NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Biologically constrained large-scale model of the Wisconsin Card Sorting Test

Kajić, Ivana; Stewart, Terrence C

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

Proceedings of the Annual Meeting of the Cognitive Science Society, 43, pp. 2295-2301, 2021-07-29

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=53eb3b7c-9ad6-49d9-882b-14de9b4cca29
https://publications-cnrc.canada.ca/fra/voir/objet/?id=53eb3b7c-9ad6-49d9-882b-14de9b4cca29

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Biologically Constrained Large-Scale Model of the Wisconsin Card Sorting Test

Permalink
https://escholarship.org/uc/item/4xg8936b

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Kajić, Ivana
Stewart, Terrence C

Publication Date
2021

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Biologically Constrained Large-Scale Model of the Wisconsin Card Sorting Test

Ivana Kajić (kivana@google.com)
DeepMind, Montreal
Canada

Terrence C. Stewart (terrence.stewart@nrc-cnrc.gc.ca)
National Research Council of Canada
University of Waterloo Collaboration Centre
Canada

Abstract

We propose a biologically constrained, large-scale neural network model that solves the Wisconsin Card Sorting Test (WCST). The WCST has been widely used in clinical and research settings to study cognitive flexibility and executive function. The model shows a good quantitative match with human responses across a number of WCST scoring indices, while consisting of neural networks that functionally and anatomically map to brain areas and structures implicated in the task, such as the prefrontal cortex and the cortico-basal ganglia-thalamus-cortical loop. We argue that the model provides a mechanistic account of WCST solving, and demonstrate its robustness by examining its performance across a range of biologically motivated parameter values.

Keywords: Wisconsin Card Sorting Test; Spiking Neural Network; Neural Engineering Framework; Semantic Pointers

Introduction

The Wisconsin Card Sorting Test (WCST) is used in clinical and research environments to study a variety of higher-level cognitive functions such as working memory, cognitive flexibility and abstract reasoning. The test consists of a deck of stimulus cards that need to be matched to 4 target cards. The target cards remain unchanged for the duration of the experiment, while the stimulus card changes in each trial. The cards differ in colour, symbol shape, and the number of symbols.

In a trial, an individual matches a stimulus card to one of the target cards and receives feedback indicating whether the match was correct or incorrect. Unbeknownst to them, after 10 correct trials, the experimenter changes the sorting rule. While a person is not told what the matching rule is, people are quick to notice that the cards can be sorted according to one of three rules: number, shape or colour. Importantly, people with cognitive deficits show a pattern of perseveration in their responses—even after being told that their match is incorrect, they continue to apply the same rule. As such, their performance is characterized by more errors and a lower number of categories achieved.[1]

Here, we take a neurocomputational cognitive modelling approach, and present a large-scale neural network model that proposes specific mechanisms across different brain regions that underlie solving WCST. The model bridges different levels of analysis—at the lower level, individual computations are realized by groups of neurons, while at the higher level it exhibits human-like behaviour on the task.

[1] A category is said to be achieved after 10 correct trials.

Existing Models

The WCST has been used as a benchmark test for computational models investigating cognitive control. While many models have been proposed that focus on different aspects of the task, here, we focus on a smaller subset of neurocomputational models. Such models aim to explain how computations performed by networks of neurons give rise to higher-level cognitive functions.

One of the first neural network models solving the WCST is that of Dehaene and Changeux (1991). Components of their model are able to adjust the activity to reflect the presence of negative reward, memorize input values and rules, as well as reason about the rules. The authors conduct ablation studies to analyze the effect of lesioned networks, showing that the model is able to reproduce aspects of the behaviour of brain-lesioned patients. While reporting perseveration rates, this work did not include a detailed comparison with human data on the task.

Another, conceptually similar, neural network model was proposed by Amos (2000). One feature of this model is that it contains units that simulate some of the basal ganglia and thalamus function, known to be involved in the feedback processing in the WCST (Monchi et al. 2001). This model also aims to coarsely match connectivity patterns of associated brain structures. However, one drawback in terms of the model's ability to incorporate neuroanatomical details is the lack of the feedback connection that completes the cortico-basal ganglia-thalamo-cortical loop (Parent and Hazrati 1995). As such this model has a limited capacity to account for the role of subcortical structures in the WCST. Another concern with this model is the lack of evaluation with respect to its variability in responses.

Finally, a model that performs a simplified version of the WCST, and that has also been used in other similar cognitive control tasks, has been proposed by Rougier et al. (2005). Similarly to previous models, this model contains distinct networks with units representing dimensions and features of interest. However, one key feature of this model is its ability to learn to perform the task in a trial-and-error way. This is done by training the model on thousands of task trials, resulting in the adjustment of weights between individual networks. Similarly to the model of Amos (2000), this model also contains a component that resembles an aspect of basal ganglia function,

namely an adaptive gating mechanism.

One feature shared among all mentioned models is their reliance on a pre-allocated role of neural units in the networks. Such units encode unique and specific features, such as dimensions or dimension values that need to be specified for the task. As such, those are all *localist* networks, contrasting the idea of distributed neural representation ubiquitous in the brain (McClelland et al. 1986). Dehaene and Changeux (1991) acknowledge that that this type of "grandmother cell" was adopted for simplicity in their model. However, it is unclear how existing models can be adapted to accommodate distributed neural representations. Such representations have important implications for scaling the models to accommodate a wider variety of tasks and inputs, as well as their robustness to deal with noise that naturally occurs in biological neural systems.

Our proposed model uses a distributed, noisy representation in a neural network whose components and connections between them functionally resemble those of the associated brain areas. Crucially, the core components of the model are agnostic to particular input features in the WCST. That is, there are no units in the model that are uniquely and exclusively representing input concepts, such as *red* or *triangle*. Instead the model extracts those concepts from the inputs, stores them in working memory, and uses the memory to adjust its response based on feedback from the environment. As well, our model performs the task in a way comparable to participants in the experiment—it performs the task continuously over time while considering the four target cards, its previous responses, and by matching a stimulus card in each trial. With our model, the whole experiment is conducted as one long simulation, which has not been done with the previous models. We also evaluate our model across a number of WCST normative scoring indices, and show that it provides a good match with human data while being robust to changes in different model parameters.

Biologically Constrained Model

Methods We use the Neural Engineering Framework (NEF; Eliasmith and Anderson 2003) to build the neural network performing the WCST. To do this, we break the WCST down into an algorithm consisting of vectors, functions applied to those vectors, and differential equations on those vectors. Vectors are represented with the pattern of activity across a group of neurons, and the functions and differential equations are implemented by finding connection weights between groups of neurons that best approximate those functions. Here we use the Leaky Integrate-and-Fire (LIF) neuron model due to its favorable trade-off between low computational complexity and rudimentary biological realism capturing neuronal membrane voltage dynamics (i.e., "spiking" activity).

In our case, desired mathematical functions are derived from a vector symbolic architecture that constitutes the Semantic Pointer Architecture (SPA; Eliasmith 2013). For example, those are the operations of vector (un)binding that are used to (de)compress vector representations. In the SPA, such vectors whose values are represented by neural activities are referred to as *semantic pointers* (SPs).[2] The SPA also includes control components that operate on such representations, such as different types of winner-take-all mechanisms, some of which are used in the proposed WCST model.

The NEF and SPA have been used to model a wide range of cognitive and neurobiological phenomena, such as spatial cognition (Komer et al. 2019; Dumont and Eliasmith 2020), affective processing (Kajić et al. 2019), structured semantic representations (Crawford et al. 2016), low-level cerebellar function (Stöckel et al. 2020), as well as the functional brain model Spaun (Eliasmith et al. 2012). Our proposed model uses the same or similar components that have been used in those other models of neurobiological and cognitive phenomena.

Our model is built using the Nengo Python library for simulating large-scale neural networks (Bekolay et al. 2014). The library provides default values for many of the model parameters (i.e., synaptic time constants, ranges of frequencies for neural responses, to name a few), some of which are set to biologically plausible ranges. Most of the parameters in the model are set to those default values, however, we needed to adjust some time constants to stabilize model behaviour. Those values, as well as the model source code are available at: `https://github.com/ikajic/cogsci21-wcst/`.

Model Architecture

The model consists of interconnected groups of neural ensembles that map to the functions of specific brain regions. While connections between ensembles in our model represent only a subset of all connections between associated brain areas, they are motivated by neuroanatomical connectivity patterns (Monchi et al. 2001). The model is schematically depicted in Figure 1 (A). Neural ensembles are shown as circles, with colours denoting whether an ensemble processes external inputs (green), produces outputs (yellow), or its dynamics is determined internally (white). Arrows denote neural connections and determine the flow of information.

Model I/O Rather than modelling the complete visual system, the visual input to our model is two vectors C_t (for target) and C_s (for stimulus) that would be the high-level output of a visual system. C_t represents the 4 target cards as follows:

$$C_t = C_1 \circledast P_1 + C_2 \circledast P_2 + C_3 \circledast P_3 + C_4 \circledast P_4, \qquad (1)$$

with individual card SPs $C_i, i \in \{1 \dots 4\}$ defined as:

$$C_1 = Color \circledast Red + Shape \circledast Square + Number \circledast One$$
$$C_2 = Color \circledast Blue + Shape \circledast Circle + Number \circledast Two$$
$$C_3 = Color \circledast Green + Shape \circledast Triangle + Number \circledast Three$$
$$C_4 = Color \circledast Yellow + Shape \circledast Star + Number \circledast Four,$$

[2] We will be interchangeably using terms *vector* and *semantic pointer*.

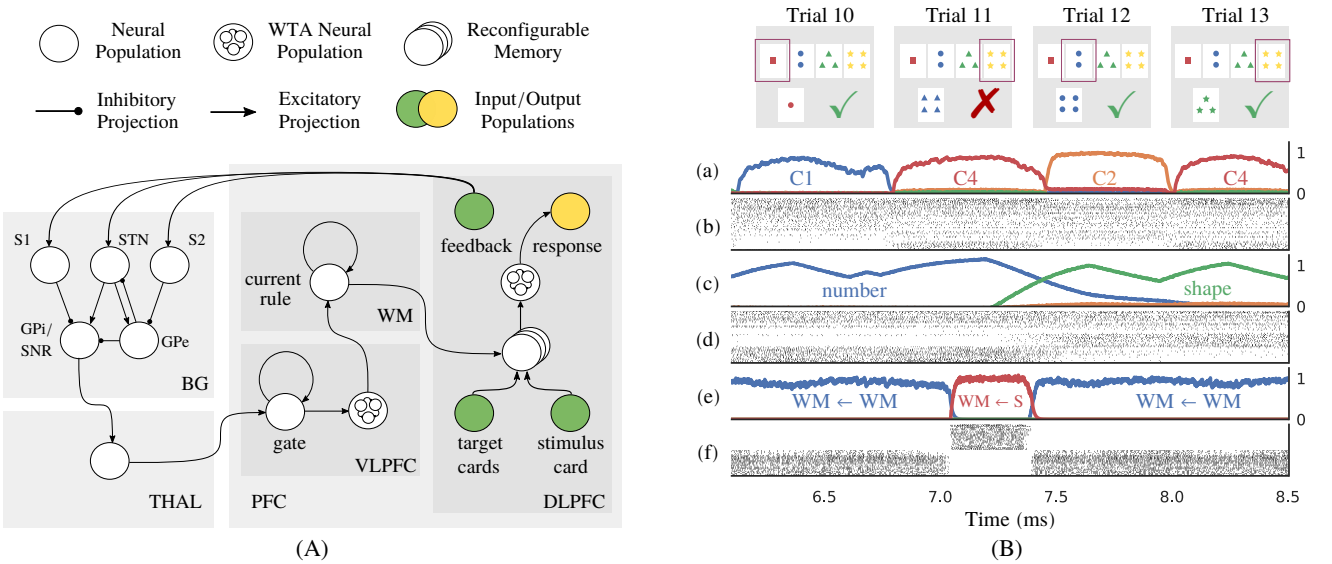

Figure 1: (A): The model. PFC: Prefrontal Cortex, DLPFC: Dorsolateral PFC, VLPFC: Ventrolateral PFC, WM: Working Memory, THAL: Thalamus, BG: Basal Ganglia, S1: Striatum D1, S2: Striatum D2, STN: Subthalamic Nucleus, GPi/GPe: Globus Pallidus internal/external; SNR: Substantia Nigra. (B): An excerpt from a simulation showing the decoded vectors (a, c, e) and spiking activity (b, d, f) in the response population (a, b), working memory (c, d), and the thalamus (e, f) over 4 trials. Decoded values are shown as dot products between the ideal vectors and the vectors represented by a neural population. Experimental rule changed in Trial 11 from *Number* to *Shape*.

where $\circledast$ denotes circular convolution, used for compressive binding of vectors. Each non-composite SP, such as *P1, Color* or *Star*, is a Holographic Reduced Representation (Plate 1995), which is a random vector with elements drawn independently from $\mathcal{N}(0, \frac{1}{n})$, where n is the number of dimensions.

The stimulus vector C_s is composed in a similar way, by controlling for all possible combinations of colour, shape and number attributes (giving 64 stimulus cards). While C_t remains unchanged for the duration of the experiment, C_s changes when the model produces a response, defined as representing one of the 4 target card SPs in the *response* neural group for at least 300 ms. We examine the effect of SP dimensionality on model performance in Experiment 1.

Key Components

Prefrontal Cortex In the context of the WCST, the prefrontal cortex (PFC) is associated with the maintenance of task-related transient information, such as set maintenance and set shifting (Monchi et al. 2001). In the model, the PFC includes networks performing the following function: maintaining the current rule/set (working memory and VLPFC), evaluating individual exemplars based on the current rule (DLPFC), and selecting the response card (DLPFC).

The novel component in charge of exemplar evaluation is the *reconfigurable associative memory*. We refer to it as *reconfigurable* since it can be used to do any task, not just WCST. Its goal is to take a complex stored vector (C_t in the WCST) and extract the 4 different features that are relevant for the current comparison. For C_t given above, and in the case where the current rule stored in working memory (V_{WM}) is *Shape*, the results should be *Square*, *Circle*, *Triangle*, and *Star*.

To do this, each of the four reconfigurable units performs the following operation:

$$feature_i = C_t \circledast P_i^{-1} \circledast V_{WM}^{-1}, \quad (2)$$

where $i = 1 \ldots 4$, V_{WM} is the vector represented in the working memory, $^{-1}$ denotes the approximate vector inverse, and C_t and P_i are as defined in Equation 1.

Then, each feature vector is compared with the stimulus feature vector (obtained in the same way) using a dot product. The 4 resulting values are fed into a WTA mechanism that selects one of the 4 target cards and projects the winning SP to the response population.

Importantly, this neural system computes this operation on *any* input vectors C_t and V_{WM}. This means it could be used for any task where a piece of information needs to be extracted from a set, regardless of whether that information is about colour, shape, size, count, age, location, or anything else that can be represented as a vector.

The cortico-basal ganglia-thalamo-cortical loop The basal ganglia (BG) and the thalamus are generally associated with control in the brain, playing a critical role in action selection and decision making. Our model uses the BG model proposed by Gurney et al. (2001) that is anatomically and physiologically consistent with the biological system.

Through the thalamus and a gating mechanism, the BG control the contents of the working memory based on the current contents as well as the experimental feedback. This feedback is an exogenous signal provided by the environment, and is shown as the green "feedback" population in DLPFC in Figure 1 (A).

If the feedback received in a trial is positive, the thalamus selects a cognitive action corresponding to "maintain the representation in working memory". In Figure 1 (B) this thalamus activity is shown as the WM←WM SP in the panel (e). If the feedback is negative, the current rule in working memory will be replaced by a different one. For example, if the current rule is *Color*, it will be replaced with *Number*, if it is *Number*, it will be replaced with *Shape*, and *Shape* will be replaced with *Color*. As a consequence, this implements a cyclical selection of rules until positive feedback is received. An example of a rule switch from *Number* to *Shape* is shown in Figure 1 (B) in the panels (c) and (e).

Results

We first evaluate the model by examining its overall performance on the WCST. To do so, we focus on two commonly used scoring indices: *the number of categories achieved* and *the number of perseverative errors*. The experiment terminates once the model achieves 6 consecutive categories, or when all 128 stimulus cards are exhausted.

Then, we compare model performance to performance of young healthy adults on a range of scoring indices. While previous modelling studies predominantly reported achieved categories and perseverative errors, here we include a more exhaustive comparison including 6 additional indices frequently reported for human data.

In all instances, we run 100 model simulations to account for possible differences in performance due to parameter initialization. Each simulation corresponds to a different seed used to initialize random parameters in the model. In the NEF, there are several such parameters, such as neural firing rate ranges, membrane thresholds, preferred input directions and encoding weights, thus modelling the variability observed in biological systems. As a first approximation, the diversity in models due to such random parameter initialization can also be considered analogous to individual differences.

Experiment 1: Categories Achieved

We investigated the effect of two parameters on the number of achieved categories: the dimensionality of semantic pointers, and the strength of memory defined by the value of the recurrent connection on the neural population representing the current rule.

Representation Dimensionality The SP dimensionality affects the number of neurons in the model and the average pairwise similarity of randomly generated vectors.[3] We investigate the effect of representation dimensionality in an attempt to quantify the robustness of our model in terms of the total number of neurons used. A model relying on vectors with many dimensions will contain more neurons to represent these additional dimensions. As we show later, our model exhibits graceful degradation in performance as we decrease this number, thus showing that even a smaller model is able to perform the task.

We run 100 simulations for each of the 4 dimensions: $128, 256, 512$ and 1024, and show the distribution of categories achieved in Figure 2 (A). We observe a bimodal distribution with a small mode corresponding to 0 categories and a larger mode corresponding to 6 categories. The model with $d = 128$ has the worst performance, as evidenced by the largest number of simulations ($N = 27$) that failed to complete even one category, but also smallest number of simulations ($N = 33$) that completed 6 categories. Increasing the SP size improves the performance, but appears to saturate beyond $d = 512$: $N = 79$ simulations complete all 6 categories for both $d = 512$ and $d = 1024$.

The approximate model sizes are as follows: 0.53M neurons for $d = 128$, 1.06M for $d = 256$, 2.1M for $d = 512$ and 4.2M for $d = 1024$. Based on these results, we use $d = 512$ in remaining experiments, as the larger model did not offer any substantial improvements and thus did not justify the additional computational expense.

Memory Strength We examine the effect of the recurrent connection on the "recurrent rule" population in WM shown in Figure 1 (A). The strength of this connection (w) determines memory characteristics. Lower values can be seen as modelling a "volatile" working memory that is unable to keep information for a sustained period of time, while higher values reinforce the current representation and saturate memory in a way which can make it difficult to integrate new information. $w = 0$ corresponds to a memory-less system, that is a feed-forward network, while $w = 1$ corresponds to a "perfect" memory[4]. The results of 100 simulations for each model using one of the five different connection strengths are shown in Figure 2 (B).

The highest number of simulations achieving 6 categories ($N = 79$), as well as the smallest number of simulations achieving 0 categories ($N = 16$) is observed with the model with $w = 0.6$. Increasing the weight beyond that value increases the number of simulations achieving 0, 1 or 2 categories, and reduces the number of simulations achieving 6 categories. None of the simulations with $w = 0.1$ achieves more than 2 categories.

Experiment 2: Perseverative Errors

In this set of experiments, we examined the effects of memory on the incidence of perseverative errors. We use the same 100

[3] Vectors randomly sampled from a higher-dimensional space are more likely to be dissimilar, as measured by some metric of pairwise similarity such as the cosine angle or the dot product.

[4] Since we still use neurons, the resulting memory will not actually be perfect, as the neurons are only approximating this ideal function.

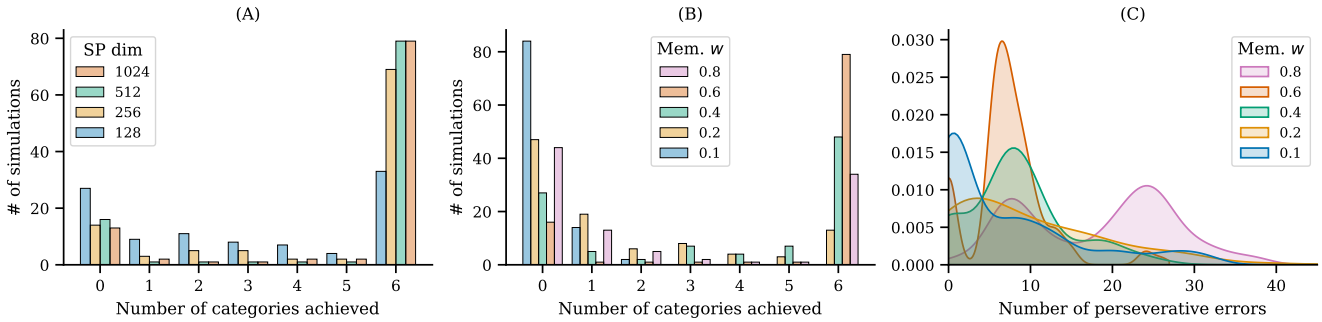

Figure 2: Number of categories achieved for models using semantic pointers of different dimensionality (A), and memory strength (B). Distribution of perseverative errors for different memory strengths (C).

simulations for each of the five different connection strengths as in the previous experiment. A KDE plot showing the distribution of perseverative errors for each connection strength is shown in Figure 2 (C). Each continuous density is estimated using a Gaussian kernel with bandwidth size 0.2, and it has been normalized so that the total area under it sums to 1. The curves have been truncated after the error value of $x = 45$ to highlight the differences in their shapes.

The resulting distributions show different quantitative properties, with a few showing prominent modes at $x = 0$. This corresponds to the majority of simulations where the model failed to achieve a single category. Next, we observe another mode at approximately $x = 7$ for several distributions, as well as one prominent one at $x = 25$ for the distribution with $w = 0.8$. For $w = 0.8$ this mode corresponds to simulations that achieved both 0 categories and more than 0 but less than 6. A small mode also at $x = 25$ for $w = 0.6$ is due to 3 simulations: one simulation where the model produced at most 8 correct responses, thus failing to complete one category; one where it produced one category and many sequences of 5+ correct responses but never 10; as well as one simulation with 6 categories and many errors.

These results lead us to conclude that the strength of recurrent memory connection has a strong impact on the distribution of perseverative errors. We observe a higher number of perseverative errors when $w = 0.8$, consistent with the hypothesis that more saturated memory interferes with the successful integration of new information.

Experiment 3: Comparison with Empirical Data

In addition to achieved categories and perseverative errors, other performance scores typically reported on WCST are: (1) trials achieved, (2) correct trials, (3) total number of errors, (4) perseverative responses, (5) trials to complete the first category, (6) failure to maintain set, (7) conceptual level responses and (8) learning to learn (Heaton et al. 1993).

We use simulation data produced by the three models that yield the highest rates of 6 categories achieved, based on the results shown in Figure 2 (B). Those are models with memory connection strengths $w \in \{0.4, 0.6, 0.8\}$. We also remove all simulations that did not complete a single category, resulting in 73 simulations for the model with $w = 0.4$, 84 for $w = 0.6$ and 56 for $w = 0.8$. These results are compared to the data on WCST performance of 25 college-aged students published in Ashendorf and McCaffrey (2008). The results are shown in Figure 3, with human averages highlighted in yellow.

To compare the results quantitatively, Bonferroni corrected 95% confidence intervals for the difference between two population means are computed, and used to test for differences in reported scores. No significant difference was found for any of the 8 scores when $w = 0.6$, while 2 and 3 significant differences were found with $w = 0.4$ and $w = 0.8$, respectively. For $w = 0.4$ those were the number of categories and failure to maintain set, and for $w = 0.8$ the same two, as well as the number of perseverative responses.

Discussion

The proposed WCST model shows a good match with human data on the test, while establishing an important link to the cognitive and neurobiological levels of explanation. For example, at the cognitive level, the proposed reconfigurable associative memory module is consistent with the two-stage process accounts of discrimination learning (Mackintosh 1965): first, attention is directed towards the relevant cue (e.g., the task rule, such as *shape*), and the cue is used to evaluate individual target cards against the stimulus card. At the neurobiological level, this module operates on non-specific, distributed, noisy representations. As such, it differs from the existing "localist" models that scale poorly and are biologically unrealistic, as they represent a concept with a specific and unique neural population. It should be noted that this module is resource-heavy, as it contains over 95% of all neurons in the model. However, we find that such cost may be justified when considering the added benefit of generality, and applicability to different discrimination tasks with a relatively small set of cues.

We have also demonstrated the robustness of the model, by showing that varying the parameters of the model results in gradual impairment in performance. For example, reducing the semantic pointer dimensionality below $d = 512$ negatively affected the number of simulations that achieve all 6 cate-

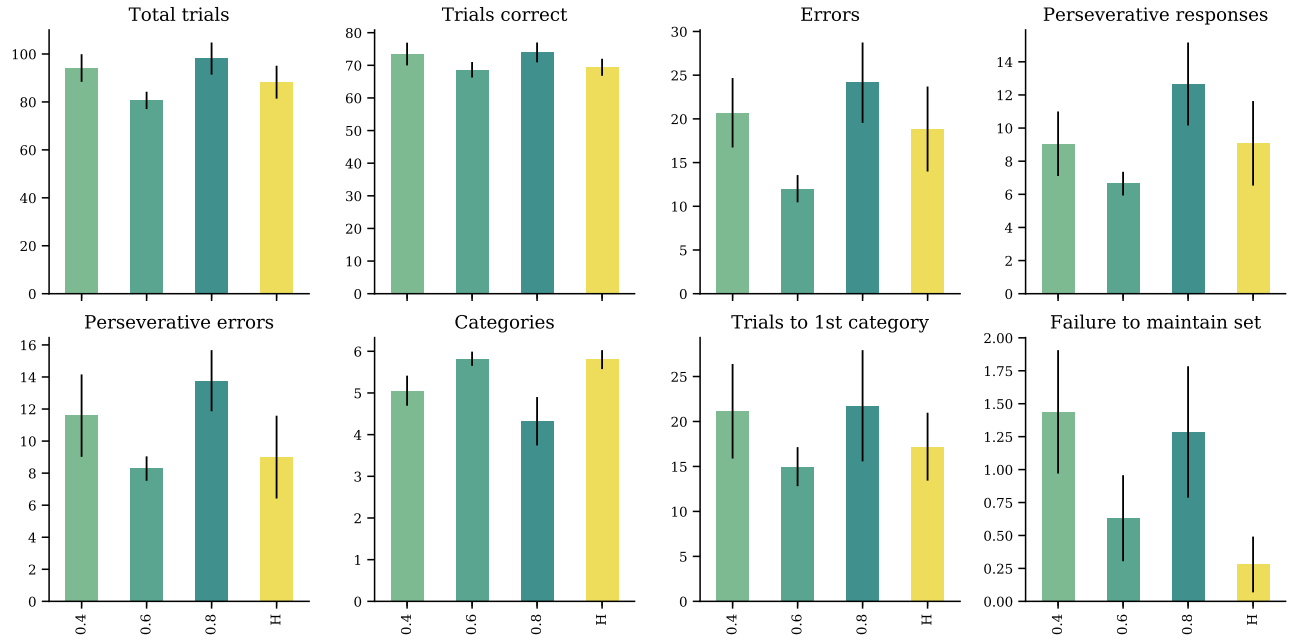

Figure 3: Mean performance and 95% confidence intervals on 8 different WCST normative scores for humans and three models ($w \in \{0.4, 0.6, 0.8\}$). H: Baseline human performance of 25 college-aged adults (Ashendorf and McCaffrey 2008)

.

gories, while no substantial improvements were found with $d = 1024$, which led us to conclude that $d = 512$ offers a good trade-off between the model size and performance. Interestingly, $d = 512$ has been found as the optimal size in related modelling work (Crawford et al. 2016; Eliasmith et al. 2012).

Varying the strength of the recurrent feedback connection on the memory module resulted in a higher incidence of perseveration errors, and a lower number of categories achieved. This decline in performance is consistent with two empirically observed accounts: first, age-related decline in the set shifting and set maintenance ability (Ashendorf and McCaffrey 2008); second, impaired performance observed in patients with dysfunctional prefrontal brain regions, such as in schizophrenia (Weinberger et al. 1986). We speculate that our experimental manipulation is more likely to reflect the poorer working memory abilities of older adults, since it is consistent with empirical evidence of age-related decline in memory updates (Hartman et al. 2001). Furthermore, the role of the basal ganglia-thalamocortical loop in the model is consistent with the proposal of its importance in determining the attentional set and overall action guidance (Monchi et al. 2001). However, additional parameter exploration is needed to comprehensively characterize impaired performance in the context of various cognitive deficits.

We believe that, compared to existing models, our model provides a more accurate match to cognitive and biological mechanisms in the WCST. However, we highlight several venues for future investigations. First, the model makes an assumption about the specific sequential order of testing for three possible task rules (i.e., *Shape, Number* or *Color*). While our results demonstrate that this is a reasonable strategy, as evidenced by our model producing responses that match human data well, it does not account for the learning stage that occurs before one settles onto the three possible rules.

Second, while the model provides a good match with human data, it should be noted that there are certain caveats associated with the statistical comparison of the human data and the model data. While, in many cases, we found no significant difference between the two scores, the inability to reject the null hypothesis (due to overlapping confidence intervals) does not imply equivalence between the model and the system being modelled. In the future, we would like to apply more conservative statistical tests that take into account the relative difference in confidence intervals when evaluating the models.[5]

Finally, given that some components of our model neuroanatomically map to biological structures, the model can be lesioned in a way which corresponds to the damage to the neural system. For example, there is empirical data available on WCST performance in Parkinson's and Huntington's patients and the model could be used to investigate the role of different subcortical nuclei.

Acknowledgments

IK conducted this research while being a graduate student at the University of Waterloo. The authors would like to thank members of the Computational Neuroscience Research

[5]For example, the equivalence test threshold proposed in Stewart and West (2010).

Group at the University of Waterloo for useful feedback on this work, Adam Santoro for a thoughtful review of the manuscript, and anonymous reviewers for numerous suggestions that helped improve this work.

References

Amos, A. (2000). "A computational model of information processing in the frontal cortex and basal ganglia". In: *Journal of cognitive neuroscience* 12.3, pp. 505–519.

Ashendorf, L. and R. J. McCaffrey (2008). "Exploring age-related decline on the Wisconsin Card Sorting Test". In: *The Clinical Neuropsychologist* 22.2, pp. 262–272.

Bekolay, T., J. Bergstra, E. Hunsberger, et al. (2014). "Nengo: a Python tool for building large-scale functional brain models". In: *Frontiers in Neuroinformatics* 7, p. 48.

Crawford, E., M. Gingerich, and C. Eliasmith (2016). "Biologically plausible, human-scale knowledge representation". In: *Cognitive science* 40.4, pp. 782–821.

Dehaene, S. and J.-P. Changeux (1991). "The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network". In: *Cerebral cortex* 1.1, pp. 62–79.

Dumont, N. S.-Y. and C. Eliasmith (2020). "Accurate representation for spatial cognition using grid cells". In: *42nd CogSci Proc.* Toronto, CA: Cognitive Science Society.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY.

Eliasmith, C. and C. H. Anderson (2003). *Neural engineering: computation, representation, and dynamics in neurobiological systems*. English. Cambridge, MA: MIT Press.

Eliasmith, C., T. C. Stewart, X. Choo, et al. (2012). "A large-scale model of the functioning brain". In: *Science* 338, pp. 1202–1205.

Gurney, K., T. J. Prescott, and P. Redgrave (2001). "A computational model of action selection in the basal ganglia. I. A new functional anatomy". In: *Biological cybernetics*.

Hartman, M., E. Bolton, and S. E. Fehnel (2001). "Accounting for age differences on the Wisconsin Card Sorting Test: Decreased working memory, not inflexibility". In: *Psychology and aging* 16.3, pp. 385–399.

Heaton, R. K., G. J. Chelune, J. L. Talley, et al. (1993). *Wisconsin Card Sorting Test (WCST): manual: revised and expanded*. Psychological Assessment Resources (PAR).

Kajić, I., T. Schröder, T. C. Stewart, et al. (2019). "The Semantic Pointer Theory of Emotion: Integrating Physiology, Appraisal, and Construction". In: *Cognitive Systems Research*.

Komer, B., T. C. Stewart, A. R. Voelker, et al. (2019). "A neural representation of continuous space using fractional binding". In: *41st Annual Meeting of the Cognitive Science Society*. Montreal, QC: Cognitive Science Society.

Mackintosh, N. J. (1965). "Selective attention in animal discrimination learning". English. In: *Psychological bulletin*.

McClelland, J. L., D. E. Rumelhart, P. R. Group, et al. (1986). "Parallel distributed processing". In: *Explorations in the Microstructure of Cognition* 2, pp. 216–271.

Monchi, O., M. Petrides, V. Petre, et al. (2001). "Wisconsin Card Sorting Revisited: Distinct Neural Circuits Participating in Different Stages of the Task Identified by Event-Related Functional Magnetic Resonance Imaging". In: *Journal of Neuroscience* 21.19, pp. 7733–7741.

Parent, A. and L.-N. Hazrati (1995). "Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamocortical loop". In: *Brain research reviews* 20.1, pp. 91–127.

Plate, T. A. (1995). "Holographic reduced representations". In: *IEEE Transactions on Neural networks* 5.2.

Rougier, N. P., D. C. Noelle, T. S. Braver, et al. (2005). "Prefrontal cortex and flexible cognitive control: Rules without symbols". In: *Proceedings of the National Academy of Sciences* 102.20, pp. 7338–7343.

Stewart, T. and R. West (2010). "Testing for equivalence: a methodology for computational cognitive modelling". In: *Journal of Artificial General Intelligence* 2.2, pp. 69–87.

Stöckel, A., T. C. Stewart, and C. Eliasmith (2020). "A Biologically Plausible Spiking Neural Model of Eyeblink Conditioning in the Cerebellum". In: *42nd Annual Meeting of the Cognitive Science Society*. Toronto, ON: Cognitive Science Society, pp. 1614–1620.

Weinberger, D. R., K. F. Berman, and R. F. Zec (1986). "Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. Regional cerebral blood flow evidence". In: *Archives of general psychiatry* 43.2, pp. 114–124.