Title

A modeling link between cognitive and biological homeostasis

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors

Jenkins, Gavin
Barnes, Jordan
Tupper, Paul
et al.

Publication Date

2017

Peer reviewed

A modeling link between cognitive and biological homeostasis

Gavin Jenkins (gjenkins@sfu.ca)

Jordan I. Barnes (jordanb@sfu.ca)

Paul Tupper (pft3@sfu.ca)

Mark Blair (mblair@sfu.ca)
Simon Fraser University
8888 University Dr., Burnaby, BC,
Canada, V5A1S6

Abstract

The problem of stability has long been a limiting factor in developing neural networks that can grow in size and complexity. Outside of particular, narrow parameter ranges, changes in activity can easily result in total loss of control. Human cognition must have reliable means of acting to stay within the stable ranges of sensitivity and activation. Learning is one such mechanism, and population dynamics are another. Here, we focus on another, often overlooked stability mechanism: cellular homeostasis through metabolism dynamics. We ran a visual change detection experiment designed to strain network stability while minimizing any learnable patterns. We fit the data using models with and without cellular energy levels as a factor, finding that the model influenced by its past history of energy use was a closer fit to the human data.

Keywords: Homeostasis; attention; visual change detection; neural modeling

Introduction

The stability of learning and development depends partly on the stability of the learner's underlying cognitive machinery. A system that is not tethered to a baseline level of activity is vulnerable to being excited out of control or perishing through complete inaction. A number of mechanisms have been uncovered that promote basic stability in neural systems across multiple timescales. For instance, neurons coupled to one another via patterns of local excitation and lateral inhibition, can interact over fractions of a second to stably form and maintain "peaks" of activation around a core value of interest (e.g. Thelen, Schöner, Scheier, and Smith, 2001). As patterns in the short term persist, memory traces can be formed that are able to project the influence of these patterns over much longer timescales, leading to overall stability across similar situations.

Stability can also be driven at a level even below that of populations of neurons. Cellular metabolic processes allow individual neural units to contribute to the stability of a population coded representation by, observing and acting on their own changes in activity, and doing so at multiple timescales. The Hebbian rule, relating changes in correlations in the activity of neurons to their degree of co-activation, under-specifies the adjustment mechanisms needed for higher order stability. Oja (1982) derives an additional term, that, if included in the instantaneous rate of Hebbian weight change, will remain within some stable range of activity while maintaining the correlation. It was suggested that this term could be thought of as a form of intrinsic "leakage" rate, η, of the materials available to the synapse.

This initial modification to the Hebbian rule was largely abstracted, however, from the precise biological interpretation. The cellular mechanism would need to toggle the adaptation of a neural unit between "labile" and "stable" dispositions toward changing connection strengths (Bienenstock, Cooper, & Munro, 1982). One suggested candidate for this process is brain-derived neurotrophic factor (BDNF) (Glaser & Joublin, 2011). Using Calcium levels as a proxy for the instantaneous levels of change at a synapse, neural units coding for changes in the level of BDNF can dynamically alter the underlying synaptic excitation/inhibition levels of cells. For instance, blocking input channels to the retina using a tetrodotoxin can affect cellular activity in ways that suggest homeostatic forces (Turrigiano, 2011). Strong excitatory inputs, when blocked, allow for a period of higher than normal activity once unblocked; likewise, the opposite is shown when inhibitory inputs are blocked. More generally, energy stores build up in the blocked neurons when receiving lower than normal input (and thus experiencing lower than normal activity themselves), or when allowed to fire higher than normal, stores are depleted. When a baseline level of energy is restored, normal conditions are eventually achieved again, but the effect in the meantime is one of an internal, cellular homeostatic force.

In the present study, we use a visual change detection paradigm to explore the capacity of the cognitive system to adapt to changes in task demands. Our model of the data are more detailed than that described by Oja (1982), but more abstract than a chemically detailed BDNF explanation. Our goal was to create perturbations of cognitive homeostasis that can produce interesting behavioral level data suitable for modeling effects beyond those observable in cellular recordings. The specifics of two computational models are then introduced for capturing the new behavioral effects. Simulations show that the model with a cellular energy term outperforms a version without one. We conclude with a discussion on the merits of including energy terms in basic neural models as a part of establishing a common language of conservation across the brain, even when exploring cognition at the level of behavior.

Change Detection

The change detection paradigm is a useful way of studying the cognitive operations necessary for several processes relevant to homeostasis. A typical trial in a change detection task consists of a sample set of items to remember, followed by a test set of items and a response prompt. The paradigm is simple yet challenging to model (Johnson, Spencer, Luck, & Schöner, 2009); so we opted for a highly simplified version with only one feature dimension of change: color. By changing anywhere from zero to six colors per trial, however, we still allowed for a straightforward manipulation of homeostasis across trials, disturbing the balance of expectation, ability, and adaptation constantly throughout our task.

Experiment

We designed a change detection experiment with the intention of placing participants in situations of rapidly changing cortical energy levels as might be consistent with a BDNF homeostatic response. The task was also designed as a situation where homeostatic control would potentially be beneficial to task performance. Two presentations of colored squares were given to participants per trial, and the number of changes between presentations was reported, changing from zero to six changes unpredictably trial by trial. No overall learnable trial-to-trial pattern was available to participants that would aid them in answering correctly, so that there was no advantage to adopting complex expectations.

Method

Participants were recruited from the Simon Fraser University psychology department subject pool, where they received course credit for 30 minutes of participation. We asked participants to think of 6 color patches as alien creatures that would change form over the course of a trial. The job of the participant was to correctly classify the number of changes as part of a national effort to better understand these aliens. Of the 33 subjects, 30 were retained for analysis. Three participants were dropped for failing to complete 106 trials of the task.

Each trial involved a masked presentation sequence designed to eliminate any relevance of spatial position of stimuli, so that color alone was the sole feature dimension of change detection. Subjects were instructed to view a fixation cross for two seconds at the start of each trial. A set of six colors was then presented for four seconds, long enough to ensure an ability to briefly encode the colors in memory. The screen was then masked to block effects of afterimages for two seconds, and a second set of colors was presented. The orientation of the colors was a vertical 2x3 grid rather than the horizontal 3x2 grid in the first presentation, to remove any clear or objective correspondence between spatial locations in the first and second presentations. All colors were also scrambled in positions, in addition to the display positions being rotated.

The second color display was left on the screen for either 2.5, 3.0, 3.5, or 4.0 seconds (counterbalanced across trials per participant) to allow for different amounts of time for any homeostatic system to adjust cellular activity rates now that the changes, if any, were visible. The intention of the time manipulation was that there would not be sufficient time to make a homeostatic adjustment in one timeframe, but enough in another, and thus that the required time for homeostatic adjustment could be identified by changes in accuracy.

A second mask was then presented for two seconds, whereafter the participant was given any amount of time to report the number of changes on a number line on screen. Participants were given explicit feedback about their answer in the form of a blue mark for the correct answer in addition to the red mark indicating the chosen response. Figure 1 depicts a graphical representation of the task procedure.

Twelve perceptually equidistant colors were assigned to each participant, offset from each other by a random amount. Color coordinates were obtained from a slice of CIEL*a*b* color space, with luminance set to a constant $L = 75$. Only coordinates in this color space that have a representation in RGB can be displayed on a monitor, so all but the largest circular portion of the color space satisfying this constraint was removed (Johnson et al., 2009; Lehky & Sejnowski, 1999). These twelve colors were then used consistently for a participant's entire experiment. Every trial randomly sampled from the participant's set of twelve colors as needed. A "zero changes" trial, for example, only required six colors (the same set of six twice), whereas a "six changes" trial required all twelve colors (two different sets of six).

Trials were counterbalanced with a customized Latin square in order to equally distribute the number of times each possible number of changes in colors (0-6) was the correct answer, while also equally distributing the number of one-back differences between correct answers on the current and previous trial. Figure 2 shows the actual number of total trials across participants of each combination of these variables. The four possible display times for the second color array were evenly distributed within each of these trial types (i.e., within each bar in Figure 2. Display times not shown in Figure).

The partial Latin square was necessary due to mathematical constraints in designing a distribution of trials that attempted to fit both criteria. Each trial is part of a one-back link to its previous trial but also to the next trial, so any change has a cascade of consequences for the options on other trials. Also, certain one-back differences are impossible; for example, if the number of changed colors on a trial is three, then the one-back difference cannot possibly be 6, because that would mean the correct answer on the previous trial was "-3 changes" which is not possible. There were 9 impossible combinations like this overall, forming a triangle of missing bars in Figure 2 (upper right portion of figure).

Trial orders were generated by a Monte Carlo algorithm that simulated many solutions to the overall trial order problem. The algorithm respected the constraints described above, while also introducing randomness in order to limit

Figure 1: The presentation sequence of a trial. A fixation cross appears for 2 seconds. A sample set of 6 colors to be remembered is then presented for 4 seconds. A diffuse colored mask intended to cancel out sensory correlation with the subsequent test array is then presented for 2 seconds followed by a set of 6 color patches rotated by 90° (and completely scrambled with no correlations in positions before vs. after masking). A final colour mask is then presented for 1 second in advance of the unconstrained response phase. Subjects would click on a value for their estimate of the number of changes, marked in red, and the correct value marked by blue.

trial order confounds between participants. A unique solution was found for each participant. Solutions were defined as trial orders where the total number of trials was nearly equal for each of the seven possible correct answers (so as many trials have an answer of "4" as have an answer of "2"), and where the number of trials was also nearly equal for each of the seven possible one-back changes in correct answer. The shape depicted in Figure 2 was the algorithm's consistent solution to this problem, with only very minor differences and asymmetries between participants.

Experiment Results

The average error across all trials and all subjects was 1.5 units (number of color changes, out of 6). A mixed effects model with subject as a grouping factor showed no significant improvement in change detection over the course of the experiment ($t = 1.6$, $p = 0.11$). In accordance with our goal of exploring sensitivity to swings in cognitive energy and activation over time, we also checked for a lag-one correlation in responding over the course of the experiment. Lag-one in this case is being measured as the correlation between responses on trial t and responding on trial t-1, i.e. the correlation between the list of responses and itself shifted one trial sooner. When this correlation is positive, it suggests that answers were given in long "runs" where a high answer would be followed by more high answers and similarly for low. When the lag-one correlation is negative, it suggests a degree of "ping-ponging" back and forth between high and low (number of changes) successive answers more often than would be expected by chance. A lag-one of zero suggests no particular persistent carryover effects of responding from one trial to the next.

For this analysis, it was necessary to control for any lag-one correlation that may have been inherent to the trial order itself. Figure 2 shows how, despite equal distribution of changes and differences in changes in colors across trials, patterns exist between these two variables. To control for such patterns during the analysis, we looked for lag-one correlation only in those trials we knew had a symmetric pattern of changes compared to previous trials: ones where the answer was exactly three changes ("Correct Answer = 3" set of bars in Figure 2). Within this restricted data set, we found a positive correlation ($\beta = 0.11, t = 2.74, p < 0.01$) in the responses between trials, indicating a minor preference for repeated "runs" of responding that was not related to any experimental design.

Timing differences in the second color presentation phase of the task did not correspond to significant differences in performance. Differences were expected, but any homeostatic effects may simply be too rapid (or too slow / occurring by memory only during the answering phase) to be distinguished by the difference between 2.5 and 4 seconds of presentation.

Model

We tested two models against our behavioral data: one biologically inspired model (in line with with the BDNF principles discussed by Glaser and Joublin, 2011) capturing cellular homeostatic principles, and the same model but with the cellular homeostatic term removed. Each was simulated in Matlab using the exact order and content of trials and color values seen by each of the 30 participants in turn. These models provide continuous real number outputs between 0 to 6, but were forced to choose exact whole number answers for number of changes, as the humans were. Especially for the model that accounted for the C.H. model, the effect of feedback could have contributed differently to the next trial's behavior if whole number answers were not required, so to ensure the most human-like between-trial patterns, model answers were also rounded.

Since our goal was to capture the levels of errors and inaccuracy in human behavior, and the difficulties of the task being between-trial consistency, we fed the signal for the number of changes on a trial to the models directly, so that the target measure of the fit was focused on the specific pattern of errors made by each human on each of their trials. Each model was fit to human responses using the method described below, which tested enough detail to capture these homeostasis straining effects over time, as well as patterns of variance in errors.

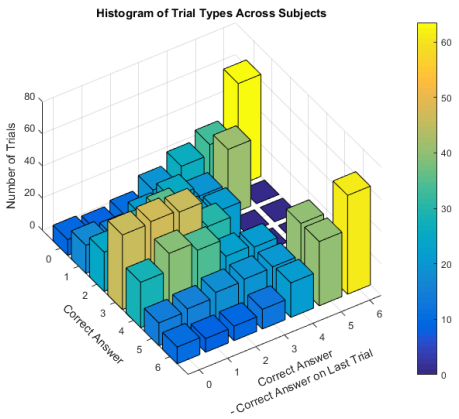

Figure 2: Histogram of different trial types in the experiment. Every trial had a correct answer, and all but the first trial had a one-back difference between its correct answer and the correct answer on the previous trial. Some of these combinations were impossible (see text). The shape seen here equally distributes correct answers overall, and also equally distributes changes between trials overall, while avoiding the impossible trial types. Bars are not perfectly symmetrical due to partial randomization between participants to avoid trial order effects. Some small asymmetries between bars are visible due to the necessity of using a Monte Carlo algorithm for this task.

Cellular Homeostasis Model

The primary model of interest displayed homeostasis as a result of cellular energy resisting extreme response rates by becoming depleted after heavy use or energized after low use. The neuron's energy reserves were its only way of tracking information across trials. Its output on a trial is given by:

$O_t = as_tE_t + b + \varepsilon n$

where s_t is the stimulus on the trial (the number of changed colors), ε is normally distributed noise, a, b, and n are freely fitted coefficients, and E is cellular energy, calculated per trial as:

$E_t = E_{t-1} + \frac{(O_{t-1}-3)c}{\tau} - \frac{(E_{t-1}-1)\frac{c}{3}}{\tau}$

The fourth free parameter of the model is c in the energy equation above. τ was not parameterized and was always set to a constant value of 10. The $O_{t-1}-3$ represents the fact that three was the most central response out of options 0-6, so any values below this were considered "low" answers that helped relatively gain cellular energy, and any values above this were "high" answers that relatively depleted energy reserves. The third term represents relaxation of the system to neutral energy over time, but 1/3 slower than the rate of energy change from responding (Toyoizumi, Kaneko, Stryker, & Miller, 2014). Setting $E_{t-1}-1$ causes energy to relax toward a neutral value of 1, where it would have no effect on the cell's output. All model answers were rounded to the nearest integer from 0-6 - the only valid answers in this task - on a per-trial basis. This was also true of the mathematical model variant below.

The energy term mimics the predictions of the BDNF chemical cell model, with as much abstraction as necessary to allow for easy application to typical cognitive behavioral data. As in the BDNF model, a decision neuron firing at a relatively low rate (0-2 changes in this task) builds up energy and fires faster than normal for a brief period in response, while a neuron firing relatively energetically (4-6 changes in this task) will progressively deplete energy and fire more slowly for a time in response.

Simple Mathematical Model

The second model had no biological motivation, but serves as a baseline comparison with the C.H. model. It replicates the C.H. model in structure, but with the energy term removed. This model had no means by which to account for previous trials, as the model had no form of memory/hysteresis. It did, however, still have all of the information needed to perform perfectly at the task according to task instructions (rather than fit to human answers). A perfect score could be achieved with parameter values $a = 1$, $b = 0$, and $n = 0$. Thus, lower performance by this model at fitting human data would be due to a lesser ability to capture human sources of error and trial to trial effects (theoretically irrelevant and distracting from optimal performance) only. The model is given by:

$O_t = as_t + b + \varepsilon n$

Fitting Method

We created three dimensional histograms of responses for humans and models to fit and compare results. Every trial across subjects was sorted into histogram bins according to the correct answer on that trial (0 to 6), the difference between the correct answer on that and the correct answer on the previous trial (-6 to 6), and the participant's (real or simulated) response (0 to 6). This produced 7 x 13 x 7 set of possibilities (637 histogram bins). The first two dimensions represent objective trial types (ones built into the design and based on actual stimuli) and since these were not perfectly evenly distributed due to mathematical constraints (see Figure 2), we weighted the importance of cells from the combinations that had more more data points, using weighted least squares:

$Fit = \sqrt{\frac{1}{637}\sum_{s,d,r}(\sum_{s,d}(N_{H_{s,d}})(H_{s,d,r,model} - H_{s,d,r,human})^2)}$

where H is the histogram, and s, d, r are stimuli at time of response (t); difference in stimuli(t) – stimuli(t-1); and response, respectively. This method captures information about trial to trial effects, main effects, interactions, general task accuracy, and patterns of variance, all in one measure.

This histogram method was chosen for the objective function to avoid the problem that fitting averaged descriptives like accuracy or standard deviation of responses, which could lead to degenerate model patterns: a mean might be fitted by 100% of model answers at exactly the mean without realistic variance, for example. Fitting the entire histogram of all relevant measures allowed for the model with the richest detailed pattern of fits across every measure.

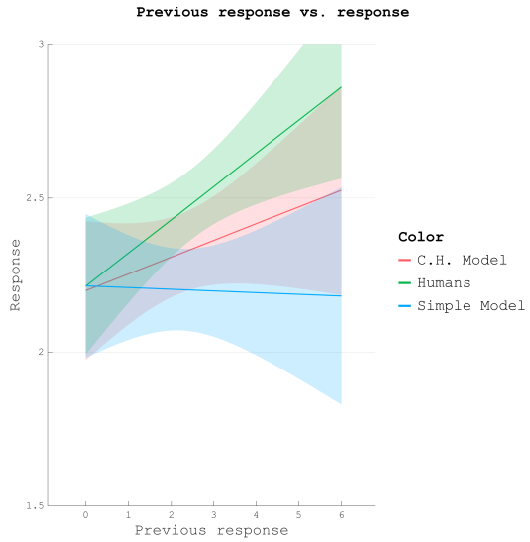

Figure 3: Change responses versus change responses on the previous trial, where the current trial had 3 changes. C.H. is the Cellular Homeostatic Model.

The models have differing numbers of free parameters (4 vs. 3), yet due to the dynamical nature of the cellular energy model, its maximum likelihood cannot be easily calculated, and simulations take non-trivial time to perform. Ultimately, the main concern of an overly complex model is failure to generalize, so instead of scoring parsimony, we ruled out over-fitting directly using cross-validation. For each model, we split the subject pool in half, and separately fit each half. We then recorded the fits for each half using only the best parameters found from fitting the other half. All results reported below are exclusively these generalization results, removing our concerns about hidden differences in generalization ability between the three and four parameter models.

We fit a 7 value range for each parameter in a grid covering all of reasonable parameter space for the task, separately for each half of participants as above. We then focused more closely near each best fit at higher precision until precision increases stopped yielding better fits.

Model fitting results

The best fitting parameters for the cellular homeostasis model were $a = -0.4$, $b = 3.5$, $c = 0.92$, and $n = 1.4$. The best fitting parameters for the simple mathematical model were $a = -0.45$, $b = 3.5$, and $n = 1.5$.

The average cross-validation weighted least squares error for the cellular homeostasis model was 8.335, while the average cross-validation weighted least squares error for the simple mathematical model was 9.436. Since these values already account for the greater potential for over-fitting with four versus three parameters, they can be compared at face value: the energy term meaningfully accounts for human behaviors above and beyond slope, intercept, and noise terms.

Although the magnitude of the effect is somewhat modest, it is noteworthy to point out again that the simple mathematical model was able to achieve 100% objective accuracy at the task as per the task instructions by simply fitting parameters $a = 1$, $b = 0$, and $n = 0$, a combination that was within the tested range of reasonable parameters during fitting. Thus, the lower performance of the simpler model is purely a result of more poorly fitted patterns of human error, possibly error in response to patterns of trials that threw off homeostatic neutrality, since the better fitting energy term in the homeostasis model varies by activity on previous trials.

Ultimately, the exact cause of the better fits of the cellular homeostasis model are unclear. Analysis of the full three-dimensional fitting histogram suggested noticeable differences between the two models and between models and human data, but these differences were too diffuse and opaque to easily interpret.

Lag-one correlations on trials with three color changes also fit human data better in the homeostasis model than in the simple mathematical model. Where $\beta_{human} = 0.11$, $\beta_{cellular\,model} = 0.05$ and $\beta_{simple\,model} = 0$ (see Figure 3). These correlations highlight the lag-one effects in particular, but lag-one effects are also built into the 3-dimensional histograms used for the main fitting results.

Discussion

Behavioral stability is often approached from the perspectives of neural population dynamics or higher-level verbal or executive control theories. Stability is also attainable, however, through more microscopic means, at an intra-cellular or synaptic level. This source of homeostasis in cognition and behavior is, by itself, simple. Activity is most likely stabilized around a static resting level, at least within the timescales afforded by a particular task. This does not necessarily match the flexibility or possible sophistication of higher level stabilizing mechanisms.

Cellular homeostasis is, however, an appreciable effect, especially when studied in a task that eliminates distracting forces and pushes the boundaries of a system's homeostasis. Even in less specialized situations, however, cellular effects are likely continuing to function and can contribute toward an understanding of behavior. This form of homeostasis may generally be playing a silent and under-appreciated role in a wide variety of cognitive activities, providing a small but important level of baseline stability that can act as a foundation for more targeted systems like learning mechanisms to explore more freely without risk of losing control of a system.

Our findings require significant further investigation to establish an exact pattern of behavior that is being captured by the cellular energy term of our model, and followup experiments are necessary to confirm those mechanisms once identified. In the meantime, we suggest cognitive modelers more often consider including cellular energy terms in neural models of not only cellular-level effects, but behavioral effects as well. All cognitive processes involve neurons, so even modest effects of such a cellular system may be of great impor-

tance collectively, for a range of effects at different levels of complexity and abstraction.

Acknowledgments

We would like to acknowledge the helpful contributions made by all members of the Cognitive Science Laboratory at Simon Fraser University. Gavin Jenkins was supported by an NSERC Discovery Accelerator Supplement. Paul Tupper was supported by an NSERC Discovery Grant and a Tier 2 Canada Research Chair.

References

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *2*(1), 32–48. doi:10.1371/journal.ppat.0020109

Glaser, C. & Joublin, F. (2011). Firing Rate Homeostasis for Dynamic Neural Field Formation. *IEEE Transactions on Autonomous Mental Development*, *3*(4), 285–299. doi:10.1109/TAMD.2011.2138705

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A Dynamic Neural Field Model of Visual Working Memory and Change Detection. *Psychological Science*, *20*(5), 568–577. doi:10.1111/j.1467-9280.2009.02329.x

Lehky, S. R. & Sejnowski, T. J. (1999). Seeing white: Qualia in the context of decoding population codes. *Neural computation*, *11*(6), 1261–1280. doi:10.1162/089976699300016232

Oja, E. (1982). A Simplified Neuron Model as a Principal Component Analyzer. *Journal of Mathematical Biology*, *1*, 267–273.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. *The Behavioral and brain sciences*, *24*(1), 1–34, discussion 34–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11515285

Toyoizumi, T., Kaneko, M., Stryker, M. P., & Miller, K. D. (2014). Modeling the dynamical interaction of Hebbian and homeostatic plasticity. *Neuron*, *84*(2), 1–40. doi:10.1016/j.neuron.2014.09.036

Turrigiano, G. (2011). Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annual review of neuroscience*, *34*, 89–103. doi:10.1146/annurev-neuro-060909-153238