Feedback Modulated Attention Within a Predictive Framework

Benjamin Cowley and John Thornton

Griffith University, School of ICT,
Institute for Integrated and Intelligent Systems
benjamin.cowley@griffithuni.edu.au
j.thornton@griffith.edu.au

Abstract. Attention is both ubiquitous throughout and key to our cognitive experience. It has been shown to filter out mundane stimuli, while simultaneously communicating specific stimuli from the lowest levels of perception through to the highest levels of cognition. In this paper we present a connectionist system with mechanisms that produce both exogenous (bottom-up) and endogenous (top-down) attention. The foundational algorithm of our system is the Temporal Pooler (TP), a neocortically inspired algorithm that learns and predicts temporal sequences. We make a number of modifications to the Temporal Pooler and place it in a framework which is inspired by predictive coding. We use a novel technique in which feedback connections elicit endogenous attention by disrupting the learned representations of attended sequences. Our experiments show that this approach successfully filters attended stimuli and suppresses unattended stimuli.

Keywords: Attention, Hierarchical Temporal Memory, Predictive Coding

1 Introduction

Attention lies at the heart of cognitive experience. It enables our conscious perception to focus upon specific elements within the vast and dynamic sensorium. It manifests in many forms: following an object along the horizon, concentrating on a melody, or mentally solving a mathematical problem. This ubiquity suggests that attentional mechanisms must be intrinsic to any truly biological approach to artificial intelligence.

It has been proposed that attention plays a role in the earliest levels of cognition and perception, acting to *filter* out stimuli that are not selected as the target of attention [3, 13]. Under this paradigm only the attended stimulus reaches the highest levels of cognition, while unattended stimuli are filtered out at the early levels of sensory perception. This filtering of stimuli is also reflected in neural recordings, where attention has been shown to enhance the responses of neurons in the neocortex that encode the attended stimulus, while simultaneously suppressing that of unattended stimuli [17, 21]. The ability to filter specific stimulus

has obvious advantages to artificial intelligence systems (e.g. reducing the problem space), as such there has been a renewed interest in applying attentional mechanisms to connectionist systems in recent years (we discuss some of this work in Section 2).

Our model fits into a broad body of work that understands the brain as a prediction machine which self-organises to form generative predictions (or hypotheses) of its current and future states. Predictive coding [19, 4] has emerged as the most promising interpretation of this theory, with top-down and lateral predictions suppressing the responses of feature encoding *error-units*. This flow of information and forming of hypotheses has since been generalised as free-energy minimisation by Friston [8].

Our model attempts to reconcile attentional filtering with predictive suppression of stimuli. Lateral predictions (formed in the same neocortical region) suppress the feedforward output of that region. Surprising stimulus (not predicted) are communicated as feedforward output to the higher regions. This forms an exogenous (bottom-up) attentional mechanism based on the Bayesian surprise theory of exogenous attention, where the least predictable stimulus is the most salient [11] (note that free-energy can also be formulated as Bayesian surprise [8]). Endogenous (top-down) attention is modulated by feedback that causes a targeted disruption in the learned representations. This disruption inhibits predictions on attended stimuli, and thus the attended stimuli is output from the region using the same feedforward pathway as endogenous attention.

For simplicity, and to focus on the mechanisms of extracting information using attention, we implement our model in a single layer system. The foundational algorithm for this system is the Temporal Pooler (TP) [9]. TP is a connectionist algorithm that has been shown to perform strongly in the domain of on-line learning anomaly detection [15]. To the TP we add feedback connections, new types of neurons, and place it in a predictive framework inspired by predictive coding; we refer to this system as Temporal Pooler plus Attention (TP+A).

The TP algorithm is based on the Hierarchical Temporal Memory (HTM) model of the neocortex [10], a predictive model similar to predictive coding, and employs a number of components directly based on neocortical biology. TP neurons self-organise using a Hebbian learning inspired method to form *synapses* to a subset of other neurons, in contrast to many deep learning systems that use less biologically plausible methods, such as backpropagation on fully connected neurons [2]. To implement attentional mechanisms we use the same basic learning policies and structures as TP, thereby inheriting the biological plausibility of the HTM approach. In this way TP+A provides a model for how attention may be implemented in the neocortex, while simultaneously providing a proof-of-concept for a system that could be incorporated into future AI systems.

2 Related work

In recent years there have been an increasing number of studies applying attentional mechanisms to connectionist systems, with much of this work focusing

on visual attention. One such approach is to select only part of an image to be processed at high resolution [24]. This method has been successfully applied to a number of domains, including object tracking [5], recognition [24, 5], and image caption generation [1]. Another method applied to connectionist systems is to use attentional mechanisms to modulate representational nodes in the system. Wang et al. [25] used two separate neural networks, one encoding the input and the other encoding top-down prior beliefs of the input's class; the output vectors of both networks were combined to produce a modified representation of the output. This approach was applied to classifying and de-noising handwritten digits. Attention inspired techniques have also been used to improve the classification of images using convolutional neural networks [23]. Here feature nodes of the network were modulated over successive time-steps using a reinforcement learning policy.

There have been a number of models that attempt to reconcile various attention related phenomena with predictive coding. Rao and Ballard [20] expanded their earlier work on predictive coding in the visual cortex [19] by showing how attentional visual search may work. They applied an outlier mask that suppressed stimuli which least conformed to a generative model, while making stimuli that were more likely under the generative model to be more salient. Subsequently Spratling [22] also expanded Rao and Ballard's original work by showing that their equations are mathematically identical to some models of bias competition, a theory that attention emerges through the modulation of representational nodes by bias weighting [6]. To demonstrate this model, feedback signals, which simulated endogenous attention, were fed into the system and resulted in phenomena consistent with binding. Perhaps the most prevalent theory of attention in predictive coding is that of precision weighting [7, 4]. This is achieved by increasing the 'gain' on error units that are predicted to provide the most precise information vis-à-vis the current environment.

Applying TP to a framework based on predictive coding is similar to work of McCall and Franklin [16], who embedded TP in a predictive coding framework and tested it for robustness to noise on random temporal sequences. Their system uses two hierarchical layers, where the feedforward output from the bottom layer is the prediction-error. This is formed by subtracting the state of the bottom layer from feedback sent from the top layer. Feedforward and feedback use bi-directional connections, in contrast to a method introduced by Kneller and Thornton [12] which uses separate, more biologically plausible, uni-directional connections. Here, feedback connections are learned using TP's learning method, while feedforward connections use the HTM spatial pooler algorithm.

3 Temporal Pooler plus Attention

TP+A is designed to perform five tasks: 1) form predictions on temporal sequences; 2) output prediction errors; 3) output temporal sequences that are the target of attention (attentional filtering); 4) learn temporal sequences; 5) learn relationships between attention signals and the temporal sequences. Tasks 1 and

4 are performed using the TP algorithm (described in subsections 3.1 and 3.4), task 2 is accomplished by embedding the TP in a framework inspired by predictive coding (described in Subsection 3.2), and tasks 3 and 5 are achieved using our attentional feedback mechanism (described in subsections 3.3 and 3.4).

3.1 Predicting temporal sequences

TP+A forms predictions on feedforward input formatted as sparse temporal sequences; we use the TP algorithm to make these predictions. TP was initially developed as part of the Cortical Learning Algorithms package (CLA) [9], which also included the HTM spatial pooler. TP compromises a number of structures named *columns*, which are based on mini-columns found in the neocortex [18]. Columns can be set into an *active-state* by feedforward boolean input. The resulting activation and deactivation of the columns over successive time-steps forms the temporal sequences on which the system learns and predicts.

Each column contains a number of artificial neurons called *prediction-cells* [9]. Prediction-cells have a number of dendrite *segments*, and each segment contains a number of *synapses*. Synapses are uni-directional connections to prediction-cells in other columns that become active when the prediction-cell they are connected to is in an active-state. When the number of active synapses in a segment is greater than the value of parameter $actiThreshold$ the segment enters an active-state. This, in turn, sets the prediction-cell into a *predictive-state*. When a column enters an active-state and one of its prediction-cells was in a predictive-state, that prediction-cell enters an active-state. If, however, a column is in an active-state and no prediction-cell was in a predictive-state then all prediction-cells in that column enter an active-state, representing that any number of temporal features could have activated the column. It is through this method that TP encodes and produces predictions on temporal sequences; for a more in-depth discussion of the algorithm see the CLA white paper [9].

3.2 Outputting the prediction error

TP+A applies the TP within a framework based on predictive coding. Our method differs somewhat from McCall and Franklin [16] who used bi-directional connections between levels to form and communicate errors and predictions. Because our study only concerns a single level we rely on only the lateral predictions (formed by TP) to detect errors. The errors are output by adding a new type of cell to the TP called an *error-cell*. Each column has one error-cell. When a column is active and this activity was not predicted by a prediction-cell then the error cell will be in an active-state. The state of each error-cell will comprise the feedforward output of TP+A. In Figure 1b we provide a diagram of a TP column and its connections embedded in this system.

3.3 Attentional filtering

Our model uses top-down signals to elicit endogenous attention. In TP+A we achieve this using feedback *axons* that input sparse codes into the system. A

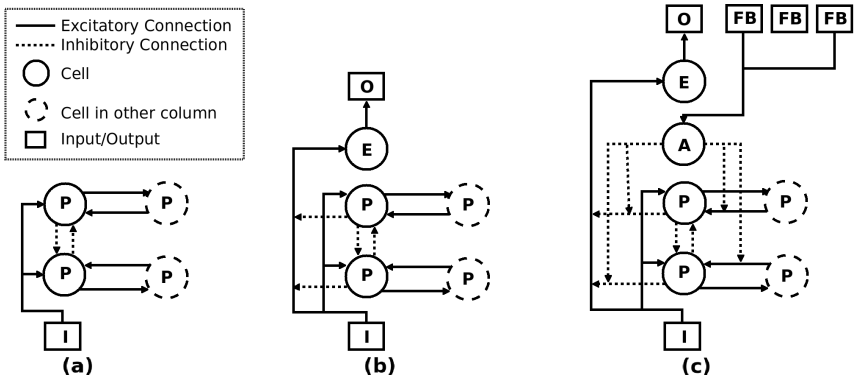

Fig. 1. Diagram of connections within columns. P: prediction-cell, I: input, O: output, E: error-cell, FB: feedback input, A: attention cell. Arrows indicate direction of information flow. (a) TP column; excitatory connections between prediction-cells can put them into a predictive state, if placed into an active state when in a predictive state, they will inhibit other prediction-cells in their column. (b) TP column embedded in a predictive architecture; output is produced by an error-cell, which is set into an active state by a connection to the input, active prediction-cells inhibit this connection. (c) TP+A column; an attention-cell is excited by feedback axons. When active, the attention-cell inhibits the inhibitory connections to the input/error-cell connection and also the dendrite segments of that synapse to prediction-cells in its column.

new type of cell, the *attention-cell*, associates this sparse code with activation of it's column. Each column has a single attention-cell and these have a number of segments with a number of synapses that can connect to an axon. The activity of these synapses and segments determine whether or not the attention-cell is placed into an active state using the same method that determines the predictive-state of prediction-cells (outlined in Subsection 3.1). The method we use for connecting to feedback is the same as Kneller and Thornton [12], however in experiments they used the Spatial Pooler algorithm and not the TP (so time-steps were not a factor); the feedback also did not elicit endogenous attention.

If an attention-cell is in an active-state, then the error-cell of its column will also be in an active-state whenever the column is active-state, even if this activity was predicted. This causes attended sequences to be output by the error-cells, where usually they would be suppressed. By using the error-cells to output attended sequences we remove any need for adding new output channels or separate data representations. A second effect elicited by an attention-cell when in an active-state is that it will *inhibit* (unable to become active) segments of prediction-cells in other columns that have synapses to prediction-cells in its column. Columns which have prediction-cells that have been inhibited in this way are more likely to be in an active-state that was not predicted, due to the disruption of the prediction process caused by the inhibition. This, counter-intuitively, is of benefit as the segments that are inhibited are likely to be forming predictions based on the attended sequence. Thus, column activations caused by an attended

sequence can be output by the error-cells even if that column's attention-cell has not learned the feedback pattern, preserving the associations between column activations learned by TP when outputting an attended sequence. Figure 1c provides a diagram of a column with the attention-cell and its connections.

3.4 Learning

Both prediction-cells and attention-cells use the same TP Hebbian-based learning algorithm [9]. In prediction-cells this algorithm governs the creation and destruction of synapses to prediction-cells in other columns, while in attention-cells it governs the connection of synapses to feedback axons. For improved clarity we have also included pseudo-code in Algorithm 1 which has been generalised for use with both prediction-cells and attention-cells.

Algorithm 1 Learning Under the Markov Assumption

Input: $column$ //column learning is to be performed on
Input: t //current time-step
Input: $newSyns, minThreshold, connThresh, permInc, permDec$ //parameters
1: **if** $column.isActiveAndNotPredicted(t)$ **then**
2: $potSyns \leftarrow getActivePotentialSynapses(t-1)$
3: $segment \leftarrow findClosestSegment(column, potSyns, minThreshold)$
4: **if** $segment = null$ **then**
5: $cell \leftarrow getRandomCell(column)$
6: $createNewSegment(cell, potSyns, newSyns)$
7: **else**
8: $addNewSynapses(segment, potSyns, newSyns)$
9: **else if** $column.isActiveAndPredicted(t)$ **then**
10: **for each** $cell$ **in** $column$ **where** $cell.inPredictiveState(t-1)$ **do**
11: **for each** $segment$ **in** $cell$ **where** $segment.active$ **do**
12: **for each** $synapse$ **in** $segment$ **where** $synapse.active$ **do**
13: $incrementPermanence(synapse, permInc, connThresh)$
14: **else if** $column.isInactiveAndPredicted(t)$ **then**
15: **for each** $cell$ **in** $column$ **where** $cell.inPredictiveState(t-1)$ **do**
16: **for each** $segment$ **in** $cell$ **where** $segment.active$ **do**
17: **for each** $synapse$ **in** $segment$ **where** $synapse.active$ **do**
18: $decrementPermanence(synapse, permDec, connThresh)$

Cells are initialised with no synapses or segments, these are first created in response to column activity. Whenever a column is in an active-state and this was not predicted by any of its prediction-cells (or, in the case of learning feedback, its attention-cell was not in an active-state), we add new synapses to a cell (lines 1-8). We search all cells for a single segment that has the greatest overlap with the set of potential synapses (other prediction-cells for prediction-cell learning, or feedback axons for attention-cell learning) that were active in the previous time-step; we then add a number of synapses up to the value of

parameter *newSyns* to the chosen segment (line 8). If no segment had an overlap above parameter *minThresh* then we add a new segment to a random cell (line 6), this segment will have the value of *newSyns* of the potential synapses. Each synapse has a *permanence* value; when the permanence value is above parameter *connThresh* the synapse is *connected* and can affect their cell's state, otherwise it will be *disconnected* and cannot affect their cell's state. Synapse permanence is decremented by the value of *permDec* whenever the synapse contributes to a cell falsely predicting its column will be active (lines 14-18; disconnecting of a synapse is handled by *decrementPermanence*()). Synapses have their permanence incremented by the value of *permInc* whenever they were in an active-state and their segment correctly predicts their column will be active, even if they are disconnected (lines 9-13; connecting of a synapse is handled by *incrementPermanence*()). This is the method that the TP uses for learning under the Markov assumption (only the current time-step can predict the next), but the TP can learn to predict further in time by engaging this method to learn the cells which were active the time step prior to a successful prediction. However, for efficiency we learn under the Markov assumption in this treatment.

4 Experiments and Analysis

To test whether the TP+A can successfully attentionally filter sequence input we performed experiments using two separate input types: *burst sequences* (which allows us to easily visualise the output) and *frequent feature sequences* (to test filtering when a subset of columns in a sequence are persistently active).

We use a similar experimental design across all tests. The TP+A has 256 columns and at each time-step is fed a sparse binary input of length 256, where one element activates one column. We use a single iteration of a 150,000 time-step training set; between time-step 100,000 and 125,000 whenever a target sequence is active we apply feedback input that simulates top-down input from a higher level. The feedback is a randomly generated sparse binary pattern with 256 elements where each element has 0.025 of been set to one. Each element of this pattern corresponds to a single feedback axon, an element set to one sets its corresponding axon to active. After training, we switch off learning and use a 10,000 time-step test set that we apply both with attentional feedback and without. We used the prediction error on the test set to tune the values of *permInc*, *permDec*, *connPerm*, and *actiThreshold* for both prediction-cell and attention-cell synapses. The number of cells used is 15; exploratory experiments showed that this number performed robustly across both sequence types. The quantitative results given are averaged across 10 experimental runs, using different random seeds for the TP+A. To quantify output for particular sequences we use the *sequence error* metric, this is the number of activations of error-cells divided by the number of column activations for each time-step after the first (we exclude the first as a sequence begins at random in our experiments).

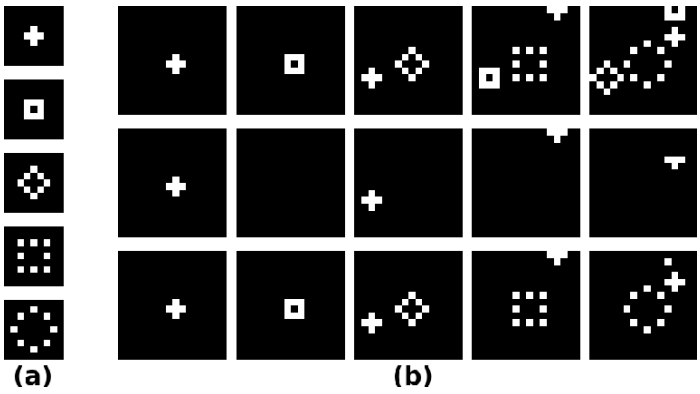

Fig. 2. (a) A sample of a sequence used during the burst sequence tests; white squares are active columns, black squares represent no column activity; the top image is the first time-step, and the bottom image is the fifth and final time-step. (b) A sample of five time-steps (running left to right) of testing; the first row is the input, second row is the system without attention active, and the third row is the system with attention on; the middle burst is target of attention.

4.1 Burst Sequences

For the first of our experiments we use feedforward input that comprises sequences which, when formatted in two dimensions, form a distinct visual 'bursting' pattern. These type of sequences were used to allow us to better display the attentional mechanisms; an example of the sequence is shown in Figure 2a. The sequences are five time-steps long and can occur at 64 possible starting positions. Given that a single TP/TP+A region is not translation invariant, each starting position constitutes a separate sequence. We used the following methodology for producing the input: at each time-step a sequence has a probability of 0.005 of becoming active (unless it is already active, in which case the probability is 0.0); if no sequence is currently active we randomly select one to become active (this is done to ensure there is column activity on every time-step). We select one sequence to be the target sequence for the attentional feedback mechanism.

A sample of the output from TP+A during testing is illustrated in Figure 2b; where the sequence located in the centre is the target of endogenous attention by way of feedback. In the second row we see that when there is no active feedback TP+A suppresses the majority of input; as sequences begin at random, the first instances will be mostly unpredicted. The third row displays similar suppression as the second, except for the target burst which is output in its entirety. However, some imperfections in the system are also illustrated; in the last time-step we can see that one element of a newly beginning sequence is suppressed without attention, while during attention it is not suppressed, as is an element of the sequence beginning the previous time-step. These irregularities are caused by the interactions between simultaneous sequences and the attention mechanism. The predicted element in the newly beginning sequence would, had the sequence not

started, be a false prediction. However, when attention is active the disruption caused by the attention cell incorrectly causes output not associated with the target sequence. This is reflected in a quantitative analysis: during testing when attention is not active sequence error for the unattended sequences is 0.02, rising to 0.22 when it is active. However countering this is the sequence error for the attended sequences where the mean average for this set is a perfect 1.0 (i.e. the entirety of the attended sequences is output). These results indicate that on this set TP+A's attention mechanisms have successfully learned to output attended sequences, although there is some residual output of unattended sequences.

4.2 Frequent Feature Sequences

The frequent feature sequences experiments are designed to test TP+A's attentional mechanisms when the sequences contain frequently recurring features. This is of interest as attended stimuli commonly has such features (e.g. a stationary object, or auditory frequencies). To build the target sequence we chose 15 distinct input elements at random; each of these elements we assign a probability, p, that it will be active on any give time-step (8 have $p = 0.2$, 4 have $p = 0.4$, 2 have $p = 0.6$, 1 have $p = 0.8$). We also have five background sequence with 30 elements chosen at random (these each have $p = 0.2$), these five sequences are concatenated to form a 100 time-step long background sequence that is fed into TP+A and continuously looped during training and testing. The target sequence will be fed into the system at random time-steps (with a probability of 0.01; or 0.0 if it is already active) and will overlap with the background sequences.

The results from these experiments show an improvement over the bursting sequences vis-à-vis the sequence error for non-attended sequences (the background) during attention: with an average of 0.02; compared with 0.01 with no attention. The target sequence averages 0.95 sequence error during attention, compared with 0.02 when not attended. These results indicate that features occurring frequently within a single sequence may improve the capability of TP+A in separating the target sequence. In Figure 3a we have included a graph of error-cell activity during endogenous attention, note that while error-cells related to target sequence are very active, those for the background are much less so.

To ascertain the exogenous attention capabilities of the system, we inserted an extra sequence (constructed with the same methodology as the target sequence) during testing. There was no training on this sequence, so the system should be 'surprised' by the sequence and output it as error. We graph these results in Figure 3b, as can be seen this sequence is highly active, while the background sequence is largely suppressed. The average sequence error for the surprising sequence is 0.97, while the background is >0.01. This shows that the TP+A outputs surprising input, while suppressing predictable input.

5 Discussion

In this work we have presented a model implemented as a connectionist system, TP+A, which is based on two separate theories of neocortical function, HTM

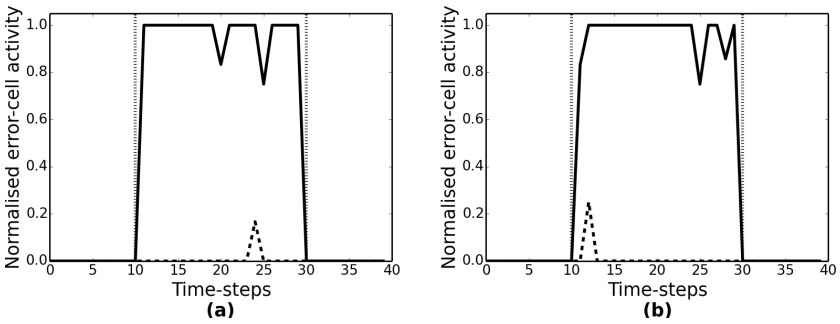

Fig. 3. Graph of error-cell activity, normalised so maximum possible activity is 1.0; solid line is target sequence, dashed line is background sequences, vertical dotted lines designate beginning and end of sequence. (a) Error-cell activity during endogenous attention. (b) Error-cell activity with a surprising sequence (exogenous attention).

and predictive coding. This system is designed to attentionally filter sequences using biologically plausible methods such as feedback, inhibition, and Hebbian learning. Results from our experiments, provided in Section 4, show that TP+A is capable of filtering out mundane (predicted) input sequences while simultaneously outputting sequences that are attended to. This paradigm of attention is in line with results from cognitive studies that show early levels of perception filter out unattended stimulus while conveying attended stimulus [13]. The use of feedback connections to illicit this type of attention makes this mechanism akin to endogenous attention, where the higher levels of cognition (or, in our case, higher levels of the hierarchy) control the attentional mechanisms of the lower levels. As well as endogenously attended input, TP+A will also relay any input that is surprising (unpredicted). Because TP+A uses the same output channel for both surprising and attended input, higher levels of the hierarchy would treat these signals identically. This is advantageous because a surprising stimulus should, and does, attract attention; studies have shown that in free viewing exercises participants attention is directed to the most 'surprising' features of scenes [11].

TP is designed to be a general purpose algorithm, capable in operating under any temporal modality. TP+A inherits this and adds to it mechanisms for both exogenous and endogenous attention. This goal of generality sets aside TP+A from many other connectionist attention systems which are specific to visual attention [5, 14]. We also use biologically inspired learning methods that exist in the original TP in contrast to systems which apply the less biologically plausbile backpropagation (such as [25]), or systems that require the combination of divergent techniques (such as backpropagation and reinforcement learning [23]).

TP+A offers two advantages over the precision weighting accounting of attention in predictive coding [7]. Firstly, TP+A has the internal resources to calculate the precision of error signals without predictive coding's need of a secondary system that learns to predict such precisions. This can be achieved by an analysis of the state of the TP (within a HTM hierarchy): here, a high precision error state

is indicated by a small number of temporally extended sequences, whereas a low precision state is indicated by a larger number of shorter sequences. Secondly, the TP+A approach does not require that we only attend to those aspects of a feature that are associated with high precision error signals. So, for example, we can endogenously attend to features that are perfectly predicted (and so emit no error signals), or we can attend to aspects of a feature associated with relatively low precision and ignore aspects with high precision errors. The phenomenology of ordinary experience suggests that we can endogenously attend in this way, but existing predictive coding models have difficulty explaining this. Our model of attention matches more closely to that of Spratling [22], who also used simulated feedback to stimulate endogenous attention. However, this model was focused on binding (where disparate features are 'bound' into a singular object), whereas ours reconciles predictive suppression with filtering. Binding in our model, could be achieved in a hierarchical system where associations between different input streams are learned at higher levels of the hierarchy. However, with our use of the TP algorithm, TP+A could be said to apply binding of *locally* encoded features, which are then fed upward due to the dendrite inhibition mechanism.

Future work will focus on the incorporation of TP+A into a hierarchy, where higher layers would need mechanisms to automatically elicit endogenous attention from lower layers, instead of simulating this feature as we did in this treatment. A fully functioning HTM hierarchy that is capable of action, attention, and recognition is still only theoretical. Through the inclusion of a mechanism for attention we believe we have made a significant step towards this goal.

6 Conclusion

We have presented a model for attention in a framework where prediction errors are suppressed. We proposed that endogenously triggered attentional filtering could be achieved through the targeted disruption of predictions. To implement this model we placed the neocortically inspired TP algorithm into a framework inspired by predictive coding. We added feedback mechanisms and a new neuron, the attention-cell; we refer to this connectionist system as TP+A. Our experiments show that TP+A successfully displayed phenomena consistent with both endogenous and exogenous attention. Future work will focus on integrating TP+A into a hierarchical system.

References

1. Ba, J., Salakhutdinov, R.R., Grosse, R.B., Frey, B.J.: Learning wake-sleep recurrent attention models. In: Advances in Neural Information Processing Systems. pp. 2575–2583 (2015)
2. Bengio, Y., Lee, D.H., Bornschein, J., Lin, Z.: Towards biologically plausible deep learning. arXiv preprint arXiv:1502.04156 (2015)
3. Broadbent, D.E.: Perception and communication. Oxford University Press (1958)
4. Clark, A.: Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press (2015)

5. Denil, M., Bazzani, L., Larochelle, H., de Freitas, N.: Learning where to attend with deep architectures for image tracking. Neural computation 24(8), 2151–2184 (2012)
6. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Annual review of neuroscience 18(1), 193–222 (1995)
7. Feldman, H., Friston, K.J.: Attention, uncertainty, and free-energy. Frontiers in human neuroscience 4 (2010)
8. Friston, K.: The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11(2), 127–138 (2010)
9. Hawkins, J., Ahmad, S., Dubinsky, D.: Hierarchical temporal memory including htm cortical learning algorithms. Techical report, Numenta, Inc, Palto Alto (2010)
10. Hawkins, J., Blakeslee, S.: On intelligence. Macmillan (2007)
11. Itti, L., Baldi, P.F.: Bayesian surprise attracts human attention. In: Advances in neural information processing systems. pp. 547–554 (2005)
12. Kneller, A., Thornton, J.: Distal dendrite feedback in Hierarchical Temporal Memory. In: Proceedings of the 2015 International Joint Conference on Neural Networks (2015)
13. Lachter, J., Forster, K.I., Ruthruff, E.: Forty-five years after broadbent (1958): still no identification without attention. Psychological review 111(4), 880 (2004)
14. Larochelle, H., Hinton, G.E.: Learning to combine foveal glimpses with a third-order boltzmann machine. In: Advances in neural information processing systems. pp. 1243–1251 (2010)
15. Lavin, A., Ahmad, S.: Evaluating real-time anomaly detection algorithms the numenta anomaly bench. In: 14th International Conference on Machine Learning and Applications (2015)
16. McCall, R., Franklin, S.: Cortical learning algorithms with predictive coding for a systems-level cognitive architecture. In: Second Annual Conference on Advances in Cognitive Systems Poster Collection. pp. 149–166 (2013)
17. Moran, J., Desimone, R.: Selective attention gates visual processing in the extrastriate cortex. Science 229(4715), 782–784 (1985)
18. Mountcastle, V.B.: The columnar organization of the neocortex. Brain 120(4), 701–722 (1997)
19. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience 2(1), 79–87 (1999)
20. Rao, R.P., Ballard, D.H.: Probabilistic models of attention based on iconic representations and predictive coding. Neurobiology of Attention pp. 553–561 (2004)
21. Reynolds, J.H., Chelazzi, L.: Attentional modulation of visual processing. Annu. Rev. Neurosci. 27, 611–647 (2004)
22. Spratling, M.W.: Predictive coding as a model of biased competition in visual attention. Vision research 48(12), 1391–1408 (2008)
23. Stollenga, M.F., Masci, J., Gomez, F., Schmidhuber, J.: Deep networks with internal selective attention through feedback connections. In: Advances in Neural Information Processing Systems. pp. 3545–3553 (2014)
24. Tang, Y., Srivastava, N., Salakhutdinov, R.R.: Learning generative models with visual attention. In: Advances in Neural Information Processing Systems. pp. 1808–1816 (2014)
25. Wang, Q., Zhang, J., Song, S., Zhang, Z.: Attentional neural network: Feature selection using cognitive feedback. In: Advances in Neural Information Processing Systems. pp. 2033–2041 (2014)