

Goal-directed top-down control of perceptual representations: A computational model of the Wisconsin Card Sorting Test

Giovanni Granato (giovanni.granato@istc.cnr.it)

Laboratory of Computational Embodied Neuroscience,
Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, Italy

Gianluca Baldassarre (gianluca.baldassarre@istc.cnr.it)

Laboratory of Computational Embodied Neuroscience,
Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, Italy

Abstract

In mammals, goal-directed behaviour, relying on an integrated network of fronto-striatal and fronto-parietal systems, supports the performance of flexible behavior. Here we focus on the contribution to such flexibility of top-down selection processes involving internal representations of percepts. We study these processes through a computational model able to solve the Wisconsin Card Sorting Test (WCST), an important neuropsychological test used for measuring cognitive flexibility. The analysis of the model behavior solving the WCST, and its errors resulting from different lesions, are comparable with those of healthy participants and patients with frontal impairments performing the test. The results represent a first validation of our hypothesis on the importance of internal representation selection for cognitive flexibility.

Keywords: Computational model, supervised RBM, goal-directed behaviour, top-down selection, cognitive flexibility

Introduction

Goal-directed behaviour (GDB; Daw, Gershman, Seymour, Dayan, and Dolan, 2011) is supported by mechanisms able to flexibly create associations between perceptions and actions. This function relies on an active internal exploration of the possible alternative courses of action based on task-independent representations of the world dynamics. Here we hypothesise that GDB supports cognitive flexibility (the ability to change strategy depending on the external feedback; Diamond, 2013) based on the selection of internal perceptual representations. This selection relies on the active top-down control of perception affecting both bottom-up and top-down perceptual processes (Findlay & Gilchrist, 2001; Vitay & Hamker, 2007).

The Wisconsin Card Sorting Test (WCST; Heaton et al., 2000) is an important neuropsychological test used to measure cognitive flexibility. Here we propose a computational model able to solve the WCST by pivoting on a novel mechanism operating a top-down selection of internal perceptual representations. We validate the model by showing how its behaviour, and the behaviour of some lesioned versions of it, resemble the behaviour of healthy and pathological human participants of the test.

Previous computational models focused on the theoretical analysis (Dehaene & Changeux, 1991), psychiatric patients (Berdia & Metz, 1998), deficit varieties (Kaplan, Şengör,

Gürvit, Genç, & Güzeliş, 2006), neuronal dynamics (Rigotti, Ben Dayan Rubin, Wang, & Fusi, 2010), sequential learning processes (Bishara et al., 2010), and the role of a specific brain component such as basal ganglia (Caso & Cooper, 2017) in the WCST. However, none of them showed the important role that the selection of internal perceptual representations, and the effect of them on bottom-up and top-down processes, might have for the solution of the task.

Task

In the WCST (Figure 1) the participant has to draw a card from a deck and match it to one of four sample cards following a visual criterion. Each card contains items with unique combinations of features. These features are grouped in three categories, each involving four attributes: colour (red, green, blue, or yellow); shape (stars, triangles, circles, or crosses); number (one, two, three, or four elements). To solve the task, the participant should move each deck card close to one of the target cards by trying to match it in terms of either the colour, the form, or the number. After each matching attempt, an external operator gives a feedback ('correct' or 'not correct') based on the current matching rule. The participant is not told this rule that so should be inferred through the feedback. Critically, the correct rule changes after a certain number of uninterrupted correct actions (usually 10) and when this happens the participant should infer the new rule and flexibly switch to it.

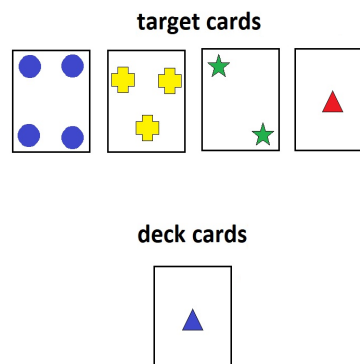


Figure 1: The Wisconsin Card Sorting Test: task setting.

The model architecture and functions

Figure 2 shows the architecture and functions of the model.

Neuro-functional components of the model

The model is formed by components implementing functions that resemble analogous functions of brain areas relevant for a GDB-based interaction with the environment: (a) a visual sensor that mimics the fovea/perifovea parts of the retina; (b) a hierarchical perceptual system that processes visual information and extracts visual features, simulating the human visual system (Konen & Kastner, 2008); (c) a working-memory that stores the goals linked to the possible matching rules, supported by the brain prefrontal cortex (Barraclough, Conroy, & Lee, 2004) and frontal-striatal loops (Baldassarre et al., 2013); (d) a motivational system, processing the external feedback, that activates the goals within the working-memory, a function mostly supported by brain basal ganglia and ventral systems (Gläscher, Daw, Dayan, & O'Doherty, 2010); (e) a selector that chooses the matching rule and hence the specific features of the cards that the system focuses on, simulating the contribution of basal ganglia (Redgrave, Prescott, & Gurney, 1999) and frontal-parietal cortex (Gazzaley & Nobre, 2012) to top-down control of internal perceptual representations; (f) a comparator that supports the visual matching (comparison of the deck card and a target card), executed by brain frontal and temporal-occipital cortices (Perani et al., 1999); (g) a motor component performing the movements of the visual sensor and the displacement of the deck card close to the target card.

Computational implementation of the model

The model components are mainly implemented with neural networks. The working-memory component is formed by recurrent units that encode the tendency to choose specific matching rules. Each unit has a self-synapse and decays to a baseline (0.5) with a decay rate ϕ . The motivational component, that changes the working-memory values on the basis of external feedback, is supported by a reinforcement learning algorithm (reward - expectation) with a learning rate μ . The hierarchical perceptual system is supported by a modified version of Deep Belief Network (DBN; Hinton, Osindero, and Teh, 2006), a bidirectional network composed by two stacked generative Restricted Boltzmann Machines (RBM; Hinton, 2012). The DBN is composed by a traditional RBM - trained by the contrasting divergence algorithm - and a supervised RBM trained with a modified version of the contrasting divergence algorithm where specific units of the external layer are clumped to different attributes of the colour, shape, and number (this learning putatively takes place before the WCST session based on the interaction with the world). Top-down selector is composed by (a) a softmax function (with temperature τ) that operates the choice of the behavioural strategy based on the working memory content, and (b) a disinhibition mechanism, mimicking basal ganglia, which inhibits all the units of the last DBN layer with the exception of the one encoding the desired rule. This allows the system to have an internal representation of each card (RBM) that is focused one one

specific attribute (either colour, shape, number). This process is sequentially applied to the deck-card and each target-card until a matching between them is found, and the deck-card displacement is triggered. The matching (based on Euclidean distance) is performed on the basis of the DBN reconstructed image focused on the selected attribute.

Results and conclusion

We used five behavioural measures to score the model and compare it with human data: (a) completed categories (CC): this identifies the number of successfully completed categories (max six categories i.e. colour, shape, number, colour, shape, number) and informs about the level of global performance; (b) total errors (TE): this identifies the total incorrect responses and is informative about the level of global deficit; (c) perseverative errors (PE): these mark a perseverative tendency to sort the cards with the same incorrect rule after a negative feedback; (d) non-perseverative errors (NPE): these occur in different situations and suggest an attentional failure or incorrect inferential reasoning; (e) Failure-to-Maintain Sets (FMS) errors: these occur after five consecutive correct matches and suggest a distraction.

We searched different parameter settings to obtain two different groups of models ('conditions) exhibiting different relevant behaviours: (a) healthy model (HM): this baseline model reproduces the behaviour of normotypical WCST participants; (b) pathological model (PM): this reproduces the behaviour of pathological WCST participants. Moreover, we applied two types of 'lesions' to the model, thus obtaining two 'extreme pathological' versions of it. The first lesion (an extreme low μ value that negatively influences the efficacy of reinforcement processing) produces a highly perseverative model (PM) with a low sensitivity to rule changes (high number of PE). The second lesion (an extreme high values of τ , the temperature of softmax, that makes the model insensitive to values differences, together with an extreme high ϕ , the decay speed of recurrent units self-synapses, that causes a forgetting of previous chosen rule) produces a distracted model (DM) that shows an opposite tendency with respect to the previous model, i.e. an extreme distracted and erratic behaviour (high number of NPE and FMS errors).

Figure 3 and Figure 4 show the statistical comparisons between human data and artificial data in healthy and pathological conditions, highlighting the similarity between the model behaviour and target empirical data. Figure 5 shows the errors of the extreme pathological models, highlighting the effects of specific neuro-inspired lesions.

The results validate the hypothesis, incorporated in the neuro-inspired model, that the selection of internal perceptual representations might represent a key mechanism supporting goal-directed cognitive flexibility measured by the WCST. In particular, the manipulation of the parameters of the model regulating its selection and memory processes, mimicking different possible lesions, reproduce the tendency of human patients to exhibit perseverative or distracted WCST errors.

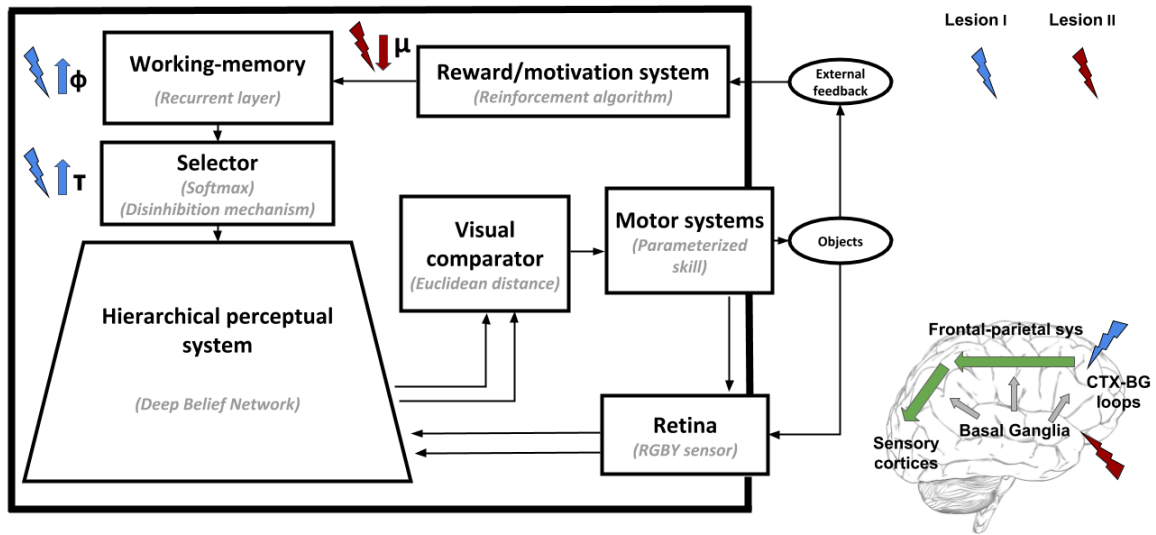


Figure 2: Schema that integrates biological, functional and computational information about constituting components. The brain figure into upper-right corner exemplifies brain systems simulated in our model, supporting the top-down manipulation of perception .

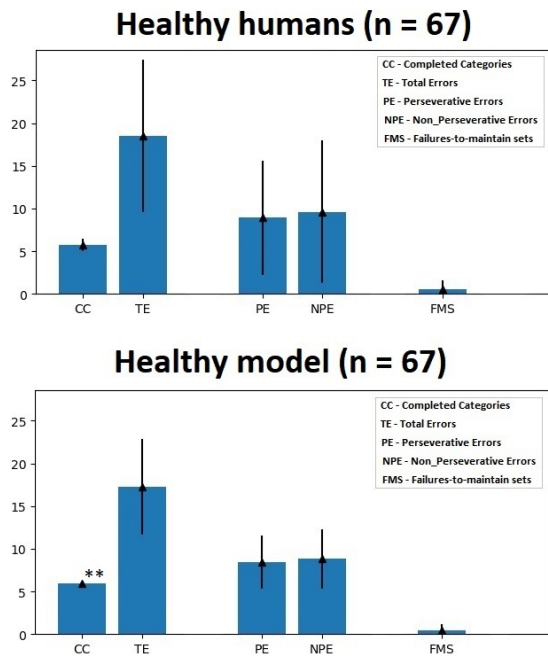


Figure 3: Statistical comparison of the errors of healthy human participants and the healthy model.

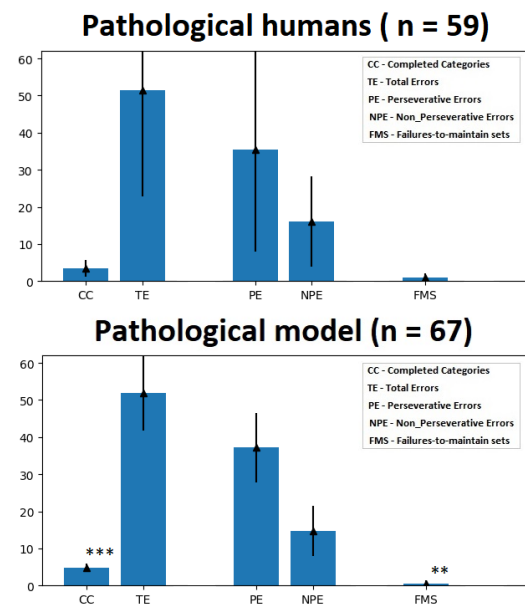


Figure 4: Statistical comparison of the errors of the human pathological participants and the pathological model.

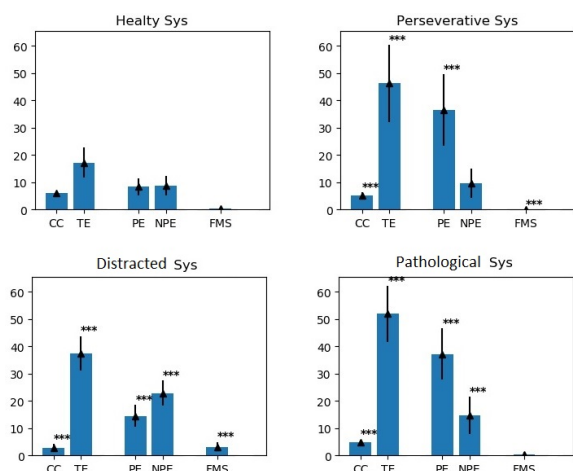


Figure 5: Statistical comparison of the errors of the healthy, pathological, extreme perseverative, and extreme distracted pathological model.

Acknowledgments

This research has received funding from the European Unions Horizon 2020 Research and Innovation Programme, under the project 'GOAL-Robots – Goal-based Open-ended Autonomous Learning Robots, Grant Agreement No 713010.

References

- Baldassarre, G., Mannella, F., Fiore, V. G., Redgrave, P., Gurney, K., & Mirolli, M. (2013). Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model. *Neural Networks*, *41*, 168-187. doi: 10.1016/j.neunet.2012.09.015
- Barracough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nature neuroscience*, *7*(4), 404.
- Berdia, S., & Metz, J. (1998). An artificial neural network stimulating performance of normal subjects and schizophrenics on the wisconsin card sorting test. *Artificial intelligence in medicine*, *13*(1-2), 123–138.
- Bishara, A. J., Kruschke, J. K., Stout, J. C., Bechara, A., McCabe, D. P., & Busemeyer, J. R. (2010). Sequential learning models for the wisconsin card sort task: Assessing processes in substance dependent individuals. *Journal of mathematical psychology*, *54*(1), 5–13.
- Caso, A., & Cooper, R. P. (2017). A model of cognitive control in the wisconsin card sorting test: Integrating schema theory and basal ganglia function. In *Proceedings of the 39th annual meeting of the cognitive science society, cogsci 2017, london, uk, 16-29 july 2017* (pp. 210–215).
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. doi: 10.1016/j.neuron.2011.02.027

- Dehaene, S., & Changeux, J.-P. (1991). The wisconsin card sorting test: Theoretical analysis and modeling in a neuronal network. *Cerebral cortex*, *1*(1), 62–79.
- Diamond, A. (2013). Executive functions. *Annu Rev Psychol*, *64*, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Findlay, J. M., & Gilchrist, I. D. (2001). Visual attention: The active vision perspective. In *Vision and attention* (pp. 83–103). Springer.
- Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in cognitive sciences*, *16*(2), 129–135.
- Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. doi: 10.1016/j.neuron.2010.04.016
- Heaton, R., Chelune, G., Talley, J., Kay, G., Curtiss, G., di Hardoy, M., ... e Cabras, P. (2000). *Wcst: Wisconsin card sorting test : forma completa revisionata : manuale*. Firenze: O.S.
- Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade* (pp. 599–619). Springer.
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527–1554.
- Kaplan, G. B., Şengör, N. S., Gürvit, H., Genç, İ., & Güzeliş, C. (2006). A composite neural network model for perseveration and distractibility in the wisconsin card sorting test. *Neural Networks*, *19*(4), 375–387.
- Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci*, *11*(2), 224–231. doi: 10.1038/nn2036
- Perani, D., Schnur, T., Tettamanti, M., Cappa, S. F., Fazio, F., et al. (1999). Word and picture matching: a pet study of semantic category effects. *Neuropsychologia*, *37*(3), 293–306.
- Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023.
- Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J., & Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Frontiers in computational neuroscience*, *4*, 24.
- Vitay, J., & Hamker, F. H. (2007). On the role of dopamine in cognitive vision. In *International workshop on attention in cognitive systems* (pp. 352–366).