A Bayesian decomposition of BAC firing as a mechanism for apical amplification in neocortical pyramidal neurons

Jim W. Kay[1], W. A. Phillips[2], Jaan Aru[3], Bruce P. Graham[2] Matthew E. Larkum[4],

1 School of Mathematics and Statistics, University of Glasgow, Glasgow, UK,
2 Faculty of Natural Sciences, University of Stirling, Stirling, UK,
3 Department of Biology, Humboldt University of Berlin, Berlin, Germany,
4 Neurocure Cluster of Excellence, Department of Biology, Humboldt University of Berlin, Berlin, Germany.

Abstract

Pyramidal cells in layer 5 of the neocortex have two distinct integration sites. These cells integrate inputs to basal dendrites in the soma while integrating inputs to the tuft in a site at the top of the apical trunk. The two sites communicate by action potentials that backpropagate to the apical site and by backpropagation-activated calcium spikes (BAC firing) that travel from the apical to the somatic site. Six key messages arise from the probabilistic information-theoretic analyses of BAC firing presented here. First, pyramidal neurons with BAC firing turn the calculation of the probability that a feature is present given the basal data into a simple Bayesian calculation of the probability that the feature is present given the basal data and the context. Second, activation of dendritic calcium spikes amplifies the cell's response to basal depolarization that occurs at about the same time as apical depolarization. Third, these analyses show rigorously how this apical amplification depends upon communication between the sites. Fourth, we use data on action potentials from a very detailed multi-compartmental biophysical model to study our Bayesian formulation in a more realistic setting, and demonstrate that it describes the data well. Fifth, this form of BAC firing meets criteria for distinguishing modulatory from driving interactions that have been specified using recent definitions of multivariate mutual information. Sixth, our Bayesian decomposition can be extended to cases where, instead of being purely driving or purely amplifying, apical and basal inputs can be partly driving and partly amplifying to various extents. These conclusions imply that an advance beyond the assumption of a single site of integration within pyramidal cells is needed. The success of neocortex may arise from the evolution of cells with the kind of context-sensitive selective amplification demonstrated here.

Author summary

The cerebral cortex has a key role in conscious perception, thought, and action, and is predominantly composed of a particular kind of neuron: the pyramidal cells. The distinct shape of the pyramidal neuron with a long dendritic shaft separating two regions of profuse dendrites allows them to integrate inputs to the two regions separately and combine the results non-linearly to produce output. Here we show how inputs to this more distant site strengthen the cell's output when it is relevant to the current task and environment. By showing that such neurons have capabilities that transcend those of neurons with the single site of integration assumed by many

neuroscientists, this 'splitting of the neuronal atom' offers a radically new viewpoint from which to understand the evolution of the cortex and some of its many pathologies. This also suggests that approaches to artificial intelligence using neural networks might come closer to something analogous to real intelligence, if, instead of basing them on processing elements with a single site of integration, they were based on elements with two sites, as in cortex.

Introduction

To cope with an ever-changing environment animals take various forms of prior knowledge into account, including knowledge of how the meaning of ascending perceptual signals and the likely consequences of descending motor signals depend on context. A currently prominent view suggests that this context-sensitive integration of prior knowledge and sensory data is processed in a Bayesian fashion, but it is not known how this is implemented at the neural level. Here, we show how a non-standard application of Bayes' theorem can be used to explain how the neocortex combines current data with prior knowledge in a way that depends on sub-cellular processes in neocortical pyramidal neurons. In brief, we show how the transmission of relevant information can be amplified, or boosted, by the sub-cellular process known as back-propagation activated calcium spike firing (BAC firing), which uses information from widely diverse sources to adapt the information transmitted by the cell to the current context. This uses the contextual information in a special super-additive nonlinear way that transcends what can be done by point neurons such as those assumed by many neuroscientists and used in most of the current generation of machine learning algorithms.

In layer 5, and perhaps in other layers also, these neurons have two functionally distinct sites of integration; one in the soma and one near the top of their apical trunk (e.g. [1]). Inputs to these two sites arise from very different sources. In particular, basal dendrites, which receive a narrow range of ascending inputs from first-order thalamic nuclei and lower cortical regions, feed directly into the soma. In contrast to that, the dendrites of the apical tuft, which receive diverse inputs from higher cortical regions, higher-order thalamic nuclei, and various sub-cortical nuclei such as the amygdala, are far from and have less direct influence upon the spike generating zone [1, 2, 3]. This clear anatomical segregation suggests that apical dendrites could use information from diverse sources to amplify transmission of ascending information that is relevant to the task, resolve various forms of ambiguity, or contradict predictions, [1, 4, 5, 6]. There is also evidence that the effects of apical input depend upon the adrenergic and cholinergic systems that regulate waking state (e.g. [7]), that they have a causal role in guiding overt perceptual detection [8], and that they may provide a common pathway for the effects of general anaesthetics, [9, 10].

The mechanism of BAC firing relies on the non-linear signaling along the apical axis of the neuron. If basal and apical dendrites are depolarized at around the same time, an action potential generated in the basal compartment lowers the threshold for Ca^{2+} spikes that can be triggered at the distal apical integration zone, and which travel rapidly to the soma, where it greatly increases the probability of more action potential(AP)s being generated within the next 10-20 ms or so (e.g. [1, 11, 12, 13, 14, 15]). Thus, unlike two neurons connected via synapses, the two dendritic compartments are linked via a privileged connection, the apical trunk, capable of non-linear, bi-directional signaling via active dendritic currents and spikes. Furthermore, signaling along the apical axis has different temporal dynamics and is regulated by a host of ion channels, such as HCN channels and those of the cholinergic and adrenergic systems.

BAC firing is less likely to be triggered by apical or somatic depolarization alone, and its amplifying effects are far greater than can be accounted for by any simple additive process (e.g. [1, 14, 15, 16]). There is a clear anatomical asymmetry between apical and basal inputs in that, whereas

basal dendrites affect the soma directly, apical inputs do so only via the apical trunk. Interestingly, recent evidence from pyramidal neurons in different brain areas [14], different stages of development [17] and different species [18] suggests that the apical dendrite can be very electrically remote from the cell body. This electrical remoteness could be a major variable determining the likelihood of apical amplification and the non-linearity of interactions between the two dendritic compartments. It is this distinctive aspect of apical function and BAC firing that is central to the perspective being further developed here.

BAC firing has previously been studied using computational models [13, 15], but has not till now been explicitly related to Bayesian inference. We therefore do so here, but not in the form in which it is most often used in neuroscience. Bayes' theorem shows how inferences can be decomposed into terms that depend on current data and prior knowledge, but the way in which it is most commonly applied to inference in neural systems, e.g. in [19], does not provide an appropriate formulation of BAC firing. This is because using the standard formulation would require the assumption that output from the somatic site depends primarily on the apical site, whereas the logic of contextual amplification and the evidence for BAC firing shows that it is the other way round, with the activation of apical site depending primarily upon output from the somatic site. So here we use an alternative application of Bayes' theorem [20], in which it is output from the somatic site that is primary, with output from the apical site depending upon that.

It has previously been shown that, given functionally distinct receptive and contextual field inputs, learning rules can change synaptic strengths so as to adapt them to the latent statistical structure in their inputs [20, 21, 22, 23]. It has been suggested that BAC firing can provide the activation functions used in those theories [5], but till now that has not been shown explicitly. Furthermore, though the pyramidal cell as a whole can be analyzed as an activation function with two inputs and one output, we show here that it can be more explicitly related to BAC firing by describing each of the somatic and apical integration zones as operating as a three-term activation function with one input from outside the cell and one input from inside the cell. This more realistic representation may not be as complicated as it sounds, however, because we show that the somatic zone, the apical zone, and the cell as a whole can all be described using the same general form of activation function.

To further simplify the analyses we build on the fact that the probabilities being estimated are of a binary event, i.e. whether or not an AP is generated. This enables us to use the ratio of these two probabilities, known as the odds, and the log of that ratio, known as the log odds. Though less intuitive, the log odds provides a simple description of a wide range of physiological and psychophysical phenomena [24]. First, we consider the prior odds in favor of an AP before reception of the current input. Second, we have probability density functions for the input given that an AP has or has not been generated, which are known as the likelihoods. They give the likelihoods of the occurrence and non-occurrence, respectively, of an AP given that input, and their ratio is known as the Bayes factor or likelihood ratio. A value of greater than 1 for this ratio favors the occurrence of an AP, while a value less than 1 favors the non-occurrence. By applying Bayes' theorem the posterior odds in favor of an AP can then be expressed as the product of the prior odds and the Bayes factor. It is simpler to take logarithms to obtain an additive decomposition of the posterior log odds as a sum of the prior log odds and the logarithm of the Bayes factor, which is known as the weight of evidence [25]. This term weighs the evidence provided by the input in favor of an AP. Thus, a positive value for the weight of evidence favors an action potential, while a negative value favors no action potential.

Such expressions of the posterior odds are found in [26, 27, 28], where their use is attributed to A. M. Turing in the 1940s; see also [29]. The weight of evidence increases as the Bayes factor increases. The additive decomposition of the posterior log odds can be easily updated as new inputs are obtained. In our application, the basal input is taken into account before the apical

input, so the posterior log odds given both these inputs can be written as a sum of three terms: the prior log odds, the weight of evidence in favor of an initiating action potential provided by the basal input, and the additional weight of evidence in favor of a second AP provided by the apical input given the basal input.

The analyses presented in the main body of the paper involve a great deal of mathematics, so, to make them clear, the key messages that we draw from these analyses are summarized here.

1. **By using an alternative version of Bayes theorem, the probability of a second action potential given both basal and apical input is shown to be decomposable into the sum of three terms.** These are the prior log odds in favor of an AP, the weight of evidence in favor of an initi-aiting AP provided by the basal input alone, and the weight of evidence in favor of a second AP provided by the apical input given the basal input. This provides a novel, simple, and general Bayesian decomposition of the effects of apical input in the BAC firing regime. Consider pyramidal neurons with two integration zones, apical 'a' and basal 'b', and with binary output z which codes for the presence ($z = 1$) or absence ($z = 0$) of a particular feature in the input to the cell. Then BAC firing enables the cell to compute the output probability, $P(z|b,a)$, given both basal and apical inputs, i.e.: $P(z|b,a) = P(z|b)p(a|z,b)/p(a|b)$. In a nutshell: Pyramidal neurons with BAC firing turn the calculation of P(feature present | basal data) into a simple context-sensitive Bayesian calculation of P(feature present | basal data, context).

2. **These analyses show how BAC firing can amplify the cell's response to basal depolarization that occurs at about the same time as apical depolarization.** This shows how BAC firing can be related in detail to the extensive physiological and psychophysical evidence for contextual modulation and disambiguation. Unlike coincidence detection, this view of apical function depends upon the marked asymmetry between the effects of apical and basal inputs that is clear in the physiological and modeling data [15]. Thus, the analyses of apical amplification presented here are consistent with previous demonstrations that contextual modulation is not a purely multiplicative interaction,which is symmetrical, but is a very special kind of asymmetrical supralinear interaction [16]. They also show that BAC firing can approximate the activation function that is central to a long-standing theory of cortical computation based on context-sensitive selective amplification [5, 21, 20].

3. **These analyses show how amplification depends upon communication between apical and basal sites of integration.** Both the somatic and apical sites of integration are shown to have two functionally distinct inputs, one from outside the cell, and one from inside the cell, and it is shown that the activation functions relating the two inputs to the outputs from each site can have the same general form. This provides a clear account of the functional consequences of communication between the two sites, and thus provides a basis for an adequate understanding of regulation of that communication by cholinergic and other modulatory systems.

4. **Application of our Bayesian formulation to binarised AP data from a detailed multi-compartment model [15], coupled with Bayesian modeling, shows that our general model provides a good description of these data, based on posterior predictive assessment.**

5. **Information theoretic decomposition of this form of BAC firing is shown to meet previously specified criteria for distinguishing modulatory from driving interactions [30].** Furthermore, application of this information decomposition to categorised AP data from [15] shows that its qualitative properties are much as expected of an amplifying interaction, and it also provides an estimate of the synergy in this system.

6. **Our Bayesian decomposition can easily be extended to include cases where, instead of being either purely amplifying or purely driving, apical and basal inputs can be partly driving and partly amplifying to various extents.** This enables application of the analyses to neurophysiological findings that provide evidence for driving effects of apical input. This extension of the Bayesian decomposition may also be useful in characterizing differences in apical function across layers, regions, species, development, and state of arousal.

The main body of the paper is organized as follows. We begin by presenting a simple idealized sketch of inputs to and outputs from the two integration zones and of the interaction between them via the apical trunk. Consideration is then given to the case in which BAC firing is initiated after a backpropagating somatic action potential has been received at the apical site, and our alternative Bayesian interpretation is defined, along with the Bayesian decompositions of the posterior log odds for both the somatic and the apical integration sites. We then provide expressions for the posterior probability of (i) an initiating AP given only basal input and (ii) a second AP given combined basal and apical input. The weight of evidence terms used in these expressions are defined in the Methods section, and they are based on a class of activation functions that is similar to those used in previous theoretical studies of contextual guidance in neocortex. These analyses are then applied to binarised data from a detailed multi-compartment model of apical function developed in [15]. We consider two models to describe the data. The first is a specially crafted threshold model, analogous to the composite model fitted to the frequency data in [15], while the second makes use of our general activation function. The posterior probability of a second AP is computed as a function of basal and apical inputs using both models and the results are compared. A categorised version of the data from [15] is then considered and analysed by means of classical information measures as well as partial information decompositions in order to clarify the effects of apical depolarization on AP generation, and of the synergistic component in particular. The technical details appear in the Methods section. Finally, results are presented for the Bayesian decompositions of posterior log odds for cases in which the basal and apical inputs may be partly driving and partly amplifying to various extents, and they are contrasted with the decomposition when interaction is additive, with the technical definitions appearing in the Methods section. This extension is motivated by evidence that that under some circumstances apical depolarization can to some extent drive APs even when not triggered by backpropagating action potentials [14, 15, 17, 8, 31].

Results

Two sites of integration that interact via backpropagation and BAC firing in layer 5 pyramidal cells

An idealization of integration within and communication between the somatic and apical sites is shown in Fig 1. This is designed to relate known physiological processes to the probabilistic analyses presented in the following sections. Each integration site receives input from two sources, one that is extracellular and one that is intracellular. The somatic site receives extracellular input from basal/perisomatic synapses, and we refer to it as the 'basal input', b. It also receives internal input, c, from the apical site. The apical site receives extracellular input from apical synapses in layer 1, which we refer to as the 'apical input', a. It also receives internal input, z, in the form of the action potential that is backpropagated from the somatic site.
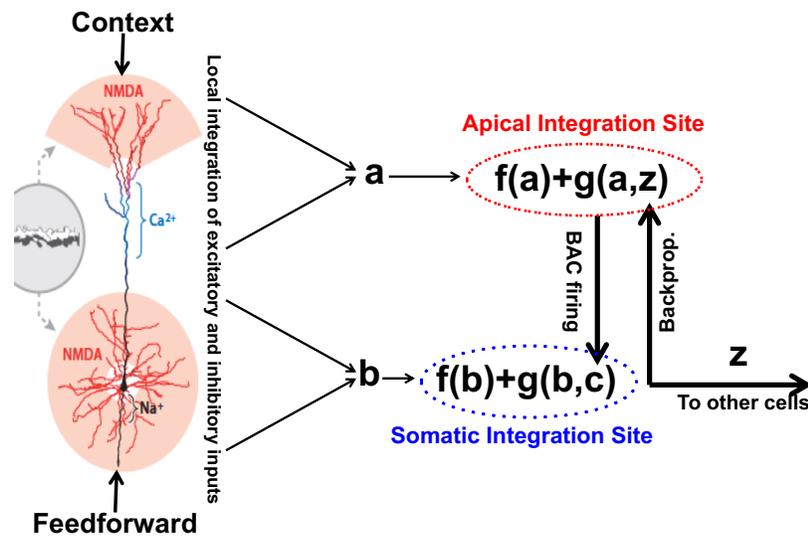
5

Figure 1: **Integration within and between somatic and apical sites.**
An idealization of integration within and communication between somatic and apical sites in the case where an initiating bAP has been generated and BAC firing is triggered. z is a binary variable that equals 1 if there has been a somatic action potential and zero if not. A backpropagated action potential amplifies transmission of information, via $c = f(a) + g(a, z)$, about the apical input, a, to the somatic integration site, where it is used to amplify transmission of information about the basal input, b, to other cells. (The layer 5 pyramidal cell on the left is from Fig 4G of [32].)

Basal dendrites typically receive their input from a few narrowly specified feedforward sources that in sensory and perceptual regions ascend the hierarchy of neocortical abstraction. Apical dendrites typically receive input from diverse sources that include feedback from higher neocortical regions, higher-order thalamus, the amygdala, and the adrenergic and cholinergic systems. This diverse set of sources provides contextual information that guides processing and learning of the feedforward data. For simplicity we refer to this set of diverse inputs collectively as 'context'. It amplifies response to feedforward signals that are relevant given that particular context. We refer to the post-synaptic locations that receive feedforward information as 'basal', however they may also include perisomatic locations, as shown in Fig 1. Positive values of a and b are taken to be analogous to net depolarization at that site, and negative values to hyperpolarization. Computation of the net basal and apical inputs are shown in Fig 1 as occurring outside their respective sites of integration, but we assume that it occurs as part of the integration within sites.

Empirical data indicates that BAC firing most frequently adds one more AP to the initiating AP; sometimes two; and rarely more. We assume that our focus on the 10 ms time-scale is justified because two or more spikes within about 10 ms is a faster and more energy efficient signal than mean spike rate over 100 ms, and because both data and models indicate that apical depolarization has more effect on bursting than on mean spike rate (e.g. [15, 33, 14]). To extend our analysis to adequately cover bursts of different lengths, inter-burst intervals, and fast regular spiking we would have to include the effects of negative feedback from inhibitory interneurons, including those specific to the tuft. The refractory nature of AP generation and synaptic facilitation/accomodation would also become relevant if we were to present the analysis as being concerned with a sequence of APs. The present analysis makes no assumptions about those things but simply considers the dependence of AP probability on a and b given that a back-propagating AP has occurred. There

are subtle simplifications in the idealization presented, e.g. a and b are considered to be stable on the brief time-course of the BAC firing effect.

A Bayesian interpretation of intra-site computation

We represent the net apical input by the continuous random variable, A, which is a weighted and summed combination of nonlinear functions of inputs from various sources in the apical dendrites. The observed value of A is the a used in Fig 1. We represent the net basal input by the continuous random variable, B, which is a weighted and summed combination of nonlinear functions of inputs from various sources in the basal dendrites. The observed value of B is the b used in Fig 1. The somatic output is given by the binary random variable, Z, which has observed value, z, which is used in Fig 1.

We consider first the scenario in which the BAC firing in the apical site is triggered by the arrival of an initiating backpropagated action potential, as illustrated in Fig 1.

A common Bayesian formulation

Bayes' theorem has been commonly used in hierarchical Bayesian inference, as in Lee and Mumford [19], who considered three classes of variable: observed, hidden and contextual. When applied in this study, the observed variable corresponds to the basal input, B, the hidden variable to the somatic output, Z, and the contextual variable to the apical input, A. In Lee and Mumford's formulation, using our notation, the conditional probability that $Z = z$ given the observed values of both the apical and basal input, $P(Z = z|a, b)$, may be written as

$$P(z|a, b) = \frac{P(z|a)\, p(b|z, a)}{p(b|a)},$$

where a generic 'p' denotes a probability density function (p.d.f.) and a generic 'P' denotes a probability.

The prior distribution of Z given the apical input a is the term $P(z|a)$, which provides a model for a-priori predictions of the output Z given the apical input. The term $p(b|z, a)$ is a generative model for the basal input b given that $Z = z$ and the observed value of the apical input a. The term $p(b|a)$ is essentially a normalizing constant.

This form of decomposition is not appropriate in our formulation, however, since it does not represent the order of the two types of communication between the somatic and apical sites; the initiating bAP emitted from the soma comes first and it consequently triggers the BAC firing by which the apical activation influences the subsequent activation in the somatic site, as illustrated in Fig 1. Hence we change perspective and use Bayes' theorem in a different manner.

An alternative Bayesian formulation

There are two phases involved when thinking of the computation from a Bayesian viewpoint. The output of the unit, Z, is a binary variable with unknown probability of taking the value $z = 1$. The aim is to determine the conditional probability that $Z = 1$ given (i) the basal input alone and (ii) given both the basal and apical input.

In Phase 1, during which no initiating action potential is generated, we have a prior distribution on Z and a generative model for the basal input, b, given that $Z = z$. These are combined using Bayes' theorem to produce the posterior probability that $Z = z$ given the basal input, b:

$$P(z|b) = \frac{P(z) \times p(b|z)}{p(b)}. \tag{1}$$

In Phase 2 of the computation, during which an initiating action potential is generated and is followed by BAC firing, this posterior probability that $Z = z$ is updated by taking into account the observed apical input, a, and using Bayes' theorem in the form

$$P(z|b, a) = \frac{P(z|b) \times p(a|z, b)}{p(a|b)}, \tag{2}$$

where $p(a|z, b)$ denotes a generative model for the apical input given the observed basal input, b, and z. In our approach, the prior is $P(z|b)$, which provides a model for a-priori predictions of Z given the basal input, and $p(a|z, b)$ is a generative model for the apical input given z and the value of the observed basal input, b. This alternative Bayesian perspective provides the correct sequence of communication between the apical and somatic sites, with the initiating backpropagating action potential (bAP) happening first to be followed by BAC firing.

While this explains the manner in which the computation could be performed using a Bayesian approach, it is not necessary to define the generative models in our work since the conditional distribution of Z given the basal and apical inputs is modeled directly without making any distributional assumptions regarding the generative models for the basal and apical inputs.

Log Odds and Weight of Evidence

In our Bayesian formulation, we change notation and consider the event S to be that an AP is generated, with $\bar{S}$ denoting the complementary event that no AP is generated. The event S is equivalent to $Z = 1$, and $\bar{S}$ is equivalent to $Z = 0$. Since S is a binary event it is simpler and customary to work with the ratio: the prior probability of the event S divided by the prior probability of the event $\bar{S}$. This is termed the prior odds in favor of the event S (before any data are available). It is of interest to determine the posterior odds in favor of the event S, given data.

In the first phase, we consider the ratio of the posterior probabilities of the events, S and $\bar{S}$, given only the basal input, which can be written using Eq (1) as

$$\frac{P(S|b)}{P(\bar{S}|b)} = \frac{P(S)}{P(\bar{S})} \times \frac{p(b|S)}{p(b|\bar{S})} \tag{3}$$

This expresses the posterior odds in favor of an initiating AP, given the basal input, as a product of the prior odds $P(S)/P(\bar{S})$ and $p(b|S)/p(b|\bar{S})$, which is the Bayes factor in favor of the event S provided by the basal input, b. This Bayes factor is a ratio of the likelihood of an initiating AP, given the basal input, and the likelihood of no AP, given the basal input. A value greater than 1 for the Bayes factor would favor the event S, as opposed to the event $\bar{S}$.

Similarly, we can use Eq (2) to update the posterior odds in Eq (3) once the apical input is observed:

$$\frac{P(S|b, a)}{P(\bar{S}|b, a)} = \frac{P(S|b)}{P(\bar{S}|b)} \times \frac{p(a|b, S)}{p(a|b, \bar{S})}. \tag{4}$$

This formula expresses the posterior odds of a second AP, given both the basal and apical inputs, as a product of the updated prior odds of an initiating AP given only the basal input and $p(a|b, S)/p(a|b, \bar{S})$, the Bayes factor in favor of the event S provided by the the apical input, a, given the basal input, b.

It is then very natural to convert Eq (4) to an additive scale, by taking logarithms, to provide a Bayesian decomposition of the posterior log odds. The logarithm of a Bayes factor is termed the 'weight of evidence' [25]. So, for example, in Eq (3) the logarithm of the Bayes factor is the weight of evidence in favor of the propagation of an initiating AP provided by the basal input, and denoted by $W(S : b)$. A weight of evidence term could be positive or negative; a positive

value would favor the occurrence of an initiating AP, while a negative value would favor the non-occurrence of an initiating AP.

Taking logarithms in Eq (3) gives the following decomposition

$$L(S|b) = L(S) + W(S:b), \tag{5}$$

where $L(S)$ denotes the prior log odds in favor of S and $L(S|b)$ denotes the posterior log odds in favor of S given only the basal input. Applying logarithms in Eq (4), we have

$$L(S|b,a) = L(S|b) + W(S:a\,|\,b) = L(S) + W(S:b) + W(S:a\,|\,b), \tag{6}$$

using Eq (5). This provides a simple Bayesian decomposition of the posterior log odds in favor of an AP, given both the basal and apical inputs, as a sum of three terms: the prior log odds in favor of an initiating AP, the weight of evidence in favor of an initiating AP provided by the basal input alone, and the weight of evidence in favor of a second AP provided by the apical input, given the basal input. This computation relates to the somatic integration site, but the term $W(S:a\,|\,b)$ would not be available without the two-way communication between the sites by which the increased apical activation is transmitted to the somatic site, thus increasing the odds of the propagation of a second AP within a time interval of around 10 ms.
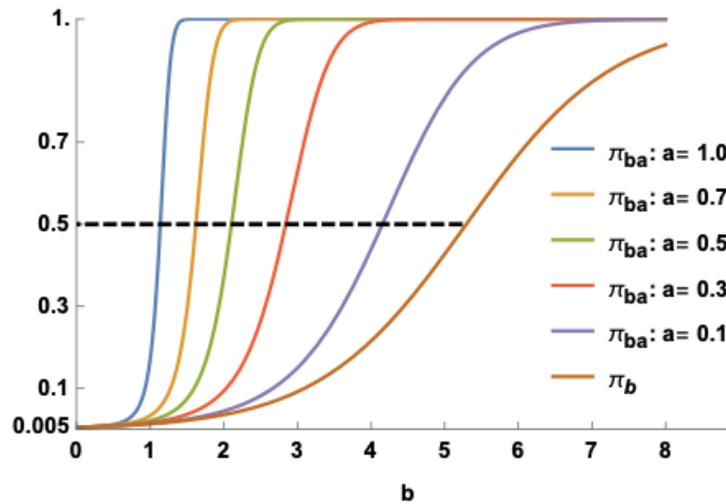

Figure 2: **Comparison of posterior probabilities.**
Posterior probabilities, π_{ba}, π_b, are plotted against positive values of the basal input, b, using five different values of the apical input, a. The dashed line indicates points on the curves at which the posterior probability is 0.5. The typical baseline firing probability, $P(S)$, is assumed to be 0.005.

Comparison of posterior probabilities of an action potential in the somatic site

The posterior probabilities of interest are (i) $P(S|b)$, the posterior probability of an initiating AP given only basal input, and (ii) $P(S|b,a)$, the posterior probability of a second AP given both basal and apical input. Using Eqs (14)-(17) with the weight of evidence terms for the somatic site that are given in Table 3 in the Methods section, we write these posterior probabilities as

$$\pi_b \equiv P(Z=1|b) \equiv P(S|b) = \frac{1}{1+\exp\left(-L(S|b)\right)},$$

$$\pi_{ba} \equiv P(Z=1|b,a) \equiv P(S|b,a) = \frac{1}{1+\exp\left(-L(S|b,a)\right)}, \tag{7}$$

9

where the respective log odds are

$$L(S|b) = -5.2933 + b \quad \text{and} \quad L(S|b,a) = -5.2933 + \tfrac{1}{2}b[1 + \exp[\tfrac{1}{2}ba(1 + e^a)]] \qquad (8)$$

for basal input b and apical input a.

The posterior probabilities π_b, π_{ba}, given only the basal input, and given both basal and apical inputs for different positive strengths of apical input, are shown in Fig 2.

We see that due to apical amplification the posterior probability, π_{ba}, of a second AP given both apical and basal input is larger than the corresponding posterior probability, π_b, given only the basal input. Apical input has little or no effect when basal input is very weak or very strong, but has a dramatic effect at intermediate values of basal input strength. This supralinear effect becomes more pronounced as the strength of the apical input increases. Defining the basal threshold as that value of the basal input for which the posterior probability is 0.5, we notice, in particular, that this threshold decreases markedly as the strength of the apical input increases. This effect of apical input on response to basal input is well matched to that sketched in Fig 2c of [1].

An analysis of the binarised action potential data from a detailed multi-compartment model

Shai et al. [15] used a multi-compartmental model to produce data on the frequency of somatic spike output for 31 given numbers of basal inputs equally spaced between 0 and 300, and 21 given numbers of apical tuft inputs equally spaced between 0 and 200. They then developed a phenomenological composite sigmoidal model for the frequency data, thus providing an explanation for coincidence detection between basal and apical tufts. They also produced data on the number of action potentials for the same given combinations of basal and apical tuft inputs, but they did not report any analysis of these data.
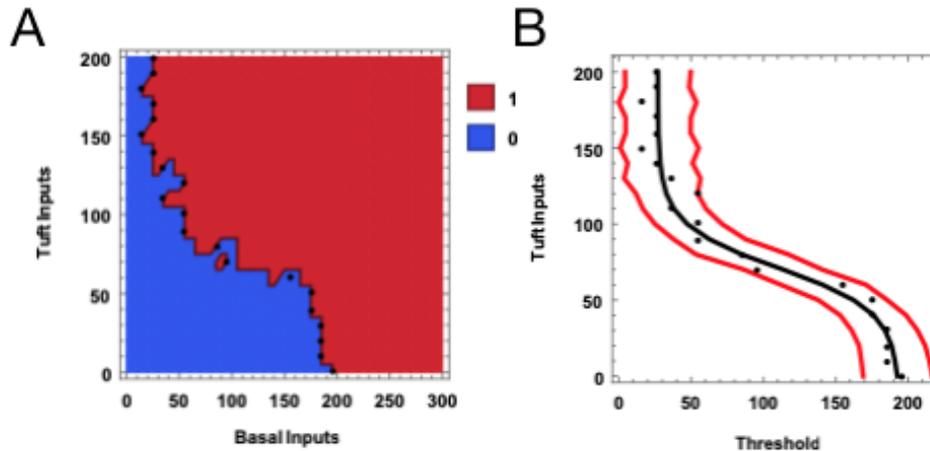

Figure 3: **The binary data & estimated thresholds**.
Illustrations of the binary AP data based on [15]. **(A)** An image of the binary data on a 21 by 31 grid, with points showing the estimated thresholds. **(B)** The predicted threshold curve given (in black) as the median of the posterior predictive threshold distribution, together with pointwise 95% posterior predictive intervals (in red) for each number of tuft inputs, obtained by applying weighted Bayesian nonlinear regression with the estimated thresholds (given as points) as responses, the number of tuft inputs as the explanatory variable, and the mean threshold given as a four parameter logistic function of the number of tufts inputs. These intervals were computed for each of the 21 numbers of apical inputs and interpolation used.

Our interest lies in modeling the posterior probability of a second AP within around 10 ms since an initial bAP has been generated. The number of APs in the data file ranges from 0 to 4. We are interested in the occurrence of a second AP, i.e. 2 APs, and this event happens also when 3 or 4 APs are observed. The data were therefore recoded by setting 0 or 1 AP to 0 and 2-4 APs to 1, thus creating a binary matrix where a '1' denotes the occurrence of a second AP and '0' means that this event has not happened. An alternative interpretation of these binary data, for each combination of the numbers of basal and apical inputs, is that a '1' indicates that bursting (2-4 APs) has happened while a '0' means 'no bursting' (0-1 APs).

The data are shown in Fig 3A. For each number of tuft inputs, the points of transition from 'blue' to 'red' along each row give a noisy indication of the threshold – the value of the basal input for which the posterior probability is equal to 0.5. It is clear from the plot that the thresholds vary according to the number of tuft inputs: the threshold is large when the number of tuft inputs is low whereas it is low when the number of tuft inputs is large. For each value of the tuft input, the transition point for each row in Fig 3A was estimated by fitting a penalized binary logistic regression model, and the estimates are shown as points in Figs 3A, 3B. A weighted Bayesian nonlinear regression model was fitted to these estimated thresholds, with mean threshold a four-parameter logistic function of number of tuft inputs. The threshold logistic curve given by the median of each posterior predictive threshold distribution is shown in Fig 3B, together with 95% pointwise prediction intervals; see the Bayesian Modeling subsection in the Methods section for further detail. The threshold logistic curve shows that the predicted threshold decreases monotonically as the number of tuft inputs increases. The pointwise posterior prediction intervals give an indication of the uncertainty of the predicted thresholds. There is greater uncertainty when the number of tuft inputs is larger than 120.

A tailor-made threshold model

The composite model in [15] is based on two logistic functions M, T of the tuft inputs. M models the maximum AP frequencies for each number of tuft inputs, while T models the thresholds. Since we consider binary output, for which the maxima are 1 for each number of tuft inputs, there is no need for M here (it is 1). We therefore model only the thresholds. The details of the modeling, weight of evidence and posterior predictive probability of a second AP are given in Eqs (19)-(22) the Bayesian Modeling subsection of the Methods section, with practical detail referenced in the Supplementary Information.

A general model

We also consider a general model by using the form of weight of evidence terms considered in Eq (10) in the Methods section. In particular, we employ the weight of evidence terms from the Somatic-Phase 2 entry in Table 2, in the Methods section, although in a more general form; see the Bayesian Modeling subsection of the Methods section for details of the construction used. A Bayesian binary logistic nonlinear regression model was fitted to the binary data of Fig 3A, with the numbers of basal and tuft inputs as explanatory variables. The details of the modeling, weight of evidence terms and the posterior predictive probability of a second AP are given by Eqs (23)-(26) in the Bayesian Modeling subsection of the Methods section, with practical detail referenced in the Supplementary Information.

A comparison of the two models

The threshold curves for the posterior predictive probability functions are displayed in Figs 4A, 4B for the threshold model and the general model, respectively.
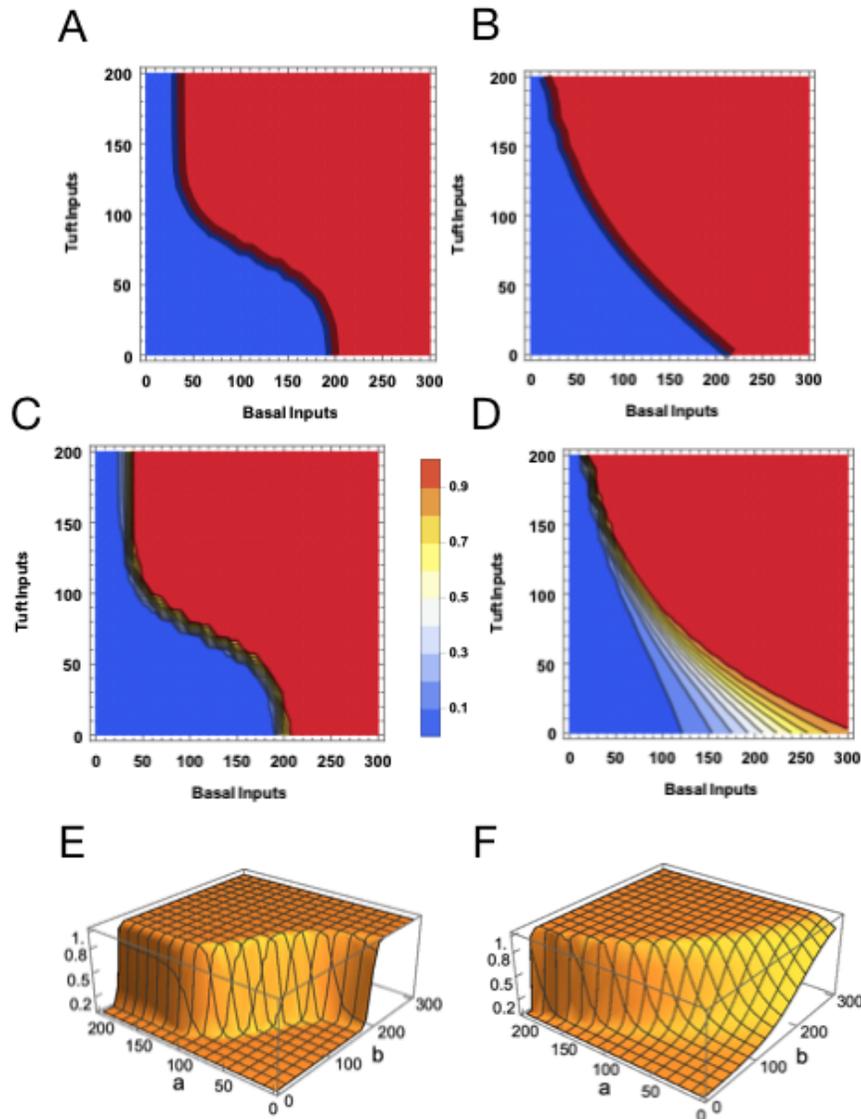
Figure 4: **Modeling the Binary Data.**
(A, B) The red (blue) regions show the set of values of the basal and tuft inputs for which the posterior predictive probability of a second AP is larger than or equal to (smaller than) 0.5, based on the threshold and general models, respectively, with the threshold curves in black. **(C, D)** Contour plots for the surface of posterior predictive probability of a second AP as a function of the numbers of basal and apical tuft inputs, based on the threshold and general models, respectively. **(E, F)** Surface plots of the posterior predictive probability of a second AP as a function of the numbers of basal and apical tuft inputs, based on the threshold and general models, respectively. In each figure, the posterior predictive probabilities were computed on a 21 by 31 grid and interpolation used.

In each plot the red(blue) region is the set of basal and apical inputs for which the posterior probability is greater than or equal to (less than) 0.5, based on the posterior predictive probabilities in Eqs (22), (26), respectively. The threshold curves in these two plots are not the same, but the red and the blue regions cover very similar combinations of basal and apical inputs, and have a large

overlap, despite the fact that no special modeling was performed to fit the general model from Eqs (23), (26). Another similarity is that the thresholds decrease monotonically with increasing apical input for both models.

Contour plots of the posterior predictive probability functions are displayed in Figs 4C, 4D. The regions of very high probability (> 0.9), or very low probability (< 0.1), are not identical but again they share similar combinations of numbers of basal and apical tuft inputs, and they have a large overlap. Maximum probability is attained not only when both the basal and apical inputs are large, which indicates a form of coincidence detection, but also when the basal input is large (200-300) while the apical input is low (0-50).

Figs 4E, 4F are surface plots of the posterior predictive probability functions. For the threshold model in Fig 4E, we notice the rather sharp transitions from almost zero probability on one side of the threshold curve to probability close to unity on the other side of the threshold curve. The general model in Fig 4F also shows such sharp transitions, especially when the apical input is large (100-200) while the basal input is low (0-100), and more gradual transitions when the apical input is lower (0-100) and the basal input is large (200-300). The two surfaces are generally similar in that for both models the sets of basal and apical inputs for which the posterior probability is close to unity, or close to zero, have a large overlap. Comparison of Fig 4E with Fig 4F shows that the posterior predictive probability surface for the general model rises more sharply for large numbers of apical and low numbers of basal input, and less sharply for lower numbers of apical input and all levels of basal input. This feature can also be noticed the contour plots in Figs 4C, 4D.

The fit of each model to the binary response data was assessed by comparing the predictions given by the model with the 651 binary responses. For the threshold model, 4.2% of the responses were misclassified, whereas the error for the general model was 5.4%. Based on this posterior predictive assessment of model fit we find that the general model performs very well, and almost as well as the threshold model. The misclassifications occur mostly near the points of transition from blue to red in Fig 3A. The application of tenfold cross-validation in order to assess 'out-of-sample' prediction produced similar results: 4.3% for the threshold model and 5.5% with the general model. This similarity between 'in-sample' and 'out-of-sample' performance is due to the structure of the binary AP data.

Application of information theory

We argue above that apical input can amplify the transmission of information about basal perisomatic input. To be made rigorous this requires quantification of information transmitted uniquely about each of the two inputs. That cannot be adequately done using classical information theory, because mutual information in that theory is defined only for a single input and a single output. There have recently been advances in the decomposition of multivariate mutual information, however, and these recent advances have now been used to specify criteria for distinguishing modulatory from driving interactions in neural systems [30]. For definition of the partial information decomposition as well as discussion of the link between Bayesian inference and Shannon's classical measure of mutual information, see the Methods section. We now apply Shannon's classical information measures, together with partial information decompositions, to a categorised version of the action potential data produced by the detailed compartmental model reported by Shai et al. [15].

The categorised AP data based on [15]

The output response variable, O, is the number of APs emitted for each combination of the basal and apical inputs. The numbers of APs range from 0 to 4, and they have been recoded into three ordinal categories, $O_1 - O_3$, containing 0-1 APs, 2 APs, 3-4 APs, respectively, since there are relatively few observations where 1 or 4 APs were obtained. The basal input was recoded into four ordinal categories: 0-60, 70-140, 150-220, 230-300 inputs, coded as $B_1 - B_4$, respectively. The apical input was recoded into four ordinal categories: 0-50, 60-100, 110-150, 160-200 inputs, coded as $A_1 - A_4$, respectively. This created a 4 by 4 by 3 contingency table of the recoded basal and apical inputs and the AP output. The data are displayed in Fig 5, which shows the proportions for the three AP categories for each combination of the four categories of basal input and the four categories of apical input. For the lowest category of basal input (0-60), the AP count is almost entirely 0-1 when the apical input is 0-100, but for an apical input of 110-200 we see that the proportion of observations with AP count 2-4 is about 50%. In the second lowest basal category (70-140), the AP count is 0-1 for the lowest apical category but 2-4 APs for observations in the higher apical categories (60-200). This trend from blue to red via green continues into the highest two basal categories where the AP count is 3-4 in the highest two apical categories (110-200).
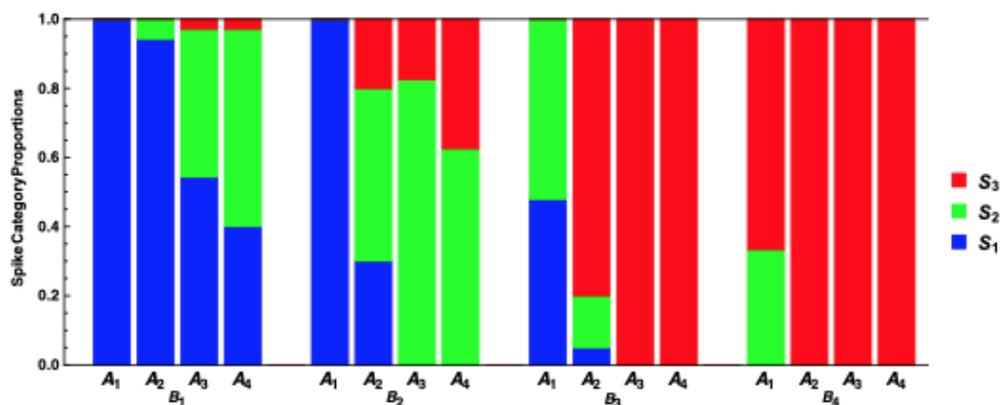

Figure 5: **AP response distributions**.
A bar chart illustrating the proportion of the observations within each of the three output AP categories O_1 (0-1), O_2 (2), O_3 (3-4) for each combination of the input basal categories B_1 (0-60), B_2 (70-140), B_3 (150-220), B_4 (230-300) and the input apical categories A_1 (0-50), A_2 (60-100), A_3 (110-150, A_4 (160-200), based on the categorised version of the AP data from [15].

The information measures were computed from these data and the values obtained are reported in Table 1. The joint mutual information between the output and the joint distribution of the basal and apical inputs is 1.0098 bits, and we notice that the mutual information between the basal input and the output, 0.5070 bits, is almost three times larger that the mutual information between the apical input and the output, 0.1783 bits.

Table 1: **Estimated information measures**.
The estimated values of the Shannon information measures (in bits) in Eqs (28)–(32) for the categorised version of the AP data from [15], where O denotes the output AP category, B denotes the basal input category and A is the apical input category. Here the output variable is the output spike category O rather than the binary output Z.

$I(O; B)$	$I(O; A)$	$I(O; B\|A)$	$I(O; A\|B)$	$I(O; B, A)$	$II(O; B; A)$
0.5070	0.1783	0.8315	0.5028	1.0098	0.3245

It is well known that these estimates are biased upwards and so estimates of the bias were obtained. Since the number of observations (651) is very large one might expect the biases to be small, and in fact they would affect only the third significant figure in each of the estimates in Table 1 if a bias correction were to be implemented.

We wish in particular to estimate the synergy in the system, since the synergistic effect of basal and tuft input is mentioned in [15]. Input from both of two distinct sources may be necessary for some transmitted information to be present. Synergy as defined within a partial information decomposition (PID) quantifies that transmission (see the Methods section), and it plays a key role in the notion of amplification because it should be strong only when the signal being amplified is present but not strong [30]. The estimate of the interaction information reported in Table 1 is approximately 0.32 bits, and so we can deduce from Eq (32) that the estimated synergy in the system is at least 0.32 bits.
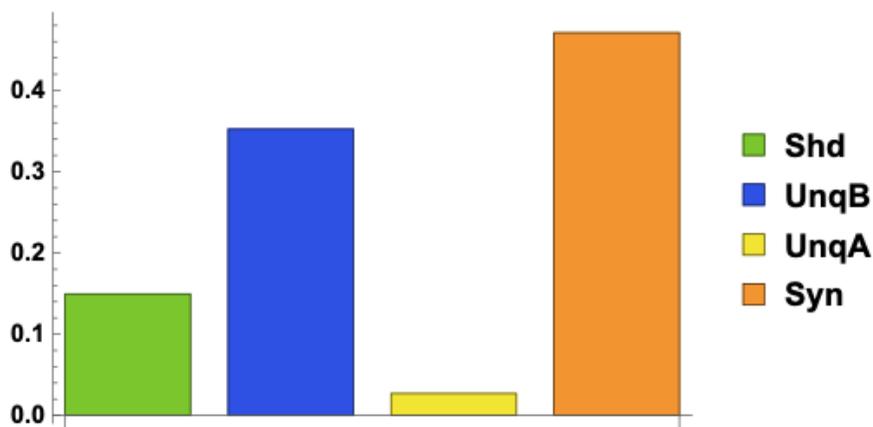

Figure 6: **Partial information decomposition**
A bar chart indicating the values (in bits) of the shared, unique and synergistic partial information components obtained using the categorised version of the AP data from [15].

The partial information decomposition [34, 35] was applied to the data and a normalised version, in which the PID components are divided by the joint mutual information $I(O; B, A)$, is shown in Fig 6. Synergy amounts to 47% of the joint mutual information, while the shared component takes up 15% of the total. There is a marked asymmetry in the estimates of the unique informations, in that the unique information due to the basal input is about eleven times larger than the unique information due to the apical input. This suggests that the apical input can amplify the information transmitted by the basal input in relation to the number of action potentials that are propagated, while conveying only a very small amount of information about itself. This

lends support to the presence of apical amplification within the system. See e.g. [30].

Four other PIDs were also applied to the data. Two of them [36, 37] gave very similar results to those quoted above. Two other methods [38, 39] produced a lower value for synergy (40%) and a larger value for UnqA. The asymmetry between UnqB and UnqA persists, however, but only by a factor of four rather than eleven. The PIDs were computed using the Python package, *dit* [40].

Alternative modes of apical function

We have so far considered the case where apical input is purely amplifying and basal input is purely driving. The distinction between drive and amplification does not depend on this dichotomy, however, and intermediate cases are likely to occur. Figure 6, for example, shows that a small amount of information was transmitted uniquely about the apical input. If the apical input were purely driving then it would not have been small, but zero. Therefore, we now consider a wider range of cases.

We first consider the unlikely, but theoretically possible, case where the functional asymmetry between apical and basal inputs is fully reversed, with apical being purely driving and basal being purely amplifying. We then consider the wide range of intermediate cases where apical and basal inputs can be partly driving and partly amplifying to various extents. Finally, we consider the case where apical and basal inputs are simply summed linearly, as is often assumed. Consideration of this wider range of cases will facilitate interpretation of any evidence of contextual feedback to layer 5 or of feedforward input to layer 1, which may occur for various reasons to be discussed in detail elsewhere. In each of the cases considered in this section (except in Figs 7C, 7D), $P(S|b,a)$ is the posterior probability of an AP given basal and apical input and not, as before, the probability of a second AP.

The first scenario is where there is basal amplification of the response to apical input when there has been no initiating bAP. The definitions of weights of evidence used here are given by Eqs (36), (37) in the Methods section.

Posterior probabilities of an AP given apical drive alone are displayed in Figs 7A, 7B whereas posterior probabilities of a second AP given an initiating bAP and consequent BAC firing are shown in Figs 7C, 7D. In Fig 7A, the posterior probability of an AP is plotted as a function of positive apical input for values of the basal input ranging from 0.1 to 2.0. Even when the basal input is very weak at 0.1, we see that the probability of an action potential approaches unity when the apical input is large; the primary drive provided by the apical input is mostly responsible for this behaviour. On the other hand, when there is appreciable basal input, saturation at unity occurs very quickly for even small values of apical input. These characteristics are also evident in Fig 7C but here the probability saturates even when the basal input is less than 1. The surface plots in Figs 7B, 7D illustrate the rate at which the posterior probability saturates at unity for various values of the apical and basal inputs. The plots indicate a slight asymmetry, but for most values of a and b they are very similar, apart from the fact that the rate of saturation is more gradual in Fig 7B than in Fig 7D for lower values of a and b. Thus our weight of evidence terms suggest that basal driving coupled with BAC firing has a stronger effect than apical drive alone, especially when the level of drive is low.
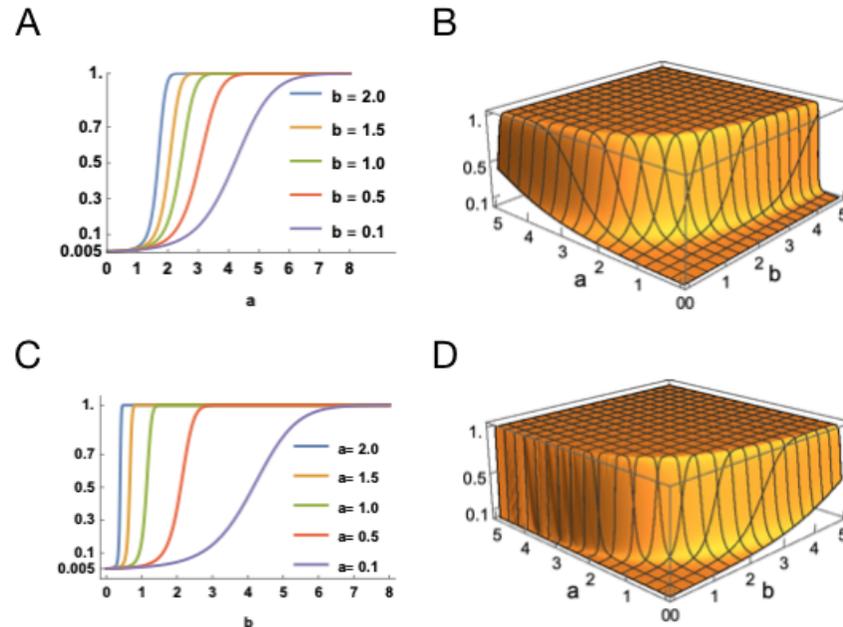
16

Figure 7: **Comparing apical drive alone with basal drive**
(A) Posterior probabilities of an AP plotted against positive values of the apical input for several values of the basal input & **(B)** a surface plot of posterior probability of an AP plotted as a function of the apical and basal inputs, both using Eqs (36), (37) in the Methods section. **(C)** Posterior probabilities of a second AP plotted against positive values of the basal input for several values of the apical input & **(D)** a surface plot of posterior probability of a second AP plotted as a function of the apical and basal inputs, both using Eqs (7), (8). In (C) & (D) , the response probability increases rapidly when a and b are both positive, which illustrates a version of 'coincidence detection'. The typical a priori firing probability, $P(S)$, is assumed to be 0.005.

The second scenario is where the amplification results from a mixture of basal and apical inputs, and we also consider the case where the basal and apical input combine additively. The definitions of weights of evidence used here are stated in Eqs (39), (40), (41) in the Methods section.

In Fig 8A, the posterior probability of an AP is plotted as a function of the basal input under an equal mixture of basal and apical amplification. The posterior probability curves saturate for small values of basal input and larger values of apical input, although larger values of basal input are required to produce saturation when the apical input is less than 1. The curves here are similar to those in Fig 7C especially when the basal input is between 0 and 1 and the apical input is between 1 and 2. These similarities given by our weight of evidence terms suggest that the probability of an AP when there is an equal mixture of basal and apical driving is similar to the probability of a second AP given an initiating bAP and consequent BAC firing, for these ranges of basal and apical input. By way of contrast Fig 8C, in which the posterior probability of an action potential is plotted as a function of basal input when there is no amplification and basal and apical inputs are combined additively, reveals a quite different picture; strengthening apical input increases the probability of an action potential but has little effect on the rate at which it increases with the strength of basal input. Comparison of the surface plots in Figs 8B, 8D indicates that the presence of amplification, in contrast to linear summation, is shown by the steepening of the probability surface in Fig 8B. This result has a lot in common with [41]
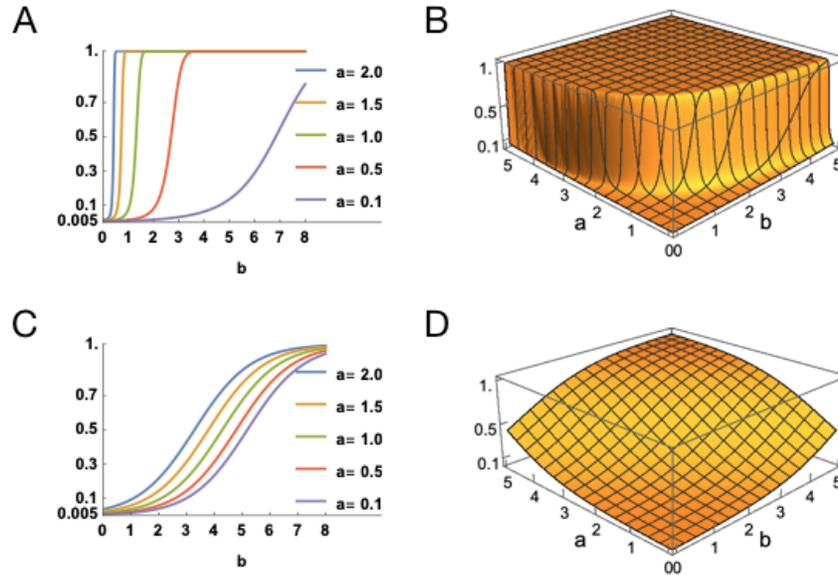
Figure 8: **Equal mixture of basal and apical drive**
(A) Posterior probabilities of a somatic action potential plotted against positive values of the basal input for several values of the apical input & **(B)** a surface plot of posterior probability plotted as a function of the apical and basal inputs, both using Eqs (39), (40) in the Methods section. **(C)** Posterior probabilities of a somatic action potential plotted against positive values of the basal input for several values of the apical input & **(D)** a surface plot of posterior probability plotted as a function of the apical and basal inputs, both using Eq (41) in the Methods section. In (D), the rate of increase of the response probability is much slower than in (C). The typical baseline firing probability, $P(S)$, is assumed to be 0.005.

Discussion

An effective way to adapt behavior to an uncertain environment is by integrating prior knowledge with ascending sensory data. This can be done by combining prior knowledge and new data to make probabilistic inferences in what is often referred to as the 'Bayesian brain' [42]. Though it is clear that probabilistic inference is central to cortical function [43], we now need to establish which particular form or forms of probabilistic inference occur in the neocortex, and how they are implemented at the subcellular and microcircuit levels. There is ample evidence that neocortex uses a form of probabilistic Bayesian inference in which ascending data and internal contextual variables interact [19], but the exact nature of that interaction and its neuronal realization are major issues in neuroscience. Our analyses here show that pyramidal neurons with BAC firing can turn the calculation of the probability that a feature is present given the basal data into a simple context-sensitive Bayesian calculation of the presence of the feature given the basal data and the apical input. We applied Bayes' theorem in a non-standard way, so it may be helpful to summarise the reasons for that. Consider again pyramidal neurons with two integration zones, apical 'a' and basal 'b' and with binary output z transmitting information about a particular feature. Our analyses show that BAC firing enables the cell to compute posterior output probability given both basal and apical inputs, i.e.,

$$P(z|b,a) = P(z|b)p(a|z,b)/p(a|b)$$

Note that all three terms in this decomposition are conditioned on b. It is also possible to use Bayes' theorem to express output probability in a way that conditions all three terms on a, as do

18

Lee and Mumford [19], i.e.,

$$P(z|a, b) = P(z|a)p(b|z, a)/p(b|a)$$

Computing output probability in that way, however, would require knowing the contribution of a to output before knowing b, which is contrary both to what we know about BAC firing and to the logic of contextual amplification [20].

Our analyses also show that our Bayesian decomposition can be more simply expressed as the sum of three terms, i.e.

$$L(S|b, a) = L(S) + W(S : b) + W(S : a|b)$$

Where $L(S|b, a)$ is the log odds of a second AP, $L(S)$ is the prior log odds, $W(S : b)$ is the weight of evidence for an initiating AP given by b, and $W(S : a|b)$ is the weight of evidence of a second AP provided by a given b. As these and the other key conclusions have already been outlined in the Introduction, the remainder of our discussion is focused on some of the many limitations and unresolved issues that arise.

A key aspect of what we propose here is that the Bayesian "calculation" is being carried out at the cellular level due to the properties and morphology of pyramidal neurons. This is important because a network of point neurons, such as is common in machine learning and common descriptions of the brain, would need to carry out the same process via multiple operations at the circuit level. By transferring this operation to the cellular level, it becomes a privileged operation that can be carried out in a massively parallel fashion. It also implies that this operation is central and intrinsic to what the brain actually computes. Whether the strategy taken by the brain to "hard wire" this operation into the fabric of the cellular tissue itself is done purely for means of efficiency or whether it qualifies as a new approach to network computing remains to be tested.

Another unresolved issue concerns the contrast between amplifying and multiplicative interactions that has been clearly demonstrated using recently developed methods for multivariate information decomposition [16]. This shows that purely multiplicative interactions transmit information only about synergistic relations between the two interacting inputs, whereas amplifying interactions use one input to modulate transmission of information uniquely about the other. We assume that both forms of interaction occur in neocortex, but there are as yet few explicit empirical attempts to distinguish between them. It is often assumed that contextual interactions are multiplicative, e.g. [44], but that does not explain the clear functional asymmetry between somatic and apical sites of integration, nor does it explain how input to one site can amplify transmission of information specifically about input to the other. The contrast between amplifying and multiplicative interactions is therefore in need of more extensive theoretical and empirical investigation.

A major limitation of the current analyses is that they considered only depolarizing inputs to basal and apical dendrites, so their extension to include the effects of hyperpolarizing inputs is clearly needed. That will require clarification of the communication of hyperpolarizing potentials between apical and somatic sites, of the time courses of recovery from hyperpolarization, and of the local inhibitory microcircuitry. Such potentials are likely to act mostly locally, so communication will be minimal. Closer attention to these issues is likely to fundamentally enhance our understanding of apical function for at least three reasons. First, the types of inhibitory interneuron that target distal apical dendrites are clearly distinct from those that target basal and perisomatic regions [45, 46]. Second, there is clear evidence for microcircuits that amplify the output of selected pyramidal cells by specifically disinhibiting the tuft [47]. Third, there is evidence for a distinctively human class of inhibitory interneuron that targets tuft dendrites in a highly specific way [48].

The idealizations analyzed here are based on evidence that has been largely, though not wholly, collected from in vitro studies of thick-tufted layer 5 cells of mature rodent sensory cortex. Though there are grounds for supposing that they have broader relevance, the extent to which these ide-

alizations can be generalized is not yet known because variations in apical function across layers, regions, species, stages of development, and states of arousal remain largely unexplored. There are already some relevant discoveries and tantalizing hints, nevertheless. For example, there is evidence that somatic and apical sites are both predominantly driving in infants, whereas the apical site has become predominantly amplifying in mature animals [49, 17]. There is also evidence that apical input remains predominantly driving in the most posterior, caudal, part of mature rodent primary visual cortex (V1), whereas it has become predominantly amplifying in other parts of V1 [14]. Cross-species variations in apical function may also be of great importance because direct electrical recordings from the somatic and apical sites of human layer 5 pyramidal neurons show that they have enhanced electrotonic separation from the soma, as compared to those from rodents [18]. Central roles for feedforward drive and feedback modulation are well-established in the hierarchical regions of posterior cortex, but, as the various regions of the prefrontal cortex do not seem to be organized into a clear hierarchy, driving and amplifying functions may be less clearly distinguished at both somatic and apical sites within prefrontal cortex. Furthermore, though there are grounds for supposing that apical function depends strongly on the state of adrenergic and cholinergic arousal [7], this issue has as yet received only scant attention.

Because of this and other evidence that, under some circumstances, apical depolarization can drive somatic output of the cell in the absence of basal drive [41], we extended our analyses to include cases where each of the somatic and apical sites can be driving and amplifying to various extents. Intermodal effects in primary unimodal sensory regions provide clear examples of cases where apical input is predominantly amplifying. For example, anatomical and physiological evidence indicates that the effects of auditory information on pyramidal cells in V1 sharpens their selective sensitivity and raises the salience of their output via the apical dendrites in layer 1 [50]. Nevertheless, as noted above, the balance between driving and amplifying effects of apical input may vary even within a single neocortical region such as V1 [14]. There is some hint of a sharp transition from a driving to an amplifying effect of apical depolarization as apical trunk-length increases in V1 [14], but whether there is a clear dichotomy or a continuum between these two forms of apical function in other regions remains unknown. Whatever the resolution to that issue, however, system-level network diagrams showing inputs to pyramidal neurons represented as single undifferentiated points of integration, though abundant, seem grossly underspecified, because it is clear that in many pyramidal neurons inputs to their basal and apical dendrites have very different effects on neuronal output, and thus on system-level dynamics.

Context-sensitive selective amplification has a central role in both learning and processing at all levels of the abstraction hierarchy, so grounding it firmly in subcellular processes has far-reaching implications: it strengthens prior hypotheses concerning the role of apical function in perception, attention, thought, and learning (e.g. [4, 9, 22, 51, 52]) by showing explicitly how BAC firing implements the primitive operations of amplification, attenuation, and disattenuation [5]; it explains how apical malfunction can have a central role in pathologies as diverse as schizophrenia [53, 54], autoimmune anti-NMDAR encephalitis [53, 55], absence epilepsy [56], and foetal alcohol spectrum disorder [57, 58]; it suggests that, in addition to being able to implement deep learning algorithms [59, 60, 61], local processors with two sites of integration can use one of them as a context that guides both learning and processing, thus enhancing the capabilities of such algorithms to an extent which, though yet to be realized, is unlikely to be small.

20

Methods

Bayesian interpretation of computation in the apical integration site

We provide just a brief version of the Bayesian decompositions of the posterior log odds for the computation in the apical site. We let the event T denote 'BAC firing is triggered', with complementary event, $\bar{T}$. The first decomposition of the posterior log odds in favor of BAC firing being triggered is

$$L(T|a) = L(T) + W(T:a).$$

This says that the posterior log odds in favor of BAC firing being triggered given the apical input is equal to the sum of the prior log odds in favor of BAC firing and the weight of evidence in favor of BAC firing provided by the apical input. This decomposition is relevant when no initiating bAP has been received from the somatic site.

Then when a bAP has been received there is amplification of the apical activation when the apical input is positive, and so the decomposition of the posterior log odds in afvor of BAC firing becomes

$$L(T|a, z) = L(T|a) + W(T:z\,|\,a) = L(T) + W(T:a) + W(T:z\,|\,a).$$

Here there is an additional term $W(T:z\,|\,a)$ due to the communication between sites, which is the weight of evidence in favor of BAC firing provided by the initiating bAP given the apical input.

A summary of the Bayesian decompositions

In Table 2, we summarise for both sites the general Bayesian decompositions of the posterior log odds.

Table 2: **Bayesian decompositions**.
Bayesian decompositions of the posterior log odds in terms of the prior log odds and weights of evidence during Phase 1 (when no initiating bAP has been propagated from the somatic site) and Phase 2 (when an initiating bAP has been generated from the somatic site and there is consequent BAC firing). General decompositions are given for both the apical and somatic sites of integration.

State	Site	General form
Phase 1	Apical	$L(T|a) = L(T) + W(T:a)$
Phase 1	Somatic	$L(S|b) = L(S) + W(S:b)$
Phase 2	Apical	$L(T|a, z) = L(T) + W(T:a) + W(T:z\,|\,a)$
Phase 2	Somatic	$L(S|b, a) = L(S) + W(S:b) + W(S:a\,|\,b)$

Having developed Bayesian decompositions of the various posterior log odds, we now consider the practical application of these formulations by defining expressions for the respective weight of evidence terms in the decompositions. In the following section, we define a general form of activation function and consider specific forms which will be used to define two three-term activation functions – one for the somatic site and one for the apical site.

Idealized forms of integration at the apical and somatic sites

We now define the general form of functions f and g, such that transmission of information about the external inputs is amplified by the internal inputs. In the following we refer to 'units' rather

21

than to 'neurons' because they are abstractions designed to show how one class of inputs can selectively amplify transmission of information about another class of inputs. Furthermore, although all inputs are assumed to be noisy, that is not explicitly studied here. We first consider a very general form of activation function which combines two real-valued inputs and produces a real-valued output. It forms the basis for the various choices to be made for the weight of evidence terms.

General form of the activation functions

The general form of activation function considered herein is

$$F(x, y) = f(x) + g(x, y), \tag{9}$$

where f is a function of the single real variable x, and g is a function of the two real variables x, y. These functions have the following general properties.

$P_1 :\ \text{sgn}[F(x, y)] = \text{sgn}[f(x)] = \text{sgn}[x]$

$P_2 :\ g(x, 0) = 0, \quad g(0, y) = 0$

$P_3 :\ \text{sgn}[g(x, y)] = \text{sgn}[y]$

Property P_1 ensures that the sign of the activation is the same as the sign of the terms $f(x)$ and x, and also that the terms $F(x, y)$ and $f(x)$ are zero when $x = 0$. Property P_2 makes clear that the activation $F(x, y)$ is equal to $f(x)$ when $y = 0$. It can be shown that Property P_3, together with Property P_1, ensures that the magnitude of the activation $F(x, y)$ is larger than the magnitude of $f(x)$ when the signs of x and y agree, whereas this activation is smaller in magnitude when the signs of x and y disagree.

In the sequel, we use the following specific choices for the functions f, g in Eq (9):

$$f(x) = x, \quad \text{and} \quad g(x, y) = (1 - s)x(\exp{(mxy)} - 1), \tag{10}$$

where $0 < s < 1$ and $m > 0$, and we have set the free constants, s, m, using the values $s = \frac{1}{2}, m = 1$, although in the work described below in the Bayesian Modeling subsection it is necessary to employ much smaller values for m.

Apical Integration Site

The apical activation function takes two different forms depending on whether or not an initiating backpropagating action potential (bAP) has been generated. When no such bAP has been received at the apical site ($z = 0$), the apical activation, and weight of evidence term, is, on setting $x = a$ and $y = 0$ in Eq (9),

$$W[T : a] \equiv F(a, 0) = f(a) + g(a, 0) = f(a),$$

by Property, P_2. Defining the weight of evidence term $W[T : z \,|\, a]$ to be $g(a, 1)$, we have that, on the other hand, when such a bAP has been received ($z = 1$) the apical activation, and weight of evidence, is on setting $x = a$ and $y = 1$ in Eq (9),

$$F(a, 1) = f(a) + g(a, 1) \equiv W[T : a] + W[T : z \,|\, a].$$

We illustrate the properties of the activation function in Eq (9) at this site. Property P_1 ensures that the sign of the activation function is that of the apical input and also equal to the sign of $f(a)$; when the apical input is zero then so is the activation. From Property P_2, we have that when

22

$z = 0$, so no bAP has been received from the somatic site, then the activation is equal to $f(a)$ and is based on the apical input alone. Property P_3, in addition to Property P_1, guarantees that when $z = 1$, and so an initiating bAP has been received from the somatic site, then the magnitude of the activation, F, is increased.

Using the special choices in Eq (10), we find that the specific forms of apical activation, and weight of evidence, become

$$F(a, 0) = a, \quad \text{and} \quad F(a, 1) = \tfrac{1}{2}a[1 + \exp(a)].$$

Somatic Integration Site

The activation function in this site takes two forms depending on whether or not an initiating bAP has been generated. When no such bAP has been generated the activation, and weight of evidence term, in this site is, on taking $x = b$ and $y = 0$ in Eq (9),

$$W[S : b] \equiv F(b, 0) = f(b) + g(b, 0) = f(b), \tag{11}$$

by Property P_2. Defining the weight of evidence term $W[S; a \,|\, b]$ to be $g(b, c)$, we have that, on the other hand, given a bAP and subsequent BAC firing, the basal activation, and weight of evidence term, is, on taking $x = b$ and $y = c$ in Eq (9),

$$F(b, c) = f(b) + g(b, c) \equiv W[S : b] + W[S : a \,|\, b], \tag{12}$$

where $c = f(a) + g(a, 1)$, and f and g are given in Eq (10).

Property P_1 ensures that the sign of the basal activation function is that of the basal input and also equal to the sign of $f(b)$; when the basal input is zero then so is the activation. From Property P_2, we have that when $z = 0$, so there has been no BAC firing, then the activation is equal to $f(b)$ and is based on the basal input alone. Property P_3, in addition to Property P_1, guarantees that when $z = 1$, and so BAC firing has been triggered, then the magnitude of the activation, F, is increased when the basal input is positive.

Using the choices of functions from Eq (10), the somatic activation function, and weight of evidence, has the particular forms

$$F(b, 0) = b, \quad \text{and} \quad F(b, c) = \tfrac{1}{2}b[1 + \exp(bc)], \tag{13}$$

when $z = 0$ and $z = 1$, respectively. Note that $c = \tfrac{1}{2}a[1 + \exp(a)]$.

The specific activation functions, and weight of evidence terms, employed are summarised in Table 3.

Table 3: **Weight of Evidence terms.**
Specific expressions for the activation functions/weight of evidence terms in the apical and somatic sites during Phase 1 (when no initiating action potential is emitted from the basal site) and Phase 2 (when an initiating bAP has been received from the basal site and BAC firing has been triggered)

Site	Phase 1 $z = 0$	Phase 2 $z = 1$
Apical	a	$\frac{1}{2}a(1 + e^a)$
Somatic	b	$\frac{1}{2}b[1 + \exp[\frac{1}{2}ba(1 + e^a)]]$

23

Posterior probabilities in the somatic site

We now develop expressions for the posterior probabilities of an initiating AP, given basal input and a second AP, given both basal and apical input, respectively, based on Eqs (5), (6). First we note that $P(S)$ gives the prior probability of an AP, which we set to 0.005.

Therefore, we may write

$$L(S) = \log \frac{P(S)}{P(\bar{S})} \equiv \beta_1 \doteq -5.2933. \tag{14}$$

Using Eqs (11), (12) with Eqs (5), (6) we have that

$$L(S|b) = \beta_1 + f(b) \quad \text{and} \quad L(S|b,a) = \beta_1 + f(b) + g(b,c). \tag{15}$$

By using the connection between probability and log odds, we can write the posterior probability of (i) an initiating AP given basal input alone and (ii) a second AP given both basal and apical input, from Eq (15), as

$$\pi_b \equiv P(Z = 1|b) \equiv P(S|b) = \frac{1}{1 + \exp\left(-L(S|b)\right)}, \tag{16}$$

$$\pi_{ba} \equiv P(Z = 1|b,a) \equiv P(S|b,a) = \frac{1}{1 + \exp\left(-L(S|b,a)\right)}. \tag{17}$$

These equations show, from Eqs (5), (6), that the posterior probabilities can be expressed directly in terms of prior log odds and weight of evidence.

We also note that the general somatic activation function, $f(b) + g(b,c)$, can now be interpreted as either the sum

$$W(S : b) + W(S : a \,|\, b),$$

or the single term $W(S : b,a)$ since [28]

$$W(S : b,a) = W(S : b) + W(S : a \,|\, b),$$

which says that the weight of evidence in a favor of a somatic action potential provided by the basal and apical inputs taken together is simply the sum of (i) the weight of evidence in favor of an initiating AP provided by the basal input alone, b and (ii) the weight of evidence provided by the apical input, a, given the basal input, b.

From Eq (6) and Eq (15) we have that $L(S|b,a) = L(S|b) + g(b,c)$ and we use this equation to describe the effect of apical amplification. When $b > 0$ and $a > 0$ it follows that $c > 0$ and that $g(b,c) > 0$, and so the posterior log odds in favor of a second AP based on both basal and apical input is larger than the corresponding prior log odds given only basal input. The effect of apical amplification is to add a positive amount to the prior log odds of an initiating AP given just the basal input, thus increasing the posterior log odds and making the firing of another AP more likely.

Bayesian Modeling

We provide details here of two statistical models for the data as well as the particular weight of evidence terms used in the construction of the posterior probability of a second AP. For further details of the Bayesian methods described, see [62, 63].

A threshold model

For each of the $n_a = 21$ apical inputs, $\mathbf{a} = \{a_i\}$, used in [15], a penalized binary logistic regression model was fitted, using 'R' [64, 65, 66, 67]. The explanatory variable was the number of basal inputs, $\{b_i\}$. The level of basal input for which the probability of a second AP is equal to 0.5 was estimated to give the estimated thresholds, $\mathbf{t} = \{t_i\}$, and the estimated standard errors associated with these estimates were noted and used to define weights in the subsequent analysis. The weight, w_i, for each threshold estimate was taken to be the estimated standard error associated with the estimated threshold, t_i. The basal and threshold values were each scaled to lie in the interval [0, 2], and the weights were changed accordingly. A weighted Bayesian nonlinear regression model, with $p(t_i|a_i, \boldsymbol{\theta})$ assumed to be a Gaussian p.d.f with mean

$$\mu(a_i) = \theta_1 + \frac{\theta_2}{1 + \exp\left(-(a_i - \theta_3)/\theta_4\right)}$$

and variance $(\sigma/w_i)^2$, where the $\boldsymbol{\theta}$ are unknown parameters. The θ_i and σ were assumed *a priori* to be mutually independent with weakly informative priors: Gaussian with mean 0 and variance 10 for the θ_i and Uniform on [0, 20] for σ. Given the context the constraints, $\theta_2 > 0, \theta_4 < 0$, were added to ensure that an increasing logistic function was fitted. This model was fitted using rstan [68], and large samples from the posterior distributions of the parameters were produced for subsequent computation of the posterior predictive probabilities of a second AP, as described below in Eqs (19) − (26). Further detail is referenced in the Supplementary Information.

It is also of interest to predict the threshold, t_{new}, given a new value, a_{new}, of the apical input. The posterior predictive distribution for t_{new} is given by

$$p(t_{\text{new}}|a_{\text{new}}) = \int p(t_{\text{new}}|a_{\text{new}}, \boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{t}, \mathbf{a})\, d\boldsymbol{\theta}, \tag{18}$$

where the dependence on $(\mathbf{t}, \mathbf{a})$ is dropped from $p(t_{\text{new}}|a_{\text{new}})$. The same Gaussian model, as described above, was assumed for each t_{new}, and the model extended to include these specifications. For each new apical input, a_{new}, a large number of values for the predicted threshold were then produced to provide a sampled version of the posterior predictive distribution in Eq 18. The prediction, $\hat{t}_{\text{new}}$, is taken to be the median of these sampled predictions, and pointwise 95% posterior predictive limits were obtained by using the 0.025th and 0.975th quantiles of the sampled posterior predictive distribution. Such calculations were used to create Fig 3B.

Let a, b be new values of the apical and basal input, respectively. By analogy with the composite model in [15], we define the weight of evidence in favor of a second AP provided by the net basal and apical inputs in terms of the parameter vector $\boldsymbol{\theta}$ as

$$W[S : b, a|\boldsymbol{\theta}] = b - \mu(a) \equiv b - \theta_1 - \frac{\theta_2}{1 + \exp\left(-(a - \theta_3)/\theta_4\right)}. \tag{19}$$

Then the posterior log odds in favor of a second AP under the threshold model is

$$L(S|b, a, \boldsymbol{\theta}) = L(S) + W[S : b, a|\boldsymbol{\theta}] \tag{20}$$

and so the posterior probability of a second AP under this threshold model is

$$P(S|b, a, \boldsymbol{\theta}) = \frac{1}{1 + \exp\left(-L(S|b, a, \boldsymbol{\theta})\right)}. \tag{21}$$

The posterior predictive probability of a second AP for new values b, a of basal and apical input, given the threshold and apical data $\{(\mathbf{t}, \mathbf{a})\}$ is then given by

$$P(S|b, a) = \int P(S|b, a, \boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{t}, \mathbf{a})\, d\boldsymbol{\theta}, \tag{22}$$

25

where $p(\boldsymbol{\theta}|\mathbf{t}, \mathbf{a})$ is the posterior p.d.f. for the parameters $\boldsymbol{\theta}$ given the threshold and apical data and, for simplicity, the dependence on these data is omitted in $P(S|b, a)$.

A general model

We first describe how the weight of evidence term is composed. Here, as in the previous section, a and b denote new values of the apical and basal input, respectively. Using Eq (10), we can write the apical activation as

$$c = \tfrac{1}{2}a(1 + \exp{(\beta_4 a)},$$

where $m = \beta_4, s = \frac{1}{2}$ and x is replaced by the apical input, a. When BAC firing has been triggered, we can then write from Eq (13) the somatic activation function as

$$\tfrac{1}{2}b[1 + \exp{(\beta_3' bc)}],$$

where $m = \beta_3'$ and x is replaced by the basal input b. Thus the expression for the activation function becomes

$$\tfrac{1}{2}b[1 + \exp[\tfrac{1}{2}\beta_3' ba(1 + \exp{(\beta_4 a)})]].$$

We then define the weight of evidence term to be

$$
\begin{aligned}
W[S : b, a|\boldsymbol{\beta}] &= \beta_2 * \tfrac{1}{2}b[1 + \exp[\tfrac{1}{2}\beta_3 ba(1 + \exp{(\beta_4 a)})]] \\
&\equiv \beta_2 b[1 + \exp[\beta_3 ba(1 + \exp{(\beta_4 a)})]],
\end{aligned}
\tag{23}
$$

where $\beta_2 \equiv \tfrac{1}{2}\beta_2$ and $\beta_3 \equiv \tfrac{1}{2}\beta_2$, and $\boldsymbol{\beta} = \{\beta_2, \beta_3, \beta_4\}$, and the weight of evidence term is a function of the unknown parameter vector $\boldsymbol{\beta}$, which has positive components.

In Fig 3A, we see that there is almost complete separation between the 0's and the 1's. Therefore it is necessary to employ regularization, and here we take a Bayesian approach by defining a prior on $\boldsymbol{\beta}$ which enforces this. There are 651 combinations of basal and apical inputs, $\{b_i, a_i\}$, and for each combination there is a binary response, Z_i, that takes the value 1, with probability p_i, when a second AP occurs and 0 otherwise. For each i, Z_i has a Bernoulli distribution with parameter p_i. The binary logistic nonlinear regression model has the form

$$\log \frac{p_i}{1 - p_i} = P(S) + \beta_2 b_i * (1 + \exp{(\beta_3 b_i a_i * (1 + \exp{(1 + \beta_4 a_i)})}))).$$

The basal and apical data were scaled to lie in the intervals [0, 2] and [0, 3], respectively. The parameters, $\beta_2, \beta_3, \beta_4$ were assumed *a priori* to be mutually independent with each following a uniform probability model on (0, 10), which provides a weakly informative prior. This interval is chosen so that the β_i parameters will not be allowed to become too large, which they would otherwise do since there is almost complete separation between the 0's and the 1's.

Therefore, the posterior log odds in favor of a second AP under the general model is

$$L(S|b, a, \boldsymbol{\beta}) = L(S) + \beta_2 * b * [1 + \exp{(\beta_3 * b * a * (1 + \exp{(\beta_4 * a)}))}]
\tag{24}$$

and so the posterior probability of a second AP under this general model is

$$P(S|b, a, \boldsymbol{\beta}) = \frac{1}{1 + \exp{(-L(S|b, a, \boldsymbol{\beta}))}}.
\tag{25}$$

The posterior predictive probability of a second AP for new values, b, a, of basal and apical input, given the values, $\mathbf{b}, \mathbf{a}$ used in [15] is then obtained by

$$P(S|b, a) = \int P(S|b, a, \boldsymbol{\beta}) \, p(\boldsymbol{\beta}|\mathbf{b}, \mathbf{a}) \, d\boldsymbol{\beta},
\tag{26}$$

where $p(\boldsymbol{\beta}|\mathbf{b}, \mathbf{a})$ is the posterior p.d.f. for the parameters $\boldsymbol{\beta}$ given the basal and apical data, where again we drop the dependence on the data $\mathbf{b}, \mathbf{a}$ in $P(S|b, a)$.

The prior log odds is taken to be $L(S) \equiv \beta_1 = -5.2933$, as before. The Bayesian modeling was conducted using rstan [68]. The posterior predictive probabilities in Eqs (22), (26) were computed using the output from rstan as Monte Carlo approximations, using (27). We describe the formulae for the general model; they are analogous for the threshold model. The posterior predictive probability in Eq (26) can be written as a posterior expectation, as follows:

$$
\begin{aligned}
P(S|b, a) &= \int P(S|b, a, \boldsymbol{\beta}) \, p(\boldsymbol{\beta}|\mathbf{b}, \mathbf{a}) \, d\boldsymbol{\beta} \\
&\equiv \mathbb{E}_{\boldsymbol{\beta}|\mathbf{b}, \mathbf{a}} \left[P(S|b, a, \boldsymbol{\beta}) \right] \\
&\doteq \frac{1}{N} \sum_{i=1}^{N} P(S|b, a, \hat{\boldsymbol{\beta}}_i),
\end{aligned}
\tag{27}
$$

where $\hat{\boldsymbol{\beta}}_i$ is the estimate of $\boldsymbol{\beta}$ generated in the ith of the N simulations. For further detail, see the files referenced in the Supplementary Information.

An information-theoretic perspective

The synergistic effect of basal and apical tufts on output frequency is mentioned in [15], and so it is of interest to consider how a Bayesian formulation can be linked to measures of Shannon information and also to provide a numerical estimate of the synergy present in the probabilistic system based on the AP data from [15].

Writing in a Bayesian context, Lindley [69] introduced a measure of the information about an unknown parameter (or random variable) θ provided by an experiment involving the observation of a random variable X. The parameter θ could be either a discrete or continuous random variable or a list of hypotheses or models; so Lindley's measure of information is very general. In our formulation this parameter is the binary random variable Z.

Expected information from the basal input

Using our notation, there is interest in learning about the random variable Z by observing the basal input B, and obtaining the realised value, b. Given a prior distribution $p(z) \equiv P(Z = z)$ for Z the prior amount information about Z is

$$
I_0 = -\sum_{z=0,1} p(z) \log p(z).
$$

After the experiment has been performed yielding the realisation b of B, the posterior amount of information about Z is

$$
I_1(b) = -\sum_{z=0,1} p(z|b) \log p(z|b).
$$

Then the average amount of information provided the experiment of observing B, with prior knowledge $p(z)$, is

$$
\mathbb{E}_b \left(I_0 - I_1(b) \right) = \sum_{z=0,1} \int p(z, b) \log \frac{p(z, b)}{p(z)p(b)} \, db \equiv I(Z; B),
\tag{28}
$$

where $I(Z; B)$ is Shannon's (non-negative) mutual information [70] between the somatic output Z and the basal input B. Lindley's measure has been termed the *expected gain of information* in [71]

where it was used to sequentially select features in diagnostic tasks, and in [72] where it was used in the exploration of an environment by an agent. It is also used as an objective function in active data selection [73].

Additional information from the apical input

Having observed the basal input, the posterior distribution of Z is $p(z|b)$. We now apply a similar argument to that above to find the additional expected gain of information about Z by observing the apical input A, with realised value a. The posterior distribution of Z then becomes, by Bayes' Theorem, $p(z|b, a)$ and the additional information we expect to obtain by observing the apical input is

$$\mathbb{E}_{b,a}\left(I_2(a|b) - I_1(b)\right) = \sum_{z=0,1} \int \int p(z, a, b) \log \frac{p(z, a|b)}{p(z|b)p(a|b)}\, da\, db \equiv I(Z; A|B), \tag{29}$$

where $I(Z; A|B)$ is the conditional mutual information between the somatic output Z and the apical input A given the basal input B.

The total gain in information that is expected to become available about the nature of the somatic output by observing both the basal and apical inputs is equal to the expected gain from observing the basal input plus an additional expected gain from the apical input, given the basal input:

$$I(Z; B, A) = I(Z; B) + I(Z; A|B). \tag{30}$$

This expected gain of information can also be written as [74]

$$I(Z; B, A) = I(Z; A) + I(Z; B|A). \tag{31}$$

There is a further information measure called the 'interaction information' [75] and this involves all three variables

$$II(Z; B; A) = I(Z; B, A) - I(Z; B) - I(Z; A) \tag{32}$$

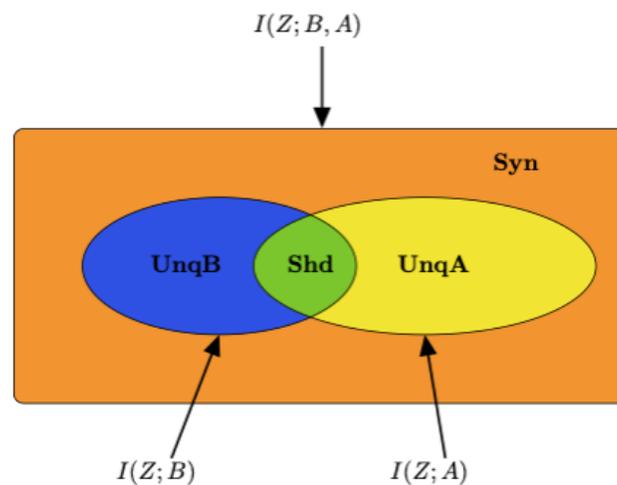and it has been used as a measure of synergy; see e.g. [76, 77].

Figure 9: **Williams-Beer diagram.**
A Williams-Beer diagram showing the decomposition of the joint mutual information between the output Z and the two inputs B and A into a sum of the four partial information components, Shd, UnqB, UnqA and Syn, as defined in the text.

Partial information decomposition

Due to seminal work by Williams and Beer [37], it is possible to provide a finer decomposition of the information that the apical and basal inputs provide about the somatic output. They decompose the joint mutual information between the output Z and the basal and apical inputs, considered jointly, into a sum of four non-negative terms: the shared information (Shd) that both the basal and apical inputs possess about the propagation of a somatic action potential (Z), the unique information (UnqB) that the basal input has about Z, the unique information (UnqA) that the apical input has about Z and the synergy (Syn) which is the information that the apical and basal inputs possess jointly about Z that cannot be obtained by observing these two variables separately. The partial information decomposition has been applied to data in neuroscience; see, e.g. [78, 79, 80, 81]. For a recent overview, see [82]. The Williams and Beer decomposition is illustrated in Fig 9, where use is made of [83]

One particular result obtained by Williams and Beer is that the interaction information can be expressed in terms of the synergy and shared information as

$$II(Z; B; A) = \text{Syn - Shd} \tag{33}$$

and this equation is used in the Results section to find a lower bound for the synergy of the system defined by the categorised AP data, based on the AP data in [15].

Alternative modes of apical function

We develop Bayesian decompositions of posterior log odds in three further hypothetical scenarios with which physiological data could be compared (e.g. [15]). We consider first the situation in which the apical input affects the probability that an action potential will be generated in the absence of a preceding initiating bAP, and where basal input amplifies somatic response to the apical input.

Basal amplification of response to apical input in the absence of a backpropagating potential

Here we consider the case where depolarization at the apical site is communicated to the somatic site in the absence of an initiating bAP and where its effect there is amplified by basal depolarization. Taken together, the apical input, a, and the basal input, b, will produce an updated posterior probability of a somatic action potential by direct use of Bayes' theorem:

$$P(S|a, b) = \frac{P(S) \times p(a, b|S)}{p(a, b)},$$

where $p(a, b|S)$ is the joint conditional probability density function for (a, b) given a somatic action potential. We convert this posterior probability into log odds form, as

$$L(S|a, b) = L(S) + W[S : a, b],$$

where $W[S : a, b]$ is the weight of evidence in favor of a somatic action potential provided by the apical input and the basal input considered jointly. In contrast with the scenario discussed above, here the posterior log odds is produced in a single step by taking both the apical and basal inputs together, rather than there being a two-step updating of the posterior log odds by b, then a, as in Eq (6) above.

Nevertheless, in a purely mathematical sense it is possible to write [28] the weight of evidence

term $W[S : a, b]$ as

$$W[S : a, b] = W[S : a] + W[S : b \,|\, a] \tag{34}$$
$$W[S : a, b] = W[S : b, a] = W[S : b] + W[S : a \,|\, b] \tag{35}$$

and so we can write the posterior log odds as

$$L(S|a, b) = L(S) + W[S : a] + W[S : b \,|\, a]. $$

We now define the weight of evidence term $W[S : a]$ to be $f(a)$, and the weight of evidence term $W[S : b \,|\, a]$ to be $g(a, b)$, where the functions f and g were defined earlier, and the specific forms are given in Eq (10). Using these specific forms, we obtain

$$W[S : a, b] = \tfrac{1}{2}a[1 + \exp(ab)]. \tag{36}$$

When a and b are both positive the weight of evidence in favor of an AP increases as the apical input increases. There could be some amplification of response to the the apical input by the basal input which would increase the weight of evidence, but if the basal input were to be very weak the weight of evidence could still be large enough to provide a large posterior probability of an AP; in this case the weight of evidence would be provided almost entirely by the apical input. It follows that the general form of the posterior probability of an AP given both apical and basal input takes the form

$$P(S \,|\, a, b) = \frac{1}{1 + \exp(L(S) + W[S : a, b])}. \tag{37}$$

The larger a positive a is then the larger the posterior probability is, provided that the basal input is also positive.

Mixtures of both apical and basal amplification

We can also provide a Bayesian interpretation of cases where both apical and basal amplification occur to various extents. Using (34), (35), the posterior log odds can be written as

$$\begin{aligned} L(S|b, a) &= L(S) + W[S : b, a], \\ &= L(S) + \alpha W[S : b, a] + (1 - \alpha) W[S : a, b], \\ &= L(S) + \alpha\big(W[s : b] + W[S : a \,|\, b]\big) + (1 - \alpha)\big(W[s : a] + W[S : b \,|\, a]\big), \end{aligned} \tag{38}$$

where the constant α lies between 0 and 1. If $\alpha = 1$ then we have the scenario in which the basal input is driving, while at the other extreme of $\alpha = 0$ the drive is from the apical input.

We now define the weight of evidence terms in the Bayesian decomposition of the log odds in Eq (38). We take $W[S : b] = f(b)$ and $W[S : a \,|\, b] = g(b, c)$. Using the general forms for the functions f and g defined in Eq (9), we take $W[S : a] = f(a)$ and $W[S : b \,|\, a] = g(a, b)$. With the specific forms of these choices in Eq (10), we find that Eq (38) becomes

$$L(S|b, a) = L(S) + \alpha\big(\tfrac{1}{2}b[1 + \exp[\tfrac{1}{2}ba(1 + \exp(a))]]\big) + (1 - \alpha)\big(\tfrac{1}{2}a[1 + \exp(ab)]\big), \tag{39}$$

and the posterior probability of an action potential is

$$P(Z = 1|b, a) \equiv P(S|b, a) = \frac{1}{1 + \exp(-L(S|b, a))}. \tag{40}$$

Somatic action potential dependent on an additive combination of basal and apical input

Finally, we include the case in which the basal and the apical inputs combine in an additive manner. This means that there are no interaction terms involving a combination of the apical and basal inputs. To model this probabilistically, we assume that the apical input, A, and the basal input, B, are conditionally independent given S and also given $\bar{S}$, so that

$$p(a, b|S) = p(a|S) \times p(b|S), \quad \text{and} \quad p(a, b|\bar{S}) = p(a|\bar{S}) \times p(b|\bar{S}).$$

Given this conditional independence, the posterior log odds simplifies and it has the following form

$$L(S|b, a) = L(S) + W[S : b] + W[S : a],$$

where we see that the weight of evidence terms involving a combination of a and b are no longer present.

We define $W[S : b]$ to be $f(b)$ and $W[S : a]$ to be $f(a)$, and taking the specific choices in (10), we have that

$$L(S|a, b) = L(S) + b + a.$$

Thus the posterior probability in favor of an AP in this case is

$$P(Z = 1|b, a) \equiv P(S|b, a) = \frac{1}{1 + \exp\left(L(S) + a + b\right)}. \tag{41}$$

Acknowledgments

References

[1] Larkum M. A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. Trends in Neurosciences 2013;36(3):141 - 151.

[2] Llinás RR. I of the vortex: From Neurons to Self. MIT press Cambridge, MA; 2001.

[3] Rubio-Garrido P, Perez-de-Manzo F, Porrero C, Galazo MJ, Clasca F. Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. Cereb. Cortex 2009;19: 2380-2395.

[4] LaBerge D. Sustained attention and apical dendrite activity in recurrent circuits. Brain Res. Rev. 2005;50: 86-99. doi: 10.1016/j.brainresrev.2005.04.004

[5] Phillips WA, Cognitive functions of intracellular mechanisms for contextual amplification. Brain and Cognition 2017;112:39-53.

[6] Anastassiou CA, Shai AS. Psyche, Signals and Systems. In: Buzsàki G, Christen Y, editors. Micro-, Meso- and Macro- Dynamics of the Brain. Cham (CH): Springer; 2016. pp. 107-156.

[7] Phillips W, Larkum ME, Harley CW, Silverstein SM. The effects of arousal on apical amplification and conscious state. Neuroscience of Consciousness 2016;1:1-13. https://doi.org/10.1093/nc/niw015.

[8] Takahashi N, Oertner TG, Hegemann P, Larkum ME. Active cortical dendrites modulate perception. Science 2016 Dec 23;354(6319):1587-1590. doi: 10.1126/science.aah6066. PMID 28008068.

[9] Bachmann T, Hudetz AG. It is time to combine the two main traditions in the research on the neural correlates of consciousness. Front. Psychol. 2014;5:940. doi: 10.3389/fpsyg.2014.00940

[10] Phillips WA, Bachmann T, Storm JF. Apical Function in Neocortical Pyramidal Cells: A Common Pathway by Which General Anesthetics Can Affect Mental State. Frontiers in Neural Circuits 2018; 12, Art. No. 50. https://doi.org/10.3389/fncir.2018.00050

[11] Stuart GJ Sackmann B. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature 1994; Jan 6;367(6458):69-72. PMID 8107777.

[12] Larkum ME, Kaiser KM, Sakmann B. Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. Proc Natl Acad Sci U S A 1999;96: 14600-14604. PMID: 10588751

[13] Larkum ME, Senn W, and Lüscher H-R. Top-down dendritic input increases the gain of layer 5 pyramidal neurons. Cerebral Cortex 2004;10: 1059-70.

[14] Fletcher LN, Williams SR. Neocortical topology governs the dendritic integrative capacity of layer 5 pyramidal neurons. Neuron 2019;101:76-90. https://doi.org/10.1016/j.neuron.2018.10.048

[15] Shai AS, Anastassiou CA, Larkum ME, Koch C. Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. PLoS Comput Biol 2015;11(3): e1004090. doi:10.1371/journal. pcbi.1004090

[16] Kay JW, Phillips WA. Contrasting information theoretic decompositions of modulatory and arithmetic interactions in neural information processing systems. arXiv. http://arxiv.org/abs/1803.05897

[17] Atkinson S Williams P. Postnatal development of dendritic synaptic integration in rat neocortical pyramidal neurons. J Neurophysiol 2009:102:735-751.

[18] Beaulieu-Laroche L, Toloza EHS, van der Goes M-S, Lafourcade M, Barnagian D, Williams ZM et al. Enhanced Dendritic Compartmentalization in Human Cortical Neurons. Cell 2018;175(3):643-651. DOI:https://doi.org/10.1016/j.cell.2018.08.045

[19] Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. JOSA A 2003;20(7):1434-1448.

[20] Kay JW, Phillips WA. Coherent infomax as a computational goal for neural systems. Bulletin of Mathematical Biology 2011;73(2):344 - 372.

[21] Kay J, Floreano D, Phillips WA. Contextually guided unsupervised learning using local multivariate binary processors. Neural Networks 1998;11;117-140.

[22] Körding KP, König P. Learning with two sites of synaptic integration. Network: Computation in Neural Systems 2000;11:1-15.

[23] Kusmierz L, Takuya IT, Toyoizumi T. Learning with three factors: modulating Hebbian plasticity with errors. Current Opinion in Neurobiology 2017;46:170-177. http://dx.doi.org/10.1016/j.conb.2017.08.020

[24] Zhang H, Maloney LT. Ubiquitous Log Odds: A Common Representation of Probability and Frequency Distortion in Perception, Action, and Cognition. Front Neurosci. 2012;6(1). doi: 10.3389/fnins.2012.00001

[25] Peirce CS. The probability of induction. Popular Science Monthly 1878;12:705-718.

[26] Wrinch D, Jeffreys H. On Certain Fundamental Principles of Scientific Inquiry. Philosophical Magazine 1921;42:369-390.

[27] Good IJ. Studies in the history of probability and statistics. XXXVII. A.M. Turing's statistical work in World War II. Biometrika 1979;66:393-396.

[28] Good IJ. Weight of Evidence: A Brief Survey. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM. editors. Bayesian Statistics 2. Elsevier Science Publishers B V. (North Holland); 1985. pp. 249-270.

[29] Gold JI, Shadlen MN. Banburismus and the Brain: Decoding Review of the Relationship between Sensory Stimuli, Decisions, and Reward. Neuron 2002; 36:299-308.

[30] Kay JW, Ince RAA, Dering B Phillips WA. Partial and Entropic Information Decompositions of a Neuronal Modulatory Interaction. Entropy 2017;19(11):560. doi:10.3390/e19110560.

[31] Suzuki M, Larkum ME. Dendritic calcium spikes are clearly detectable at the cortical surface. Nature Communications 2017;8, 276. doi: 10.1038/s41467-017100282-4.

[32] Larkum ME, Nevian T, Sandler M, Polsky A, Schiller J. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. Science 2009;325: 756-760. doi: 10.1126/science. 1171958 PMID 19661433.

[33] Siegel M, Körding KP, König P. Integrating Top-Down and Bottom-Up Sensory Processing by Somato-Dendritic Interactions. Journal of Computational Neurosceince 2000;8(2):161-173.

[34] Bertschinger N, Rauh J, Olbrich E, Jost J, Ay N. Quantifying Unique Information. Entropy 2014;16:2161 - 2183.

[35] Griffith V, Koch C. Quantifying synergistic mutual information. In: Guided self-organization: Inception. Emergence, complexity and computation 9. Berlin/Heidelberg: Springer, 2014, pp.159 - 190

[36] Harder M, Salge C, Polani D. Bivariate measure of redundant information. Phys. Rev. E 2013; 87. doi:10.1103/PhysRevE.87.012130.

[37] Williams PL, Beer RD. Nonnegative decomposition of multivariate information. arXiv 2010, arXiv:1004.2515. Available online: https://arxiv.org/abs/1004.2515 (accessed on 20 February 2019).

[38] Ince RAA. Measuring multivariate redundant information with pointwise common change in surprisal. Entropy 2017;19:318.

[39] James RG, Emenheiser J, Crutchfield JP. Unique Information via Dependency Constraints. arXiv 2018, arXiv:1709.06653. (accessed on 20 February 2019).

[40] James RG, Ellison CJ, Crutchfield JP. a Python package for discrete information theory. The Journal of Open Source Software 2018; 25:738. https://doi.org/10.21105/joss.00738.

[41] Murayama M, Pérez-Garci E, Nevian T, Bock T, Senn W, Larkum ME. Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. Nature 2009 Feb 26; 457: 1137-41.

[42] Friston K. The history of the future of the Bayesian brain. Neuroimage 2012; 62:1230-1233. doi: 10.1016/j.neuroimage.2011.10.004

[43] Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. Nature Neuroscience 2013;16;1170-1178.

[44] Häusser M, Mel B. Dendrites: bug or feature? Current opinion in neurobiology 2003;13(3):372-383.

[45] Kubota Y, Karube F, Nomura M and Kawaguchi Y. The Diversity of Cortical Inhibitory Synapses. Front. Neural Circuits 2016;10:27. doi: 10.3389/fncir.2016.00027

[46] Larkum ME. The yin and yang of cortical layer 1. Nature Neuroscience 2013;16:114-115.

[47] Wang XJ, Yang GR. A disinhibitory circuit motif and flexible information routing in the brain. Current Opinion in Neurobiology 2018;49:75-83. doi.org/10.1016/j.conb.2018.01.002

[48] Boldog E, Bakken TE, Hodge RD, Novotny M, Aevermann BD, Baka J, et al. Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. Nat. Neurosci. 2018;21:1185-1195. doi.org/10.1038/s41593-018-0205-2

[49] Zhu JJ. Maturation of layer 5 neocortical pyramidal neurons: amplifying salient layer 1 and layer 4 inputs by Ca^{2+} action potentials in adult rat tuft dendrites. J Physiol. 2000;526(3):571-87.

[50] Ibrahim LA, Mesik L, , Ji XY, Fang Q, Li HF, Li YT, Zingg B, et al. Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. Neuron 2016;89:1031-1045. doi: 10. 1016/j.neuron.2016.01.027

[51] Spratling MW, Johnson MH. A feedback model of visual attention. Journal of Cognitive Neuroscience 2004;16(2):219-37.

[52] Spratling MW, Johnson MH. A feedback model of perceptual learning and categorisation. Visual Cognition 2006;13:129-165.

[53] Phillips W, Clark A, Silverstein SM. On the functions, mechanisms, and malfunctions of intracortical contextual modulation. Neuroscience and Biobehavioral Reviews 2015;52:1-20. https://doi.org/10.1016/j.neubiorev.2015.02.010.

[54] Silverstein SM, Wibral M, Phillips W. Implications of Information Theory for Computational Modeling of Schizophrenia. Computational Psychiatry 2017;1:82-101. https://doi.org/10.1162/CPSY-a-00004.

[55] Hughes EG, Peng X, Gleichman AJ, Lai M, Zhou L, Tsou R, et al. Cellular and synaptic mechanisms of anti-NMDAR encephalitis. J. Neurosci. 2010;30:5866-5875.

[56] Kole MH, Brauer AU, Stuart GJ. Inherited cortical HCN1 channel loss amplifies dendritic calcium electrogenesis and burst firing in a rat absence epilepsy model. J Physiol. 2007;578:507-25.

[57] Granato A, Palmer LM, De Giorgio A, Tavian D, Larkum ME. Early exposure to alcohol leads to permanent impairment of dendritic excitability in neocortical pyramidal neurons. The Journal of Neuroscience 2012;32(4): 1377-1382. doi: https://doi.org/10.1523/JNEUROSCI.5520-11.2012

[58] De Giorgio A, Granato A. Reduced density of dendritic spines in pyramidal neurons of rats exposed to alcohol during early postnatal life. International Journal of Developmental Neuroscience 2015;41:74-79.

[59] Guerguiev J, Lillicrap TP, Richards BA. Towards deep learning with segregated dendrites. eLIFE 2017. doi: https://doi.org/10.7554/eLife.22901.001

[60] Richards BA, Lillicrap TP. Dendritic solutions and the credit assignment problem. Current Opinion in Neurobiology 2018;54:28-36. https://doi.org/10.1016/j.conb.2018.08.003

[61] Sacramento J, Costa RP, Bengio Y, Senn W. Dendritic cortical microcircuits approximate the backpropagation algorithm. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R. (Eds.), Advances in Neural Information Processing Systems 31 (NIPS 2018), Montréal, Canada.

[62] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis (Third Edition); CRC Press: Boca Raton, USA, 2013.

[63] Lunn D, Jackson C, Best N, Thomas, A, Spiegelhalter D. The BUGS Book: A Practical Introduction to Bayesian Analysis; CRC Press: Boca Raton, USA, 2013.

[64] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[65] Heinze G, Ploner M. logistf: Firth's Bias-Reduced Logistic Regression 2018. R package version 1.23. https://CRAN.R-project.org/package=logistf

[66] Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Statistics in Medicine 2002;21(16): 2409-2419.

[67] Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80(1): 27-38.

[68] Stan Development Team (2018). RStan: the R interface to Stan

[69] Lindley DV. On a measure of information provided by an experiment. Ann. Math. Stat. 1956;27:986-1005.

[70] Shannon CE. A Mathematical Theory of Communication. Bell Syst. Tech. J. 1948:27:379-423. doi:10.1002/j.1538-7305.1948.tb01338.x.

[71] Aitchison J, Kay JW. Principles, practice and performance in decision making in clinical medicine. In: White, D. J. and Bowen, K. C. (Eds.), The Role and Effectiveness of Theories of Decision in Practice 1975: 252-272. Hodder & Stoughton, London.

[72] Sun Y, Gomez F, Schmidhuber J. Planning to be surprised: optimal Bayesian exploration in dynamic environments, in Artificial General Intelligence (Mountain View, CA: Springer) 2011: 41-51. doi: 10.1007/978-3-642-22887-2-5.

[73] MacKay DJC. Information Theory, Inference and Learning Algorithms; Cambridge University Press: Cambridge, UK, 2003.

[74] Cover TM, Thomas JA. Elements of Information Theory. Wiley-Interscience: New York, NY, USA, 1991.

[75] McGill WJ. Multivariate Information Transmission. Psychometrika 1954; 19(2):97-116.

[76] Schneidman E, Bialek W, Berry MJ. Synergy, Redundancy, and Population Codes. J. Neurosci. 2003;23:11539–11553.

[77] Gat I, Tishby N. Synergy and redundancy among brain cells of behaving monkeys. In: Proceedings of the 1998 conference on Advances in neural information processing systems 2. MIT Press Cambridge, MA, USA, 1999. pp. 111-117.

[78] Timme N, Alford W, Flecker B, Beggs JM. Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. J. Comput. Neurosci. 2014; 36:119-140. doi:10.1007/s10827-013-0458-4.

[79] Wibral M, Finn C, Wollstadt P, Lizier JT, Priesemann V. Quantifying Information Modification in Developing Neural Networks via Partial Information Decomposition. Entropy 2017;19:494.

[80] Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG. A Statistical Framework for Neuroimaging Data Analysis Based on Mutual Information Estimated via a Gaussian Copula. Hum. Brain Mapp. 2017;38:1541-1573.

[81] Park H, Ince RAA, Schyns PG, Thut G, Gross J. Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. PLOS Biology 2018;16(8), e2006558. doi:10.1371/journal.pbio.2006558. PMID 30080855

[82] Lizier JT, Bertschinger N, Jost J, Wibral M. Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. Entropy 2018;20(4):307. https://doi.org/10.3390/e20040307

[83] Tantau T. Graph Drawing in TikZ. In: Proceedings of the 20th International Conference on Graph Drawing. Springer-Verlag, Berlin: 2013. pp. 517-528.