

POSTER PRESENTATION

Open Access

Coordination of adaptive working memory and reinforcement learning systems explaining choice and reaction time in a human experiment

Guillaume D Viejo^{1,2*}, Mehdi Khamassi^{1,2}, Andrea Brovelli³, Benoît Girard^{1,2}

From The Twenty Third Annual Computational Neuroscience Meeting: CNS*2014
Québec City, Canada. 26-31 July 2014

Contemporary behavioral learning theory provides a comprehensive description of how we and other animals learn, and places behavioral flexibility and automaticity at heart of adaptive behaviors. However, to our knowledge, the computations supporting the interactions between *deliberative* and *habitual* decision-making systems are still poorly understood. Previous functional magnetic resonance imaging (fMRI) results suggest that the dorsal striatum host complementary computations that may differentially support *deliberative* and *habitual* processes [1] in the form of a dynamical interplay rather than a serial recruitment of strategies. From the same instrumental task, we develop a dual-system computational model of the two systems that can predict both performance (i.e., participant choices) and modulations in reaction times during learning. The instrumental task is a trial-and-error learning task requiring participants to find the correct associations between color stimuli and finger responses.

To model the habitual system, we use a simple Q-learning algorithm (QL) [2] whose properties are fast responses, but slow convergence. For the deliberative (i.e goal-directed) system, we propose a new Bayesian Working Memory (BWM) which searches for information in the history of previous trials and stops as soon as the uncertainty on the action to perform decreases below a certain threshold. Last, we also propose a model for QL and BWM coordination. Currently, most models of system selection tend to control action selection concurrently, using either the deliberative or habitual model according to uncertainty criteria [3,4]. Only one model

has investigated the relation between working memory and reinforcement learning [5] without, however explicitly modeling the temporal aspect of memory manipulation. In our approach, we propose a model for QL and BWM coordination. QL and BWM are merged such that the expensive memory manipulation is under control of, among others, the level of convergence of the habitual learning. Consequently, we also predict specific reaction times for each model that can be compared with the evolution of reaction times in instrumental learning tasks.

Models are optimized for each subject with the NSGA-2 multi-objective evolutionary algorithm. The first fitness function is the Bayesian Information Criterion for individual choices. The second fitness function is also a likelihood that maximizes the probability of performing reaction times similar to humans. We compare the ability of the new model to explain human behavior with the QL or BWM only, as well as with a combination of these models based on [4], which reveals that the proposed model is in general more accurate. To conclude, we suggest that a close combination of BWM and QL better explains both choices and reaction times for most participants.

Authors' details

¹Sorbonne Universités, UPMC, Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France. ²CNRS, UMR 7222, ISIR, F-75005, Paris, France. ³Institut de Neurosciences de la Timone (INT), UMR 7289, CNRS - Aix Marseille Université, Marseille, France.

Published: 21 July 2014

References

1. Brovelli A, Nazarian B, Meunier M, Boussaoud D: **Differential roles of caudate nucleus and putamen during instrumental learning.** *NeuroImage* 2011, **57**(4):1580-1590.

* Correspondence: guillaume.viejo@isir.upmc.fr

¹Sorbonne Universités, UPMC, Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France

Full list of author information is available at the end of the article

2. Watkins C, Dayan P: **Q-Learning**. *Machine Learning* 1992, **292**(8):279-292.
3. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nature Neuroscience* 2005, **8**(12):1704-1711.
4. Keramati M, Dezfouli A, Piray P: **Speed/accuracy trade-off between the habitual and the goal-directed processes**. *PLoS computational biology* 2011, **7**(5):e1002055.
5. Collins A, Frank MJ: **How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational and neurogenetic analysis**. *European Journal of Neuroscience* 2012, **35**(7):1024-1035.

doi:10.1186/1471-2202-15-S1-P156

Cite this article as: Viejo et al.: Coordination of adaptive working memory and reinforcement learning systems explaining choice and reaction time in a human experiment. *BMC Neuroscience* 2014 **15**(Suppl 1):P156.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

