Introducing the CORTEX database: COntext-dependent Reinforcement learning and Transfer EXperiments

Isabelle Hoxha (isabellehoxha@gmail.com)
Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Stefano Palminteri (stefano.palminteri@gmail.com)
INSERM, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Abstract

Learning - Transfer paradigms are particularly relevant to study value learning, as they can uncover contextual value learning. However, there is no consensus within and across laboratories on the implementation of the experimental variable nor the exact computational processing underlying context-dependence, thus making it unclear whether models such as reference-point and range adaptation genuinely generalize across contexts or are merely tailored to specific task structures. We therefore created an extremely large (n>2500 human participants) yet extremely curated dataset to establish a standardized framework and systematically evaluate the applicability of competing value learning models. We showed that the range adaptation model is the best fitting model for half of the participants, with task specifications such as feedback type modulating the relevance of in-context learning models.

Keywords: open science; reinforcement learning; range adaptation; reference-point centering

Introduction

Increasing evidence suggest that learning values in multi-armed bandit tasks is done in context, i.e. the values are not learned with their absolute magnitude but rather in relation to the alternatives (Palminteri et al., 2015; Bavard et al., 2018; Lebreton et al., 2019; Hayes & Wedell, 2023b). However, task specifications such as feedback type, number of alternatives, or magnitudes displayed, seem to alter the predictive power of contextual learning models. Here, we addressed this issue by compiling to a unique format 31 experiments from 10 published studies, gathering more than 2500 human participants and 700,000 trials, with transfer testing after the initial learning phase. These datasets were subsequently fit using three reinforcement-learning model variants which take into account the context in different ways.

Methods

Data format. The dataset only gathers bandit tasks in published experiments. The datasets were selected on the basis of the presence of a learning phase, where fixed ensembles of contingencies are presented together, and a transfer phase, where all pairs of stimuli are presented together. While all experiments have their specificities, a common subset of variables was kept: the experiment value (i.e. the expected value across the full experiment), the agent number, the trial number, the trial order within the presented ensemble, the session number, a binary marker of whether the trial was a transfer or not, the stimuli presented, the expected values of the stimuli, the availability of the stimuli (i.e. whether there was a forced choice or observational trial), the choice, the outcomes (obtained and foregone when available), the response times, and the regret associated with the trial. Any missing information was replaced by a NaN entry.

Models. At the present time, we compared the fitness of three reinforcements learning models on the collected data.

Absolute model. This model corresponds to a classical Q-learning rule. At each trial, when participants receive feedback r, Q-values are adapted according to:

$$Q_i \leftarrow Q_i + \alpha_i(r_i - Q_i)$$

We set $\alpha_i = \alpha_c$ when option i was chosen, and $\alpha_i = \alpha_u$ when option i was not chosen.

Reference-point centering. Before the updating Q-values, reference-point centering assumes that the rewards are compared to the mean of received rewards V within a context:

$$\tilde{r} = r - V$$

The referenced rewards $\tilde{r}$ are subsequently used to update Q-values. The mean value is then updated based on the mean of observed rewards $\bar{r}$:

$$V \leftarrow \alpha_{ref}(\bar{r} - V)$$

Range adaptation. The rewards are compared to the range of the rewards received within a context, with the range adapting at each trial. In this work, the following variant was implemented:

$$\tilde{r} = \left(\frac{r - R_{\min}}{R_{max} - R_{\min} + 1}\right)^{w_r}$$

The range boundaries are updated at rate α_R if the rewards are outside of the range, following:

$$\begin{cases} R_{min} \leftarrow R_{min} + \alpha_R(\min(r) - R_{min}) \\ R_{max} \leftarrow R_{max} + \alpha_R(\max(r) - R_{max}) \end{cases}$$

Results

Datasets. The datasets included are summarized on Table 1. In total, data from 10 published studies, spanning 31 experiments and 2534 agents have been included, for a total of 700,950 trials.

Table 1: Included datasets

Paper	N.exp	N.agents
Bavard & Palminteri, 2023	4 (1-4)	500
Hayes & Wedell, 2023	1 (5)	64
Vandendriessche et al., 2023	1 (6)	26
Bavard et al., 2018	2 (7, 8)	60
Gueguen et al., 2024	1 (9)	44
Hayes & Wedell, 2023b	2 (10,11)	222
Hayes & Wedell, 2023c	1 (12)	50
Salem-Garcia et al., 2023	10 (13-22)	207
Bavard et al., 2021	8 (23-30)	800
Anlló et al., 2024	11 (31)	561

Fitting results. We present here the best winning model on the basis of the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). We represent the share of each winning model per participant across all experiments and within experiment.

On the basis of the AIC, the range adaptation model was the best for 1254 participants, the reference-point centering model for 186 participants and the absolute model for 1094 participants. On the basis of the BIC, the absolute model was this time the best, being the best fitting model for 1481 participants, followed by the range adaptation model (916 participants) and the reference-point centering model (137 participants).

Importantly, the best fitting model depended on participants and experiment. We observed that the absence of feedback in the transfer phase enhanced context-dependent learning (over global learning).

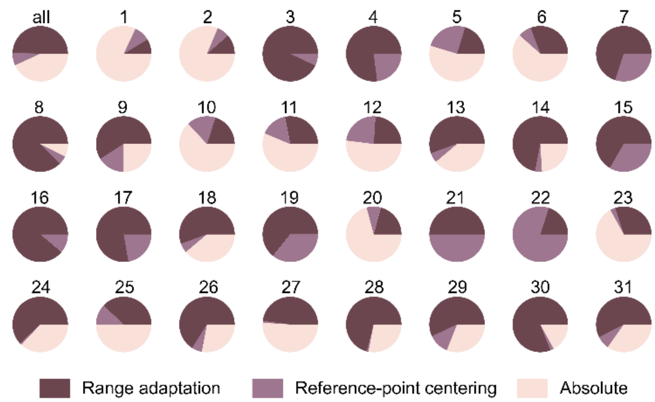

Figure 1: Proportion of best fitting model on the basis of AIC for all experiments and split per experiment.

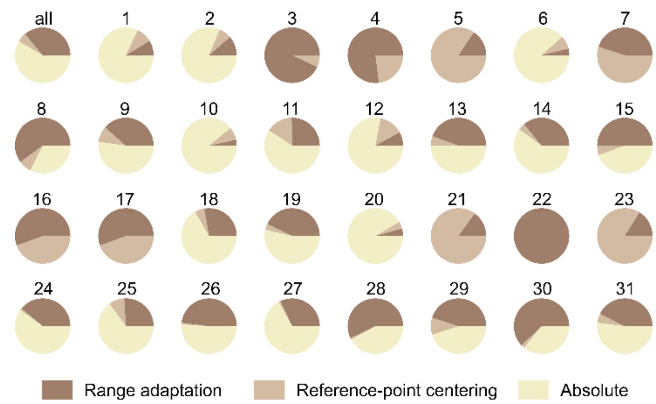

Figure 2: Best fitting model on the basis of BIC for all experiments and split per experiment.

Conclusion

Over the 31 selected experiments, we showed that the range adaptation model was the best selected model. We noted that the selection of the best model depended on the task, and therefore the absolute encoding model was sometimes better suited, in particular when a unique range was presented to participants. The next steps of this study will include an analysis of parameter recovery as a function of task specifications, as well as other variants of in-context value learning.

References

Anlló, H., Bavard, S., Benmarrakchi, F., Bonagura, D., Cerrotti, F., Cicue, M., Gueguen, M., Guzmán, E. J., Kadieva, D., Kobayashi, M., Lukumon, G., Sartorio, M., Yang, J., Zinchenko, O., Bahrami, B., Silva Concha, J., Hertz, U., Konova, A. B., Li, J., … Palminteri, S. (2024). Comparing experience- and description-based economic preferences across 11 countries. *Nature*

Human Behaviour, *8*(8), 1554–1567. https://doi.org/10.1038/s41562-024-01894-9

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature Communications*, *9*(1), 4503. https://doi.org/10.1038/s41467-018-06781-2

Bavard, S., & Palminteri, S. (2023). The functional form of value normalization in human reinforcement learning. *eLife*, *12*, e83891. https://doi.org/10.7554/eLife.83891

Bavard, S., Rustichini, A., & Palminteri, S. (2021). Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances*, *7*(14), eabe0340. https://doi.org/10.1126/sciadv.abe0340

Gueguen, M. C. M., Anlló, H., Bonagura, D., Kong, J., Hafezi, S., Palminteri, S., & Konova, A. B. (2024). Recent Opioid Use Impedes Range Adaptation in Reinforcement Learning in Human Addiction. *Biological Psychiatry*, *95*(10), 974–984. https://doi.org/10.1016/j.biopsych.2023.12.005

Hayes, W. M., & Wedell, D. H. (2023a). Effects of blocked versus interleaved training on relative value learning. *Psychonomic Bulletin & Review*, *30*(5), 1895–1907. https://doi.org/10.3758/s13423-023-02290-6

Hayes, W. M., & Wedell, D. H. (2023b). Reinforcement learning in and out of context: The effects of attentional focus. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(8), 1193–1217. https://doi.org/10.1037/xlm0001145

Hayes, W. M., & Wedell, D. H. (2023c). Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition*, *230*, 105280. https://doi.org/10.1016/j.cognition.2022.105280

Lebreton, M., Bacily, K., Palminteri, S., & Engelmann, J. B. (2019). Contextual influence on confidence judgments in human reinforcement learning. *PLOS Computational Biology*, *15*(4), e1006973. https://doi.org/10.1371/journal.pcbi.1006973

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Salem-Garcia, N., Palminteri, S., & Lebreton, M. (2023). Linking confidence biases to reinforcement-learning processes. *Psychological Review*, *130*(4), 1017–1043. https://doi.org/10.1037/rev0000424

Vandendriessche, H., Demmou, A., Bavard, S., Yadak, J., Lemogne, C., Mauras, T., & Palminteri, S. (2023). Contextual influence of reinforcement learning performance of depression: Evidence for a negativity bias? *Psychological Medicine*, *53*(10), 4696–4706. https://doi.org/10.1017/S0033291722001593