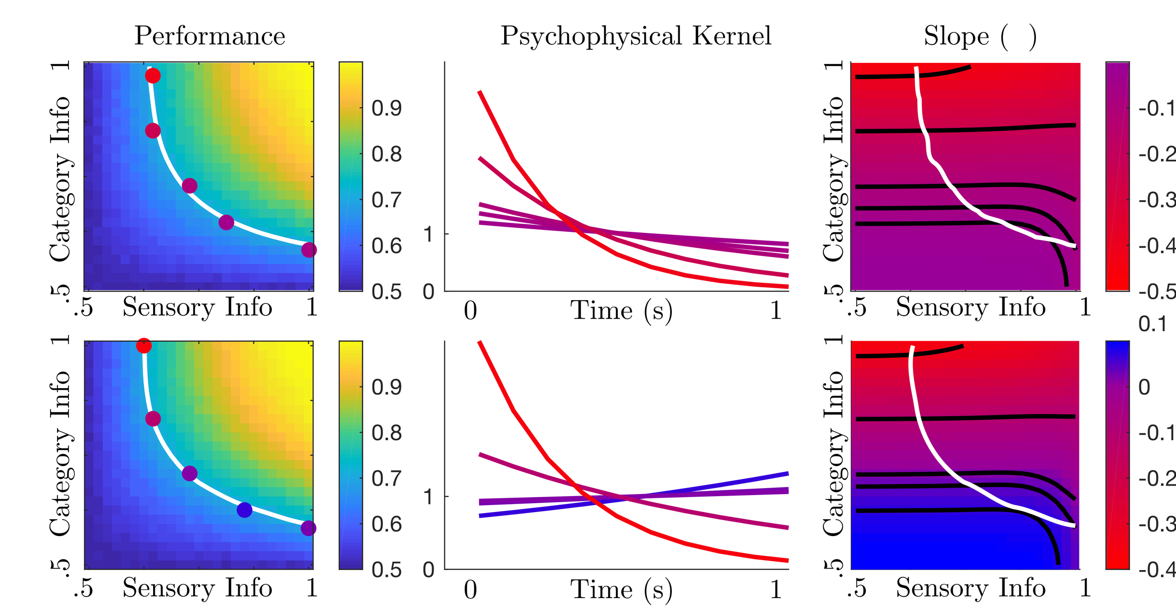
Inference is be done using Mean Field Variational Bayes, which passes messages between q(C) and q(x) to approach a minimum of KL(q||p). Here, the updates (with an auxiliary variable z) are:

$$\log q(C) = \log p(C) + \mu_z\mu_x/\sigma_x^2 \qquad \log q(z) = \log p(z) + \mu_C\mu_x/\sigma_x^2$$

$$q(x) = \mathcal{N}\left(x; \frac{\sigma_e^2\mu_C + \sigma_x^2\mu_z e}{\sigma_x^2+\sigma_e^2}, \frac{\sigma_x^2\sigma_e^2}{\sigma_x^2+\sigma_e^2}\right)$$

Variational Model gamma = 0

Variational Model gamma = 0.1



Conclusions

Feedback of priors is required for sensory areas to represent full posteriors, but potentially at the cost of biased evidence accumulation due to "double counting" the prior.

This **perceptual confirmation bias** effect explains discrepancy in past studies as well as within-subject differences in the present study.

References

[1] Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature, 459(7243), 89–92.
[2] Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. The Journal of Neuroscience, 8(12), 3017–3029.
[3] Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science, 340(6128), 95–8.
[4] Wyart, V., Gardelle, V. De, Scholl, J., & Summerfield, C. (2012). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. Neuron, 76(4), 847–858.
[5] Haefner, R. M., Berkes, P., & Fiser, J. (2016). Perceptual Decision-Making as Probabilistic Inference by Neural Sampling. Neuron, 90, 649–660.
[6] Fiser, J. J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences, 14(3), 119–30.
[7] Raju, R. V., & Pitkow, X. (2016). Inference by Reparameterization in Neural Population Codes. Advances in Neural Infromation Processing Systems, 30.
[8] Murphy, K. (2012) Probabilistic Machine Learning.