Unifying Neural Delay Representations in Cognitive Tasks: A Joint Human Behavioral and Recurrent Neural Network Study

Daniel Ehrlich (daniel.ehrlich@yale.edu)
Interdepartmental Neuroscience Program, 333 Cedar Street
New Haven, CT 06511 USA

John D. Murray (john.murray@yale.edu)
Department of Psychiatry, 300 George Street
New Haven, CT 06511 USA

Abstract

In this study we define contingency representations, a representational schema for delay tasks in which neural states encode prospective choice points, and demonstrate how such a representation unifies seemingly contradicting sensory-, action- and rule-based representations reported for prefrontal cortex neurons in different delay tasks. Further, we describe a novel experimental paradigm, the conditional delayed logic (CDL) task, in which we investigate competing theories of representational structures as they are utilized to perform varied working memory tasks. We trained a recurrent neural network to perform the CDL task, identifying a contingency representation subspace and testing its functional and mechanistic properties. Human subjects tested on the CDL task demonstrated behavior consistent with the contingency-based representational schema and inconsistent with many leading models of working memory. Contingency representations, in addition to clarifying neuronal delay tuning, provide a novel hypothesis for mixed selectivity as well as dynamic tuning observed during many working memory tasks. Lastly, we present a set of falsifiable predictions and analyses for neural data sufficient to differentiate contingency representations from alternative representational theories.

Keywords: Working Memory; Prefrontal Cortex; Representations; Recurrent Neural Networks

Introduction

The observation that neurons in the prefrontal cortex (PFC) retain a stimulus-selective persistent memory trace forms the backbone of working memory research in neuroscience (Funahashi, Bruce, & Goldman-Rakic, 1989). Since this discovery, however, the narrative of what it is PFC neurons are tuned to has has been complicated by multiple studies indicating that neurons in the PFC could be selectively tuned for future actions, expected upcoming stimuli, and task set/rule, in addition to "retrospective" tuning to past sensory information. The traditional explanations for these results have been either (a) that the PFC is multi-faceted and different subpopulations participate in different varieties of persistent encoding, or (b) that depending on the context, PFC neurons can be tuned to different task features. In this study we will show that a single representational schema, the contingency representation, can actually explain these disparate findings, with a single population encoding all relevant information and no switching in single-unit tuning.

The contingency representation is defined by a set of possible future actions. States in which an expected future stimuli will require the same response will be grouped together, while states with different expected response contingencies will be separated. By designing the conditional delayed logic (CDL) task, to cleanly separate sensory information, rule information, action information and contingency information we can identify the true underlying representational structure. Then through mapping subtasks of the CDL paradigm into subsets that match previous experimental paradigms we can demonstrate how a contingency representation gives rise to experimental observations mentioned above.

Conditional Delayed Logic Task

In order to illustrate the properties of different representations, we focus on a family of binary delayed logic tasks, for which the object is to apply some classification to two binary stimuli separated by a delay. For example, the OR task asks a subject to separate cases in which either stimuli is "1", from cases in which they are both "0". In all there are 16 possible separations, however, we will focus on 10 eliminating the two tasks in which the response is constant independent of stimuli, always respond "0" or "1", and the four cases in which the specific order of the observed stimuli matters. The full set of remaining tasks with their stimuli-output mappings are described in table 1. While the binary delayed logic task family will be the primary focus of this study, all theory expands to a much more general set of cases in which optimal action is contingent on future expected stimuli.

Importantly the CDL task family contains functional analogs to many well studied paradigms. The MEM task parallels the computation necessary for the Oculomotor Delayed-Response (ODR) Task (Funahashi et al., 1989), the XNOR and XOR task represents the Delayed Match to Sample Task and Match to non-Sample, respectively (Skinner, 1950), and the REPORT and Anti-REPORT mirror arbitrary stimulus-action response paradigms (Rainer, Rao, & Miller, 1999). By training a single network to do all of the above tasks we will be able to investigate how the same representation could subserve all these apparently different paradigms.

Table 1: Conditional Delayed Logic Subtasks

Rule	0,0	0,1	1,0	1,1
XOR	0	1	1	0
XNOR	1	0	0	1
OR	0	1	1	1
NOR	1	0	0	0
AND	0	0	0	1
NAND	1	1	1	0
MEM	0	0	1	1
Anti MEM	1	1	0	0
REPORT	0	1	0	1
Anti REPORT	1	0	1	0

Contingency Representations

In contrast to sensory or action representations, contingency representations are defined by a set of possible future actions. For example, in the AND task, if cue A is "1" then the contingency representation would uniquely identify the responses to either possible value of cue B. In this case if cue B is "0" the correct response would be "0" and if cue B is "1" the correct response would be "1". We would therefore say this is a [0,1] cue B contingency representation (Fig. 1). In this paper we will refer to cue B contingency as merely contingency for convenience.

Computing Through Contingency

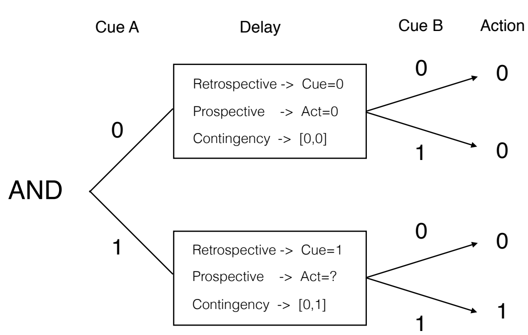

Figure 1: Retrospective Sensory, Prospective Action and Contingency Representations

The contingency representation is a fundamentally compositional solution to the CDL problem. Instead of computing the response as 10 separate functions of the two stimuli, using 4 intermediate products, [0,0], [0,1], [1,0], and [1,1] contingency states, every cue B can be resolve into its correct response. Contingency representations are then compositional in-between conditions in that different tasks will re-use the same contingency representations and compositional in-between stages of the task as a first function transforms the cue A and rule data into a correct contingency state and then a second function can be applied independent of how the system reached that state.

It is important to note that it is possible for a subject to complete the task with perfect accuracy without ever learning or utilizing a contingency-based representation of the CDL task. In fact the traditional sensory focused models would be sufficient under the right circumstances (Fig. 1).

Results

Human Behavior

To investigate the feasibility of contingency representations as an explanation for human behavior, we had human participants perform the CDL task. Subjects we're first shown a rule cue, describing the trial type which remained on the screen throughout the task. Following the rule cue a transient stimulus, cue A, depicting either a "0" or a "1" appeared on screen. After 200 ms the cue A stimulus disappeared. Following a 1000 ms delay, a second "0"/"1" stimulus, cue B, appeared on screen. Subjects could respond only after cue B appeared. The subjects were instructed to respond as quickly and as accurately as possible and there was a time-out penalty for responses over 2 s.

One important prediction of sensory representations is that the same classifier solves the problem of identifying the correct output independent of the cue A identify. In the contingency model, however, since different cue A by task combinations lead to different contingency states this is no longer the case. For this reason we would expect that performance on a task should be independent of cue A if the sensory representation is utilized, but not necessarily so for the contingency representation.

We found that this is the case. For trials in which the rule was OR, cue A = 0 trials had substantially longer response times (RTs) than cue A = 1 trials. In contrast for trials in which the rule was AND, cue A = 0 trials had substantially shorter RTs than cue A = 1 trials.

In a more detailed analysis we found this to be explained by the fact that [0,0] and [1,1] contingencies had shorter RTs than [0,1] or [1,0] contingency trials. A linear analysis of RTs revealed that significantly more variance was explainable by contingency than by rule, cue, or response.

Identifying Contingency Representations in an RNN Model

Further to investigate the mechanistic logic of a contingency representation implemented in a functional circuit we trained a task-optimized RNN on the CDL task and analyzed the state space trajectories of the resulting network.

Using a linear subspace identification method, we identified a plane in which we could separate trials by contingency. This demonstrated that the representation in our RNN was sufficient to perform contingency operations.

We used unsupervised dimensionality reduction (UMAP) and clustering methods (agglomerative clustering) to identify that contingency explained the global structure of the data better than alternative models both in euclidean and topological space. Using Adjusted Rand Score we identified that clusters were better explained by contingency, than either stimulus, action or rule.

Finally, we tested the functional relevance of contingency representations through a perturbation experiment, in which we slightly perturbed the state of the network just prior to cue B onset. We found that perturbations in contingency subspace caused significantly greater deficits in performance than did perturbations of equal magnitude orthogonal to the contingency subspace.

Identifying Sensory, Action and Rule Representations as sub-parts of Contingency

A key motivating features of this study, is the goal of explaining and unifying different PFC delay representations identified in the literature. By taking limited subsets of the CDL tasks, one can show how contingency representations can be made to appear as different forms of tuning.

For example if one were to only investigate the MEM condition of the CDL task, the contingency under cue A = 0 would be [0,0], while the cue A = 1 contingencies would be [1,1]. This separates into two differentiable persistent states in the contingency representation. For this reason an analysis of unit tuning would identify any contingency tuned cells as stimulus tuned. Sensory tuning will be found for an analysis of any single CDL task in which the contingency state depends on cue A.

Rule tuning can be made to appear as a function of the task as well in cases in which two or more CDL subtasks are analyzed together. In a task with just the NOR and AND conditions of the CDL task, since those tasks cover three contingency states [0,0],[0,1] and [1,0] units would appear to be tuned to both rule and cue. These hypothetical results match those reported in the rule tuning and task set literatures (Wallis, Anderson, & Miller, 2001). This is true for any pair of tasks in which there are three contingency states utilized.

To generate prospective action/response tuning, we can design a task with only the MEM and anti-MEM condition. Since the outputs after the contingency representations will always be ”1” on contingency [1,1] trials, and ”0” on contingency [0,0] trials delay activity will appear tuned to a prepared action(Rainer et al., 1999). This is true for any pair of tasks that both only move to [0,0] and [1,1] contingency representations after cue A.

By building subsets of the CDL task with different correlations between contingency states, cues, and rules, data with rule tuning, cue tuning, mixed tuning or neither tuning can be generated. Despite this when examined over the entire CDL task, which was designed to decorrelate cues, rules, actions and contingencies, the single contingency representation emerges.

Mixed Selectivity

One recent phenomena of significant computational interest is the propensity of neurons in PFC to encode a combination of multiple task features, or to be of nonlinear mixed selectivity. While nonlinear mixed selectivity does provide computational advantages (Rigotti et al., 2013), one alternative explanation could be that neurons code non-linear interactions of features as part of their network solution. Due to the fact that contingency representations are a cue $\times$ rule interaction, the units appear to be partially tuned to many different features of the task. Despite this fact many neurons are purely contingency selective, and only appear mixed because of the prior assumption that they are encoding sensory, rule, or motor features.

Dynamic Working Memory

Recently, even the persistent stable nature of working memory has begun to be questioned. Many researchers have begun to report dynamic encoding of stimuli between the cue epoch and delay (Cavanagh, Towers, Wallis, Hunt, & Kennerley, 2018), and subsequently computational models of this phenomenon have begun to emerge (Murray et al., 2017) . In contrast to those models which attempt to instantiate dynamic encodings, delay dynamics emerge naturally from the contingency representation RNN. This is because the mapping from cue A and rule to contingency state takes time, and this trajectory involves a rotation in sensory and rule readouts. If analyzed from the perspective of sensory tuning this appears to be a dynamic sensory code, when in reality it is a transformation from a primarily sensory representation during the cue epoch to a contingency representation in the delay epoch.

Predictions for Neural Data

Crucially, contingency representations make unique predictions to differentiate it from alternative models of working memory, both high dimensional randomly connected network (RCN) models (Rigotti et al., 2013) as well as from more traditional sensory models (Wong, 2006). First the dimensionality of contingency representations are low, as they are able to use compositional states that reduce total representational states. The effective dimensionality is lower than the number of conditions, and an order of magnitude lower than RCNs.

Second, contingency representations make separable predictions for inter-conditional similarity in a CDL-type task. Using representational similarity analysis, we show neural measures can by classified as more in line with a contingency similarity representation than a sensory similarity representation or rule representational schema.

Lastly, the problem of high dimensional noise and random non-linear mixtures makes many predictions vulnerable to a variety of assumptions. In order to avoid these problems, we applied a non-parametric variance partition analysis to our network. By separating the data into linear variance, nonlinear contingency variance, and other non-linear variance, and comparing contingency variance explained against permuted random non-linear interactions we demonstrate how we can identify whether an arbitrary system has more contingency variance then would be expected for that model class.

Discussion

The contingency representation can be thought of as an intermediate product compositional solution to the CDL task. Instead of trying to do the task in one step from the joint sensory cue representation, by transforming the computation into two stages a single complex classification problem can be turned into two relatively more straightforward ones.

As such this representation utilizes an important concept of time in recurrent systems as computation. It has long been established that recurrent neural networks are analogous to deep feedforward networks under certain constraints (Liao & Poggio, 2016). For this reason it would seem logical that any system would attempt to use this extra computational depth. This may be a possible explanation for why similar phenomena, of forward looking pre-computation, are identified in our human subjects and task optimized RNN.

Independent of the reason for the representation itself, however, identifying a possible single representation that can account for the diversity of persistent delay tunings observed in PFC neurons moves us in the direction of a unifying theory of PFC function.

Lastly, our novel CDL task capable of dissociating different possible representational schema will allow experimental groups to test the existence of contingency representations in humans in a neuroimaging context or with invasive recordings in an animal model.

Acknowledgments

We would like to acknowledge support from NIH R01MH112746. Further we would like to thank Dr. Alan Anticevic, Nicole Santamauro and Naja Hill for their assistance in setting up our behavioral experiment.

References

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., & Kennerley, S. W. (2018, August). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nature Communications*, *9*(1), 3498. doi: 10.1038/s41467-018-05873-3

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989, February). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349. doi: 10.1152/jn.1989.61.2.331

Liao, Q., & Poggio, T. (2016, April). Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex. *arXiv:1604.03640 [cs]*. (arXiv: 1604.03640)

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2017, January). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *114*(2), 394–399. doi: 10.1073/pnas.1619449114

Rainer, G., Rao, S. C., & Miller, E. K. (1999, July). Prospective Coding for Objects in Primate Prefrontal Cortex. *Journal of Neuroscience*, *19*(13), 5493–5505. doi: 10.1523/JNEUROSCI.19-13-05493.1999

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013, May). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. doi: 10.1038/nature12160

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*(4), 193–216. doi: 10.1037/h0054367

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001, June). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953. doi: 10.1038/35082081

Wong, K.-F. (2006, January). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, *26*(4), 1314–1328. doi: 10.1523/JNEUROSCI.3733-05.2006