

# WHAT CAN NEURONS DO FOR THEIR BRAIN? COMMUNICATE SELECTIVITY WITH BURSTS

DAVID BALDUZZI AND GIULIO TONONI

**ABSTRACT.** Neurons deep in cortex interact with the environment extremely indirectly; the spikes they receive and produce are pre- and post-processed by millions of other neurons. This paper proposes two information-theoretic constraints guiding the production of spikes, that help ensure bursting activity deep in cortex relates meaningfully to events in the environment. First, neurons should emphasize selective responses with bursts. Second, neurons should propagate selective inputs by burst-firing in response to them. We show the constraints are necessary for bursts to dominate information-transfer within cortex, thereby providing a substrate allowing neurons to distribute credit amongst themselves. Finally, since synaptic plasticity degrades the ability of neurons to burst selectively, we argue that homeostatic regulation of synaptic weights is necessary, and that it is best performed offline during sleep. selectivity and synaptic plasticity and information theory and credit assignment

## 1. INTRODUCTION

To survive an organism must choose favorable actions in a great variety of situations. This responsibility falls primarily on the central nervous system and in particular on its millions or billions of neurons. When deciding how to act neurons face a fundamental problem. A typical neuron does not interface directly with the environment, but only with other neurons. All it can “see” is a large number of input patterns, constantly changing, on its thousands of synapses, but what these input patterns may mean or represent is unknown to it. Worse, all a neuron can do, in essence, is choose just one of three actions: stay silent, spike once, or burst.<sup>1</sup> In other words, a neuron is extraordinarily “stupid” [1]: it does not know what its inputs mean, it does not know what it is communicating, to whom, and for what purpose, and in any case it has very little to say. How, then, can neurons possibly act in the interest of their brain?

A large body of evidence suggests neurons learn by modifying their synapses according to the distribution of pre- and postsynaptic spikes, modulated by chemical signals such as dopamine and noradrenaline [2–5]. Spikes play a privileged role in most models of neuronal learning: the distribution of spikes (or inter-spike intervals) determines when synapses are modified. For example in Hebbian learning synaptic plasticity is a function of correlations between spiking activity, whereas in spike-timing dependent plasticity (STDP) the precise timing of pre- and postsynaptic spikes determines whether synapses are potentiated or depotentiated [6–12]. Spikes therefore seem to provide a mechanism the brain uses to assign distribute credit and blame amongst neurons: synapses that transmit many spikes are proportionally potentiated or depotentiated, suggesting that spiking neurons and synapses are credited or blamed for good or bad outcomes experienced by the organism. Using bursts to assign credit and blame only makes sense if bursts contribute *actively* to total brain activity, in contrast to silent (or near silent) neurons, which form a more *passive* background.

This paper makes two main contributions. First, it information-theoretically characterizes the distinction between bursting foreground and silent background in terms of selectivity (quantified as

---

<sup>1</sup>We use the term “burst” loosely to indicate many spikes emitted in a short time.

*effective information*). Second, it proposes that neurons *communicate selectivity with bursts*. By this we mean: (i) neurons should use bursts to *emphasize* outputs that depend selectively on their inputs, and few or no spikes for outputs that depend vaguely on their inputs and (ii) neurons should *propagate* selective inputs by responding to spiking inputs with spiking outputs. Bursts should be: (i) selective and (ii) impactful.

In the results we show that highly selective outputs are responsible for almost all of the information transferred by a neuron. Moreover, we show that communicating selectivity satisfies a necessary condition for ensuring that selectivity is preserved by composite channels (such as pairs of neurons or neuronal populations). We then consider the implications of communicating selectivity for credit assignment and learning, showing that (de)potentiating synapses in response to selective outputs (i.e. bursts) yields finer control over how neurons deep in cortex adapt to sensory stimuli, since selective responses are more traceable.

Finally, we discuss how communicating selectivity may be enforced. Since synaptic strengths change constantly, ensuring bursts are selective requires ongoing effort. We argue that sleep, when the brain is offline and its activity is not task-dependent, provides an ideal time to align bursts with effective information and balance the relationship between input and output spikes.

**1.1. Related work.** Many models of learning and inference in distributed systems have been developed, starting perhaps with Selfridge’s Pandemonium of “shrieking demons” [13]. Recent approaches have focused on Bayesian models [14–17] where, typically, the number of spikes outputted by a neuron or the likelihood of a neuron outputting spikes corresponds to the probability of some event. Our approach is complementary to these since, after imposing the two constraints required for communicating selectivity, neurons have many remaining degrees of freedom regarding when they should spike. Neurons are free to use their spikes to predict neuronal or external events, so long as the events they focus on are *specific*.

Our work builds on observations that cortical representations of sensory inputs are *sparse* [18, 19]. Indeed communicating selectivity is a necessary condition for bursts to be sparse in cortex. However, rather than focus on sparsity at the population level, we investigate the more basic notion of selectivity, which is a local (specific to individual neurons) information-theoretic requirement for global sparsity.

## 2. METHODS

Neurons share the same repertoire of outputs – spiketrains – but differ in how they categorize their inputs. The most basic fact about a category is how sharp or *selective* it is: the fraction of inputs it contains. This section introduces effective information as a tool for quantifying selectivity.

We model neurons as abstract elements with finite alphabets of inputs and outputs (or situations and actions) denoted by  $S$  and  $A$  respectively. The probability that element  $n_k$  outputs  $a \in A$  in response to input  $s \in S$  is  $p_k(a|s)$ . Time is discretized into bins of fixed length which we leave unspecified (somewhere between 10 and 100ms). The input and output alphabets consist of patterns of 0s and 1s corresponding to silences and spikes.

**2.1. Quantifying selectivity.** The information generated when an element produces an output is quantified following prior work [20, 21]. Let the *potential repertoire*  $p_{unif}(S)$  be the set of potential inputs equipped with the uniform distribution.

The *actual repertoire*  $\hat{p}_k(S|a)$  of inputs that cause (lead to)  $a$  is computed by applying Bayes’ rule

$$(1) \quad \hat{p}_k(s|a) := \frac{p_k(a|do(s)) \cdot p_{unif}(s)}{p(a)},$$

where  $p(a) = \sum_s p_k(a|do(s)) \cdot p_{unif}(s)$ . The  $do(-)$  notation refers to Pearl’s calculus for working with causal interventions [22]. It stipulates that  $p_k(a|do(s))$  is computed by imposing input  $s$  onto

neuron  $p_k$  and observing the distribution of responses. Thus, the actual repertoire is computed based on interventions rather than observations – which is why we denote it by  $\hat{p}$  rather than  $p$ .<sup>2</sup>

Intuitively, an action is selective if it is chosen in response to few out of a large set of potential inputs. Formally, the *effective information* generated by an action is the Kullback-Leibler divergence between the actual and potential repertoires:

$$(2) \quad ei(S \xrightarrow{n_k} a) = H[\hat{p}(S|a) \parallel p_{unif}(S)].$$

Kullback-Leibler divergence  $H[p||q] = \sum_i p_i \log_2 \frac{p_i}{q_i}$  is non-negative, and is zero if and only if  $p = q$ . Effective information lies in range  $[0, n]$ , where  $n = \log_2(\# \text{ inputs in } S)$ . The effective information a neuron generates when it outputs  $a$  is high if few inputs cause (lead to) the neuron choosing that output. Conversely, effective information is low if output  $a$  is chosen for a large fraction of potential inputs.

Effective information is an action-specific quantity, unlike mutual information. The expectation of effective information  $\sum_{a \in A} p(a) ei(S \rightarrow a)$  is the mutual information  $I(S_{unif}; A)$  where inputs are given the uniform distribution.<sup>3</sup>

**Deterministic elements.** The above treatment simplifies considerably for a deterministic function  $f : S \rightarrow A$ . Define Markov matrix

$$p_f(a|s) = \begin{cases} 1 & \text{if } f(s) = a \\ 0 & \text{else.} \end{cases}$$

The actual repertoire is then

$$\hat{p}_f(s|a) = \begin{cases} \frac{1}{|f^{-1}(a)|} & \text{if } f(s) = a \\ 0 & \text{else.} \end{cases}$$

The support (the set of inputs with  $p > 0$ ) of the actual repertoire is the set  $f^{-1}(a)$  of inputs that function  $f$  sends to output  $a$ . Alternatively, we can describe  $f^{-1}(a)$  as a category implicitly defined by the function  $f$ , since  $f$  assigns the same label  $a$  to all elements of the pre-image.

The effective information generated by a deterministic function is

$$(3) \quad ei\left(S \xrightarrow{f} a\right) = \log_2 \frac{|S|}{|f^{-1}(a)|} = \log_2 \frac{(\text{total } \# \text{ of inputs})}{(\# \text{ of inputs causing } a)}.$$

Thus, an action  $a$  by function  $f$  is *selective* if it specifies a small category  $f^{-1}(a)$  in a large state space  $S$ . Conversely, the larger  $f^{-1}(a)$  is relative to the repertoire of potential inputs, the vaguer the action.

Fig. 1 shows how a firing and silent AND-gate categorizes its inputs and generates information. Of particular importance is “tracing back”. When the AND-gate fires, it specifies a unique cause: input 11. On the other hand, when the AND-gate is silent the specification is more vague: the input could have been any of 00, 01 or 10. A firing AND-gate thus specifies its input more sharply than a silent AND-gate.

To simplify the exposition we consider deterministic elements in the remainder of the paper – except for section §3.1.

<sup>2</sup>In this paper, where we consider effective information generated by a single neuron and the mechanism is known, the *do*-calculus is redundant. We retain the notation to maintain consistency with prior and future work, where applying causal interventions is necessary.

<sup>3</sup>We use the uniform distribution since, as shown in Eq. (3), it precisely captures the fraction of inputs causing an output.

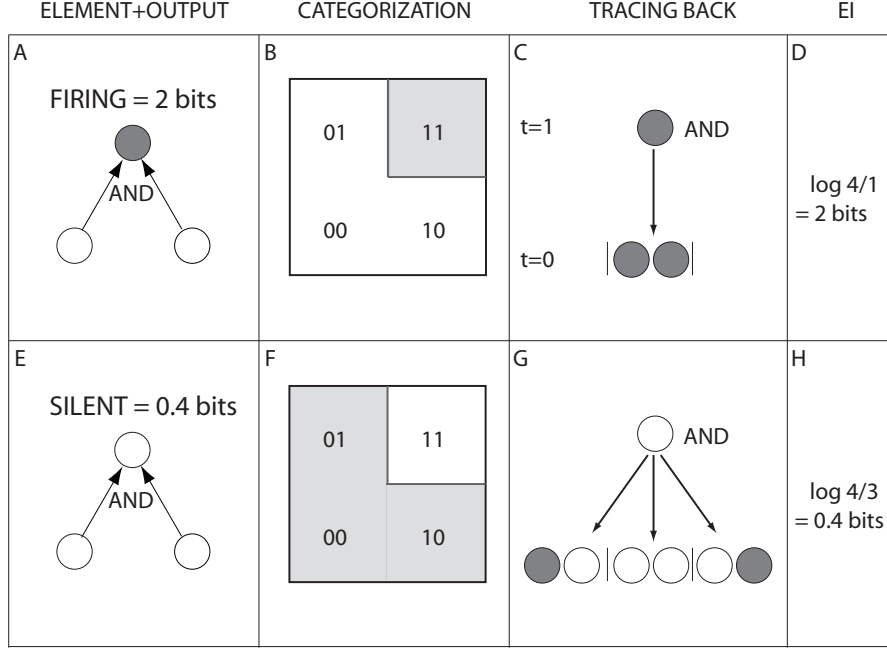


FIGURE 1. **Categorizing inputs.** (AB): An AND-gate fires iff it receives input 11. (C): The gate firing at time  $t = 1$  points to a single cause at time  $t = 0$ : 11. (D): The firing gate generates 2 bits of effective information. (EF): The gate is silent iff its input is one of 00, 01, 10. (G): The silent AND-gate at time  $t = 1$  points to three potential causes at time  $t = 0$  that are indistinguishable to it. (H): Silence is vaguer than firing, generating 0.4 bits of effective information.

**2.2. Communicating selectivity.** We propose that neurons communicate the selectivity of their outputs by: (i) emphasizing *selective outputs* with bursts and (ii) propagating selectivity by bursting in response to *selective inputs*. Fig. 2 shows examples of communicating and *not* communicating selectivity.

**Constraint 1** (emphasize selectivity).

*A neuron emphasizes selective outputs with spikes if for each output  $a \in A$ ,*

$$\text{firing-rate}(a) \approx g(ei(S \rightarrow a))$$

*for  $g$  some monotonically increasing function such as the identity, exponential or sigmoid function.*

Neurons have extremely asymmetric input/output ratios: their inputs alphabets are vastly larger than their output alphabets. They can therefore categorize inputs asymmetrically, so that bursts are much more selective than responses containing few or no spikes. This fits experimental evidence showing that neurons burst selectively for highly specific stimuli whereas the vast majority of inputs elicit little or no spiking response [23].

**Constraint 2** (propagate selectivity).

*A neuron propagates selectivity to the extent that its input and output firing rates are aligned on average,*

$$\text{avg firing-rate}(\text{inputs}) \approx \text{firing-rate}(\text{output}).$$

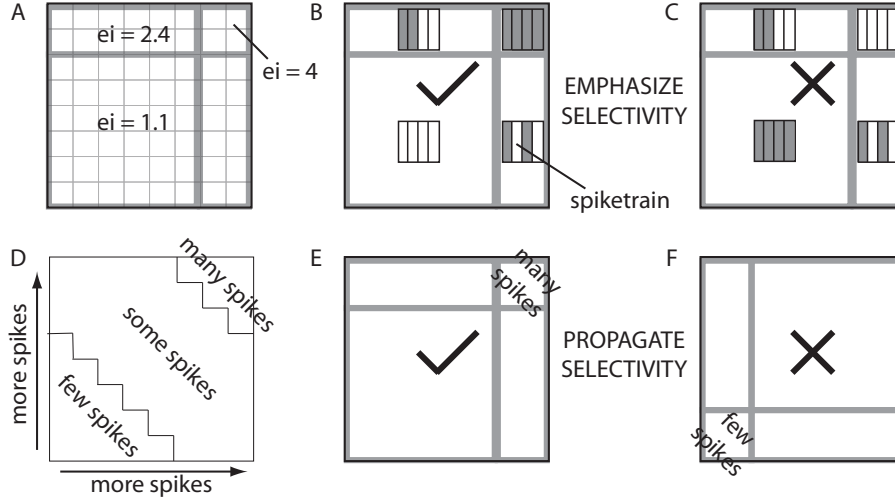


FIGURE 2. **Communicating selectivity.** An element (not shown) has 6 binary input lines, and so can receive  $2^6 = 64$  input patterns. (A): The element categorizes the 64 firing patterns into 4 categories. (B): To emphasize selectivity, the element assigns more spikes to sharper categories. (C): Selectivity is *not* emphasized if vague categories receive more spikes than sharp categories. (D): The 64 input patterns are arranged so that the number of spikes increases from the bottom left corner to the top right. (E): To propagate selectivity the element assigns more output spikes to inputs containing many spikes and conversely. (F): An element fails to propagate selectivity if it assigns many output spikes to inputs containing few spikes.

Most synapses are excitatory, so that presynaptic spikes tend to cause postsynaptic spikes. Propagating selectivity requires further that the ratio of input to output spikes is actively regulated.

Neurons can fail to propagate selectivity in two ways. First, they may overspike, bursting when they receive few spikes. In this case, they will bias the brain towards epileptic seizures. Second, they may underspike, responding to many bursting inputs with silence. In this case, they will tend to ignore important events in the brain and the environment, possibly to the detriment of the organism.

The two constraints are complementary. Emphasizing selectivity requires that sharp categories correspond to bursts. Propagating selectivity requires that neurons burst when they receive many spikes. Together they imply that neurons conserve effective information – i.e. neurons treat their input-spikes, on average, as selectively as the neurons that generated them:

$$(4) \quad \frac{1}{|\{j : j \rightarrow k\}|} \sum_{j \rightarrow k} g\left(ei\left(S_j \xrightarrow{n_j} a_j\right)\right) \approx g\left(ei\left(S_k \xrightarrow{n_k} a_k\right)\right) \text{ for each element } n_k.$$

We highlight some features of the proposal. First, treating spikes selectively depends only on the neuron: its mechanism and its output. Selectivity can be computed locally. Second, effective information maps naturally onto firing rates since it is a non-negative scalar.<sup>4</sup> Third, emphasizing selectivity conserves energy since costly outputs, which generate high effective information, are chosen for a small fraction of potential inputs. Spikes and, more generally, synaptic activity caused by spikes,

<sup>4</sup>We do not propose a rate code in the sense that only firing rates are meaningful. Rather, we suggest that firing rates have a specific, standardized meaning. Additional information can be encoded in the precise timing of spikes.

is metabolically expensive, accounting for much of the brain’s large energy budget, whereas silence is less expensive [24]. If spikes are expensive, they should be used as little as possible.

The choice of function  $g$  in Eq. (1) is not information-theoretically crucial, so long as it is monotonically increasing and consistently applied across the system. The relationship between the number of spikes received and produced by neuronal populations has implications for how activity propagates through cortex [25, 26], which may impose additional constraints on the choice of function.

Inhibitory neurons do not propagate selectivity since inhibitory (GABA) synapses suppress postsynaptic postsynaptic spiking activity: increasing inhibitory input decreases spiking output. Inhibitory neurons thus operate according to different principles than those proposed here (although note they do appear to emphasize selectivity) and are deferred to future work.

**Implications for the neural code.** It remains unclear what neurons encode into their spiketrains or how they decode and usefully exploit the information they find there. Certainly, propagating symbols – such as, for example, distributed patterns of spiking activity – is difficult because of the asymmetric ratio between neuronal inputs and outputs: thousands of input wires are compressed into a single output wire.

What neurons can easily do is burst when they receive many bursting inputs (Constraint 2) and ensure that bursts are selective (Constraint 1). In this way, neurons ensure that, on average, they produce selective outputs when they receive selective inputs. Additional information may be encoded into the precise timing of neuronal firing as an overlay on top of the more basic firing rate code advocated here.

Thus, even if neurons have difficulty propagating symbols, *they can at least propagate selectivity*, a property of symbols that we argue below is important for credit assignment.

### 3. RESULTS

**3.1. Selectivity and information transfer.** This subsection presents a rigorous justification for the constraints proposed above. It is well known that the mutual information quantifies the amount of information transferrable across a channel [27]. It is useful to view neurons, or populations of neurons, as information-theoretic channels in the brain. The first result then shows that the effective information generated by highly selective outputs approximates mutual information up to first order:

**Theorem 1** (selective outputs dominate information transfer).

*Suppose outputs by neuron  $n_k$  are grouped under two labels,  $a_0$  and  $a_1$ , and  $p_k(a_1) \ll 1$ . Then the total information transferred by the neuron is approximated to 1st order by the information it transfers using  $a_1$  alone:*

$$(5) \quad I_k(S_{unif}; A) = p_k(a_1) \cdot ei(S \xrightarrow{n_k} a_1) + O(p_k(a_1)^2).$$

*Proof.* See Appendix of [28]. □

We apply the theorem by grouping together highly selective outputs (bursts) as  $a_1$  and other outputs as  $a_0$ . It follows that almost all of the information transferred by a neuron is carried by its selective outputs. Theorem 1 and Constraint 1 jointly provide the first step towards an explanation of why synaptic plasticity depends so strongly on pre- and post-synaptic spikes – because, if neurons communicate selectivity, then it is bursts that carry signals and are therefore useful for learning. Indeed, a recent study has shown that hippocampal neurons rely heavily on bursts to transfer information to downstream brain structures and to encode memories when learning [29].

Importantly, by using the uniform distribution we quantify the information transferred by the neuron itself. Using a different prior would inject additional information that is not available to the neuron.

The second result provides an upper bound on the effective information generated by composite channels:

**Theorem 2** (effective information for composite channels).

Let channels  $n_1$  and  $n_2$  have Markov matrices  $p_1(y|x)$  and  $p_2(z|y)$  on finite sets  $X$ ,  $Y$  and  $Z$ . Let  $p_{12}(z|x) = \sum_{y \in Y} p_2(z|y) \cdot p_1(y|x)$  denote the composite channel. Then

$$(6) \quad ei(X \xrightarrow{n_2 \circ n_1} z) \leq \sum_{y \in Y} c(y|z) \cdot ei(X \xrightarrow{n_1} y), \text{ where } c(y|z) := p_2(z|y) \cdot \frac{p_1(y)}{p_{12}(z)}.$$

*Proof.* Appendix A1. □

Theorem 2 provides a *necessary* condition for the composite of two channels,  $n_2 \circ n_1$ , to generate high effective information when outputting  $z$ : it is necessary that the mass of probability distribution  $c(y|z)$  is concentrated on outputs  $y$  with high effective information under channel  $n_1$ . In the cortex, channels  $n_1$  and  $n_2$  could correspond to two (populations of) neurons. For output  $z$  to transfer a large amount of information through the composite channel it is necessary that

- (\*) outputs  $y$  with high effective information  $ei(n_1, y)$  under channel  $n_1$ ,  
 have a high probability  $p_2(y|z)$  of causing output  $z$  under channel  $n_2$   
 – and conversely, to keep  $p_{12}(z)$  low.

Communicating selectivity provides a method for ensuring (\*) holds:

- (1) outputs with high effective information are tagged with bursts (emphasizing selectivity) and are therefore
- (2) more likely to cause bursts whereas, conversely, vague outputs are tagged with few spikes and are less likely to cause bursts (propagating selectivity).

**Example 1** (application to bursting neurons).

Consider when  $z = \text{burst}_2$  is a particular burst by neuron  $n_2$ . Then the effective information generated about  $X$  by channel  $n_2 \circ n_1$ ,

$$ei(X \xrightarrow{n_2 \circ n_1} \text{burst}_2) \leq \sum_{y \in Y} c(\text{burst}_2|y) \cdot ei(X \xrightarrow{n_1} y),$$

is bounded by the average of the effective information generated by outputs  $y$  of neuron  $n_1$  that cause  $\text{burst}_2$ . Note that  $c(\text{burst}_2|y)$  quantifies the probability that  $\text{burst}_2$  was caused by  $y$ . Thus, for the left-hand side of the equation to be high, it is necessary that the  $y \in Y$  causing  $\text{burst}_2$  have high effective information – i.e. are themselves bursts.

**3.2. Credit assignment.** Figuring out how to assign credit is a problem faced by any distributed system. When a hungry mouse reaches for cheese only a small fraction of neurons are actively involved. Most neurons are specialized for unrelated activities. It follows that not all neurons in the brain should be rewarded when the mouse sates its hunger. This section describes how communicating selectivity helps neurons distribute credit amongst themselves by providing a way to identify which neurons and synapses actively contributed to global outcomes.

**Explanatory power.** Effective information quantifies how well an outputs fits an input. It can be shown that

$$(7) \quad \begin{aligned} ei(S \xrightarrow{n_k} a) &= \frac{H(p_{unif}(S))}{\text{(total bits available)}} - \frac{H(\hat{p}_k(S|a))}{\text{(bits indistinguishable to } n_k)} \\ &= \left( \frac{\# \text{ bits output } a \text{ explains}}{\text{(total bits available)}} \right), \end{aligned}$$

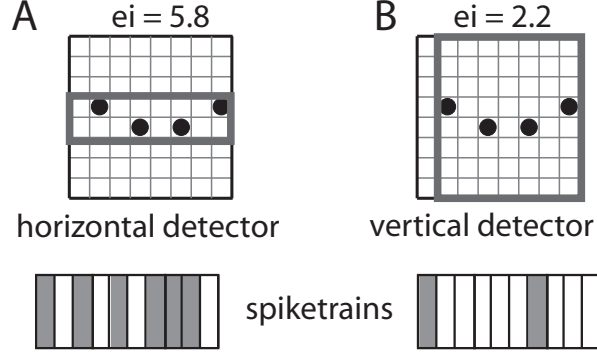


FIGURE 3. **Explanatory power.** (AB): Two elements receive the same configuration of dots. The first fits horizontal rectangles to configurations; the second fits vertical rectangles. The configuration is one of 11,760 fitting in a horizontal rectangle of height 2, and one of 139,040 fitting in a vertical rectangle of width 7. Since the elements emphasize selectivity, the first element signals its superior explanation with more spikes than the second element.

where  $H(p) = -\sum_i p_i \log_2 p_i$  is Shannon entropy. Outputs with higher effective information have more explanatory power. Alternatively, they fit the input data tighter.

It is useful to reinterpret the results above in terms of explanatory power. Theorem 1 says that outputs with high explanatory power account for most of the information transferred by an element. Theorem 2 provides a necessary condition for conserving explanatory power when composing elements.

Fig. 3 illustrates explanatory power using two elements loosely modeled on orientation columns in visual cortex. Inputs are configurations of dots on an  $8 \times 8$  pixel grid. Element  $n_1$  categorizes configurations by height whereas  $n_2$  categorizes configurations by width. The configuration in Fig. 3 has height 2 and width 7. The horizontal detector generates  $ei(n_1, 2) = 5.8$  bits and the vertical detector generates  $ei(n_2, 7) = 2.2$  bits, see appendix for computations. Element  $n_1$  generates more effective information since fewer configurations fit in a  $2 \times 8$  rectangle than a  $8 \times 7$  rectangle: the horizontal explanation fits the data better than the vertical explanation.

- If the elements communicate selectivity then  $n_1$  will produce more spikes than  $n_2$ . Thus, goodness-of-fit is signaled with spikes.
- Moreover, if the elements connect to motor neurons that communicate selectivity, then  $n_1$  yields a stronger motor response than  $n_2$ . Thus, neurons that fit their input data better exert more control on downstream activity than those that do not.

If the organism subsequently experiences a negative outcome, the motor neuron responding to  $n_1$ 's burst is more to blame than the motor neuron not responding to  $n_2$ 's few spikes. Communicating selectivity thus provides a minimal substrate for credit assignment.

**Traceability.** The spiking input patterns neurons received by neurons deep in cortex have been preprocessed by millions of other neurons. It is these spiking patterns that directly determine which synapses are modified – rather than what is actually going on in the environment. It is therefore crucial that spikes deep in the brain relate meaningfully to events in the environment. Traceability refers to the extent to which spikes relate to external events. Theorems 1 and 2 provide an abstract framework for understanding how spikes deep in a system relate to external events. Below, we provide a concrete explication of the mechanics of traceability in the simplest possible case. For detailed computations see Appendix A3.



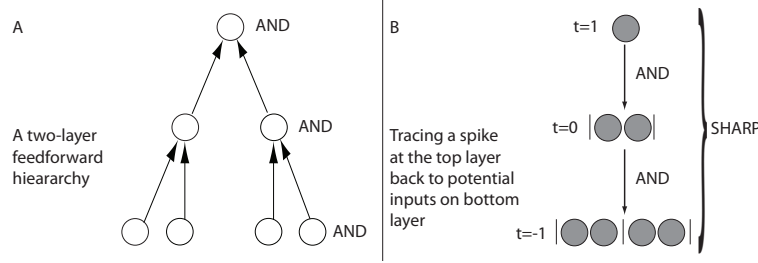


FIGURE 4. **Communicating selectivity**  $\implies$  **traceability**. What does the spiking of the top element imply about the input two layers down? (A): A 2-layer hierarchy of AND-gates. (B): The system of AND-gates both emphasizes and propagates selectivity, so spikes can be traced back through multiple layers and time-steps. The input must have been 1111.

We consider AND, OR and NOR gates. Inputs and outputs are 0s and 1s, where 0s corresponds to silence and 1s to spiking (or bursting).

	AND	OR	NOR	
Emphasize	✓	×	✓	(spikes $\rightarrow$ high $ei$ )
Propagate	✓	✓	×	(spikes-in $\rightarrow$ spikes-out)

Figs 4 and 5 consider how spikes relate to inputs two layers away in a minimal model. Fig. 4 shows a system of AND-gates. These emphasize and propagate selectivity since (i) spikes generate 2 bits of information whereas silences generate 0.4 bits and (ii) AND-gates only spike when they receive 1 spike on *each* wire.

The spike at the top of the system directly relates to a specific environmental event since it implies the two gates below both spiked, which in turn implies the input was 1111. Fig. 5B considers the same setup with AND-gates replaced by OR-gates. The spike at the top now implies little about the input two layers further down. Finally, Fig. 5C considers NOR-gates. Here, spikes trace back one step but not two, since NOR-gates emphasize selectivity but do not propagate it. In panels BC many different inputs at the bottom layer cause a spiking response at the top, so the spike at the top relates nonspecifically or vaguely to the environment.

Although AND-gates are much simpler than neurons, they share two critical features. First, AND-gates spike for less than half their inputs ( $ei(\text{spike}) > 1$ ). Although neurons do not burst for a unique pattern, they appear to burst highly selectively (for far fewer than half of their physiological inputs,  $ei(\text{spike}) \gg 1$ ), typically in response to some preferred, biologically meaningful stimulus such as an edge, face, or particular person. Second, AND-gates only spike when they receive spikes. Here again, the result generalizes to excitatory neurons which require spiking activity on many of their synapses before they burst. These are the only two features neurons need share with AND-gates for bursts to be more traceable than silences.

Communicating selectivity ensures bursts relate more directly to environmental inputs and motor outputs than silences. Emphasizing selectivity brings bursts into the foreground (the silent background does not relate to the specifics of the current situation) and propagating selectivity ensures the bursting foreground plays the major role in determining what neurons do next (bursts cause bursts). Bursts therefore provide a local signature of responsibility so that bursting neurons and synapses should receive credit or blame.

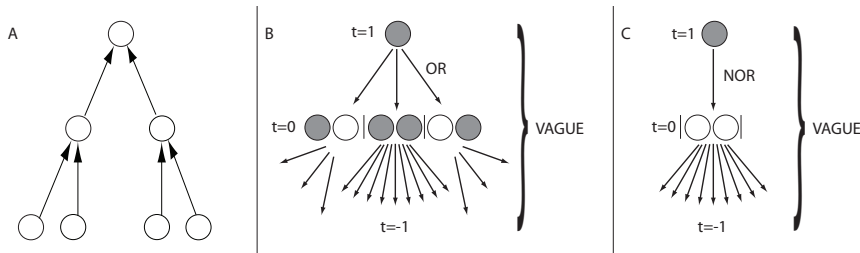


FIGURE 5. **Not communicating selectivity  $\implies$  no traceability.** (A): System as in Figure 4 with AND-gates replaced by ORs and NORs respectively. (B): Spikes are not selective in a system of OR-gates since they do not emphasize selectivity. A spike at the top of the network could have been caused by many input patterns 2 layers down. (C): Spikes are selective in a system of NOR gates, but selectivity is not propagated (silences cause spikes, rather than spikes causing spikes). Consequently, although spikes trace back a single step, after two steps they become much less traceable.

This fits a large body of experimental evidence showing that spikes and spike-timing play decisive roles in synaptic plasticity [2, 4, 30]. Communicating selectivity with bursts constrains comparatively few of a neuron’s thousands of degrees of freedom. In essence it forces them to specialize on a small fraction of their (bursting) inputs. In this way, the brain ensures spikes deep in the brain meaningfully relate to activity in the environment. Our approach is thus complementary to many existing models of synaptic plasticity [6, 10, 11].

**Selectivity determines the granularity of synaptic plasticity.** Finally, we consider how traceability affects the “spatial resolution” of synaptic plasticity. Most Hebbian and spike-timing dependent learning rules entail that neurons potentiate or depotentiate their spiking synapses shortly before or after postsynaptic spikes [3–6]. We illustrate the importance of selectivity by considering the *indirect* effects of spike-dependent learning. When a neuron modifies one of its synapses, it modifies its response to the sensory input that caused the synapse to spike – but also its response to all other sensory inputs causing the synapse to fire. The size of this set of sensory inputs is the *granularity* of the synaptic potentiation.

Fig. 6 considers granularity in 1080 randomly generated three-layer feedforward networks. Each neuron receives connections from 20 other neurons, with synaptic weights sampled from the uniform distribution and then renormalized to average values shown on the  $x$ -axis (54 values ranging from 0.04 to 0.0559). Effective information was approximately computed by sampling from 10,000 input patterns on layer  $L_1$ .

As synaptic weights were increased, the  $ei(L_1 \xrightarrow{net} l_3)$  generated by layer  $L_3$  about layer  $L_1$  progressively decreased: selectivity was lost, see  $\times$ ’s. Moreover, see  $\bullet$ ’s, as effective information decreased, the number of sensory inputs on  $L_1$  resulting in a given plastic event in  $L_3$  increased: synaptic potentiations in  $L_3$  treated more and more inputs at  $L_1$  as indistinguishable, resulting in less fine-grained learning.

To summarize: as selectivity decreases, the responses of neurons deep in cortex lump more and more sensory inputs together. For example, a neuron may modify itself after the organism is exposed to fire, but at the same time also change its response to many completely unrelated situations. Controlling the granularity of learning deep in cortex is therefore necessary. Since learning is driven by spikes, and

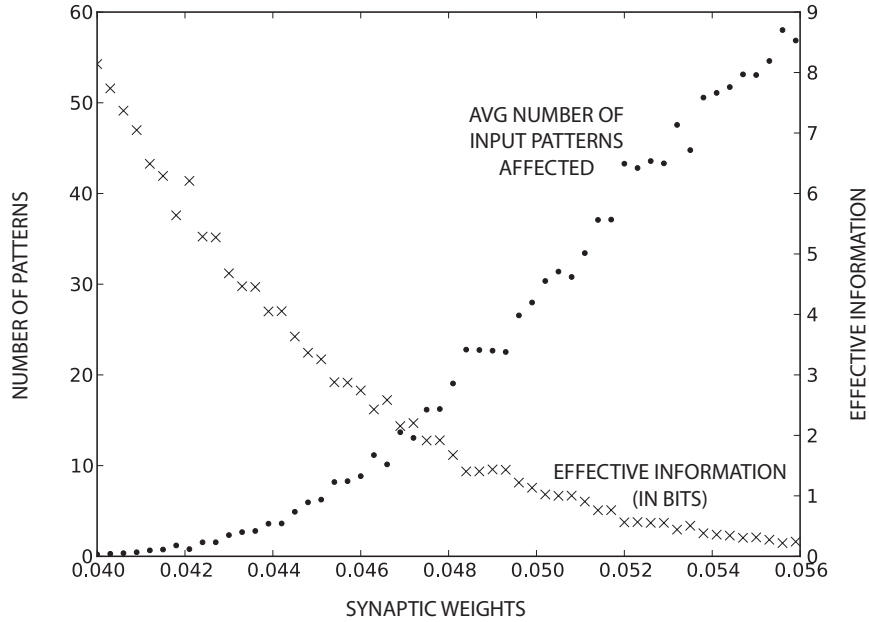


FIGURE 6. **Selectivity determines the granularity of synaptic plasticity.** Average synaptic weights plotted against effective informations ( $\times$ 's) and the average number of input patterns affected by individual synaptic modifications ( $\bullet$ 's).

even more so by bursts of spikes, we have argued that this can be done by controlling the selectivity of bursts.

Although we only considered feedforward networks, it is worth noting that recurrent networks can be viewed as feedforward networks when unfolded over time. The response of a recurrent network at time  $t = 3$  generates effective information about its state at time  $t = 2$  and also  $t = 1$ , so that the granularity of a system's response to its own prior state therefore depends on the selectivity of its responses.

**3.3. Plasticity and synaptic homeostasis.** Observed properties of cortical excitatory neurons – that (i) burst-firing is selective and (ii) the more spikes a neuron receives the more it produces, with average cortical firing stable over time – are compatible with the hypothesis that neurons communicate selectivity. However, neurons in cortex unceasingly modify the weights of existing synaptic contacts in response to a highly non-stationary environment [31]. It follows that active effort is necessary to ensure bursts remain selective as the brain is constantly rewires itself.

**Synaptic potentiation reduces the selectivity of bursts.** There is increasing evidence that, while a number of plasticity mechanisms in various brain regions can lead to both strengthening and weakening of synapses, overall synaptic strength tends to increase in the course of waking activities [32, 33]. Synaptic potentiation degrades the selectivity of bursts by increasing the number of inputs that cause burst-firing, see Fig. 7. Eventually, going to the logical – if not physiological – extreme, a fully potentiated neuron would fire constantly so that its spikes have no information-theoretic value at all. In practice, synapses would saturate long before this extreme, severely compromising learning.

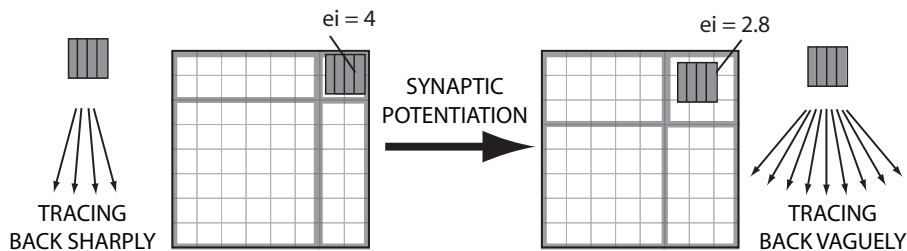


FIGURE 7. **Synaptic potentiation reduces the selectivity of bursts.** Increasing synaptic strengths causes elements to fire more frequently in more situations. Consequently, burst categories become larger (from 4 potential causes to 9) and bursts become less informative and harder to trace back.

Moreover, since spikes and excitatory post-synaptic potentials are metabolically expensive, a progressive increase in firing rates and connection strengths is costly and ultimately unsustainable. The brain consumes a disproportionate amount of the body's energy ( $\approx 15\%$ ), and it is estimated that up to 75% of the brain's budget goes to maintain synaptic activity [24]. Stronger synapses occupy more space, require more supplies, and may lead to cellular stress [34]. Consequently a system containing billions of plastic elements should regulate the relationship between inputs and outputs.

**Renormalization is best performed offline.** We argue that the selectivity of bursts is best regulated during sleep. Regulating selectivity requires computing effective information. This can be done by sampling inputs from the uniform distribution and counting how many inputs fall into each output category. In practice, neurons never receive uniformly distributed inputs. However, sampling from a large number of uncorrelated or weakly correlated firing patterns approximates effective information across physiologically relevant input patterns.

Sampling during wakefulness is problematic. The inputs a neuron receives while its organism engages in behavior form an extremely biased sample. For example, the inputs sampled by a motor neuron during a day spent performing mostly one kind of activity (say typing), provide biased estimates on the distribution of spiking activity. If synaptic strengths were downscaled during sleep using the same distribution over inputs that caused them to potentiate during the wakefulness, then downscaling would depotentiate exactly what was potentiated.

Synaptic renormalization is therefore best performed off-line, most notably during sleep, when neurons receive inputs uncoupled from the immediate needs dictated by environmental interactions [35, 36]. Indeed, a paramount fact about the sleeping brain is that it is spontaneously active, often at levels similar to those observed during wakefulness [37]. Moreover, this spontaneous activity seems to faithfully reflect the underlying connectivity at multiple levels [38], and to change as a function of experience, that is, in response to changes in synaptic strength [39]. During sleep, then, the cerebral cortex perturbs and samples from itself for many hours with many different firing patterns and, crucially, does so in a task-independent manner.

Experimental evidence suggests that average firing rates and net synaptic strength increase in both the cerebral cortex and the hippocampus during wakefulness [32, 33]. By contrast both average firing rates and net synaptic strength, as indexed by both molecular and electrophysiological markers, decrease after periods of sleep. During sleep many neuromodulators are released at their lowest level, and phasic, burst release is notably absent [40]. Low neuromodulation may, first of all, prevent the occurrence of synaptic potentiation when neural activity is decoupled from behavior (you would not want to learn your dreams) and, second, favor a net depression of synaptic strength to achieve an

overall renormalization. The occurrence of slow oscillations during NREM sleep, characterized by periods of activity (UP-states) followed by profound hyperpolarization (DOWN-states) every second or so, may also favor synaptic depression and renormalization [41].

#### 4. DISCUSSION

How does a population of individually stupid neurons make collectively intelligent decisions? This paper argues a necessary condition is that neurons burst selectively, so that bursting neurons take credit for their accomplishments and responsibility for their mistakes.

**What can a neuron do for its brain?** Loosely speaking, the brain’s goal is to act usefully in any situation. It is composed of hundreds of billions of neurons; its actions consist of their actions. It is up to each neuron to figure out how to act usefully based on local data, spikes, and global data, neuromodulatory signals. Since spikes are produced for neurons by neurons, it is plausible that there are constraints in place guiding their production and reception. We therefore investigate how selectively neurons categorize their inputs, focusing on spikes.

Neurons are specialized. Some respond preferentially to visual stimuli, others to auditory ones, still others to motor commands. Within the visual systems, some respond more to shapes and others to colors. Clearly, different neurons categorize inputs differently, and thus inputs that are “important” for one neuron may not be so for another. In any given situation, then, which neuronal outputs are important? From the point of view of the system as a whole it is necessary to emphasize important categorizations right now (e.g. there is probably a face) as opposed to currently unimportant categorizations (e.g. there is probably *not* a car). Moreover, it is advisable for other neurons (say in the planning and motor systems) to pay attention to important rather than to unimportant inputs when deciding and learning. How can this be achieved?

We proposed that neurons communicate – that is, emphasize and propagate – important or selective outputs using bursts. Neurons emphasize selectivity by mapping effective information onto spike trains so that the greater the number of spikes, the greater the selectivity of the output. Emphasizing selective outputs with spikes makes metabolic sense since spikes are energetically expensive, suggesting they should be used rarely, in response to highly specific (i.e. selective) inputs. Neurons propagate selectivity by outputting many spikes when they receive many input spikes. Bursts thus exert more control over downstream activity – and ultimately behavior.

We then showed that communicating selectivity highlights outputs with high explanatory power: burstiness reflects goodness-of-fit. The better a neuronal output fits its input, the more spikes it produces and the more impact it has on downstream neuronal activity. Neurons must learn how to act with only spikes and neuromodulators for guides – their interactions with the external environment are extraordinarily indirect. Decisions that lead to positive outcomes for the organism should be reinforced. However, in a system consisting of billions of neurons and trillions of synapses, it is not *a priori* obvious who is responsible for a successful outcome. Indeed, responsibility must be distributed across many neurons. But not all. Some neurons and synapses are more responsible than others. Communicating selectivity provides a simple way to track responsibility.

Although communicating selectivity is a constraint placed on neurons individually, its main implications are for their collective dynamics. In particular, if bursts are selective, it ensures that transient coalitions of bursting neurons are also selective, and therefore useful for learning, inference, and behavior.

**Testing the proposal.** Neurons are known to burst for specific stimuli such as faces and vertical or horizontal edges, suggesting they may emphasize selectivity. Further, excitatory neurons spike-for-spikes, suggesting they may propagate selectivity. However selectivity, as quantified by effective information, refers to the proportion of potential inputs that cause an output and not the specificity

of a neural response to stimuli in the environment, which is a system property and does not depend on any single neuron. Estimating the effective information generated by neuronal outputs requires manipulating the inputs received by a significant fraction of its thousands of synapses and observing the responses over a wide range of physiologically relevant inputs. Directly testing whether neurons communicate selectivity with bursts is thus technically challenging.

However, some implications of communicating selectivity are more accessible. First, it should be investigated to what extent communicating selectivity provides a useful substrate for learning and cognition; specifically for assigning credit. Spike-timing dependent plasticity has proven to be difficult to work with because of a tendency to learn to overspike, spiraling into epileptic seizures. A depotentiation bias is therefore necessary [7]. Introducing a sleep-phase where synaptic weights are downscaled to counteract overspiking opens up new computational possibilities for STDP during wakefulness.

Second, whether (and how) neurons regulate selectivity should be investigated. We have argued that homeostatic regulation of synaptic strengths is necessary and that such regulation is best performed during sleep. Thus, the hypothesis can be tested by investigating how synaptic strengths are modified during sleep and learning. A growing body of evidence suggests that synaptic strengths are downscaled during sleep [32, 33, 40, 42]. How closely this regulation ties to the selectivity of bursts remains to be seen.

**Acknowledgements.** We thank Michel Besserve for useful comments and Theorem 1. Supported in part by NIH Director’s Pioneer Award and Conte Center National Institute of Mental Health (P20MH077967) to GT, and by Defense Advanced Research Projects Agency, Defense Sciences Office (DSO), Program: Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE).

#### REFERENCES

- [1] Dennett DC: **Darwin’s “strange inversion of reasoning”**. *Proc Natl Acad Sci USA* 2009, **106 Suppl 1**:10061–10065.
- [2] Markram H, Lübke J, Frotscher M, Sakmann B: **Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs**. *Science* 1997, **275**(5297):213–5.
- [3] Abbott L, Nelson S: **Synaptic plasticity: taming the beast**. *Nature Neuroscience* 2000, **3**:1178–1183.
- [4] Caporale N, Dan Y: **Spike Timing-Dependent Plasticity: A Hebbian Learning Rule**. *Annu Rev Neurosci* 2008, **31**:25–46.
- [5] Pawlak V, Wickens JR, Kirkwood A, Kerr JND: **Timing is not everything: neuromodulation opens the STDP gate**. *Front. Syn. Neurosci* 2010, **2**(146).
- [6] Hebb DO: *Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons 1949.
- [7] Song S, Miller KD, Abbott LF: **Competitive Hebbian learning through spike-timing-dependent synaptic plasticity**. *Nature Neuroscience* 2000, **3**(9).
- [8] Song S, Abbott LF: **Cortical Development and Remapping through Spike Timing-Dependent Plasticity**. *Neuron* 2001, **32**.
- [9] Rao RP, Sejnowski TJ: **Spike-timing-dependent Hebbian plasticity as temporal difference learning**. *Neural Comput* 2001, **13**(10):2221–2237.
- [10] Seung HS: **Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission**. *Neuron* 2003, **40**(1063-1073).
- [11] Izhikevich E: **Solving the distal reward problem through linkage of STDP and dopamine signaling**. *Cerebral Cortex* 2007, **17**.
- [12] Legenstein R, Pecevski D, Maass W: **A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback**. *PLoS Comput Biol* 2008, **4**(10):e1000180.
- [13] Selfridge OG: **Pandemonium: a paradigm for learning**. In *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physics Laboratory* 1958.
- [14] Lee TS, Mumford D: **Hierarchical Bayesian inference in the visual cortex**. *J Opt Soc Am A* 2003, **20**:1434–1448.
- [15] Rao RPN: **Bayesian computation in recurrent neural circuits**. *Neural Comput* 2004, **16**:1–38.
- [16] Ma WJ, Beck JM, Latham PE, Pouget A: **Bayesian inference with probabilistic population codes**. *Nature Neuroscience* 2006, **9**(11):1432–1438.

- [17] Deneve S: **Bayesian spiking neurons I: inference.** *Neural Comput* 2008, **20**:91–117.
- [18] Rolls ET, Tovee MJ: **Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex.** *J Neurophysiol* 1995, **73**(2):713–26.
- [19] Olshausen BA, Field DJ: **Sparse coding of sensory inputs.** *Curr Opin Neurobiol* 2004, **14**(4):481–7.
- [20] Balduzzi D, Tononi G: **Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework.** *PLoS Comput Biol* 2008, **4**(6):e1000091.
- [21] Balduzzi D, Tononi G: **Qualia: the geometry of integrated information.** *PLoS Comput Biol* 2009, **5**(8):e1000462.
- [22] Pearl J: *Causality: models, reasoning and inference.* Cambridge University Press 2000.
- [23] Quiroga RQ, Reddy L, Koch C, Fried I: **Invariant visual representation by single neurons in the human brain.** *Nature* 2005, **435**(1102-1107).
- [24] Attwell D, Laughlin SB: **An energy budget for signaling in the grey matter of the brain.** *J Cereb Blood Flow Metab* 2001, **21**(10):1133–45.
- [25] Beggs JM, Plenz D: **Neuronal avalanches in neocortical circuits.** *J Neurosci* 2003, **23**(35):11167–11177.
- [26] Plenz D, Thiagarajan TC: **The organizing principles of neuronal avalanches: cell assemblies in the cortex?** *Trends Neurosci* 2007, **30**(3):101–110.
- [27] Cover TM, Thomas JA: *Elements of information theory.* John Wiley & Sons 2006.
- [28] Balduzzi D, Ortega PA, Besserve M: **Metabolic cost as an organizing principle for cooperative learning.** *Submitted.*
- [29] Xu W, Morishita W, Buckmaster PS, Pang ZP, Malenka RC, Südhof TC: **Distinct neuronal coding schemes in memory revealed by selective erasure of fast synchronous synaptic transmission.** *Neuron* 2012, **73**(5):990–1001.
- [30] Dan Y, Poo MM: **Spike timing-dependent plasticity: from synapse to perception.** *Physiol Rev* 2006, **86**(3):1033–1048.
- [31] Bhatt DH, Zhang S, Gan WB: **Dendritic Spine Dynamics.** *Annu Rev Physiol* 2009, **71**:261–282.
- [32] Vyazovskiy VV, Cirelli C, Pfister-Genskow M, Faraguna U, Tononi G: **Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep.** *Nat Neurosci* 2008, **11**(2):200–8.
- [33] Vyazovskiy VV, Olcese U, Lazimy Y, Faraguna U, Esser SK, Williams JC, Cirelli C, Tononi G: **Cortical firing and sleep homeostasis.** *Neuron* 2009, **63**(6):865–78.
- [34] Chklovskii DB, Schikorski T, Stevens CF: **Wiring optimization in cortical circuits.** *Neuron* 2002, **34**(3):341–7.
- [35] Tononi G, Cirelli C: **Some considerations on sleep and neural plasticity.** *Arch Ital Biol* 2001, **139**(3):221–41.
- [36] Tononi G, Cirelli C: **Sleep and synaptic homeostasis: a hypothesis.** *Brain Res. Bull.* 2003, **62**:143–150.
- [37] Steriade M: *The Intact and Sliced Brain.* MIT Press 2001.
- [38] Tononi G, Edelman G, Sporns O: **Complexity and coherency: integrating information in the brain.** *Trends Cog Sci* 1998, **2**(12):474–484.
- [39] Wilson MA, McNaughton BL: **Reactivation of hippocampal ensemble memories during sleep.** *Science* 1994, **265**(5172):676–9.
- [40] Tononi G, Cirelli C, Pompeiano M: **Changes in gene expression during the sleep-waking cycle: a new view of activating systems.** *Arch Ital Biol* 1995, **134**:21–37.
- [41] Czarnecki A, Birtoli B, Ulrich D: **Cellular mechanisms of burst firing-mediated long-term depression in rat neocortical pyramidal cells.** *J Physiol* 2007, **578**(2):471–9.
- [42] Huber R, Ghilardi M, Massimini M, Tononi G: **Local sleep and learning.** *Nature* 2004, **430**:78–81.

## A1. PROOF OF THEOREM 2

Given distribution  $p(x)$  on  $X$ , let  $H(p) = -\sum_{x \in X} p(x) \cdot \log p(x)$  denote the entropy.

### Theorem 2.

Let  $p_1(y|x)$  and  $p_2(z|y)$  be Markov matrices on finite sets  $X$ ,  $Y$  and  $Z$ . Let  $p_{12}(z|x) = \sum_{y \in Y} p_1(y|x) \cdot p_2(z|y)$  denote the composite channel. Then

$$ei(X \xrightarrow{n_2 \circ n_1} z) \leq \sum_{y \in Y} c(y|z) \cdot ei(X \xrightarrow{n_1} y), \text{ where } c(y|z) := p_2(z|y) \cdot \frac{p_1(y)}{p_{12}(z)}.$$

*Proof.* As usual,  $p_1(y) := \sum_{x \in X} p_1(y|do(x)) \cdot p_{unif}(x)$  and  $p_{12}(z) := \sum_{x \in X} p_{12}(z|do(x)) \cdot p_{unif}(x)$ .

The actual repertoire is

$$(8) \quad \hat{p}_{12}(x|z) = \frac{p_{12}(z|do(x)) \cdot p_{unif}(x)}{p_{12}(z)} = \frac{\sum_{y \in Y} p_1(y|do(x)) \cdot p_2(z|y) \cdot p_{unif}(x)}{p_{12}(z)}.$$

Observe that

$$(9) \quad \hat{p}_1(x|y) = \frac{p_1(y|do(x)) \cdot p_{unif}(x)}{p_1(y)}.$$

Combining Eq. (8) and (9) obtains

$$\hat{p}_{12}(x|z) = \sum_{y \in Y} c(y|z) \cdot \hat{p}_1(x|y).$$

It is easy to check that  $c(y|z)$  induces a probability distribution on  $Y$  so that, by convexity of relative entropy [27],

$$\underbrace{H \left[ \sum_{y \in Y} c(y|z) \cdot \hat{p}_1(X|y) \middle\| \sum_{y \in Y} c(y|z) \cdot p_{unif}(X) \right]}_{ei(X \xrightarrow{n_2 \circ n_1} z)} \leq \sum_{y \in Y} c(y|z) \cdot \underbrace{H \left[ \hat{p}_1(X|y) \middle\| p_{unif}(X) \right]}_{ei(X \xrightarrow{n_1} y)}.$$

□

## A2. EFFECTIVE INFORMATION FOR FIG. 3

Effective information for the two detectors can be computed by exhaustively perturbing each with all possible configurations. However, since we understand their mechanisms, it is easy to replace exhaustive perturbation of the elements with combinatorics. First, since we have imposed the condition that every configuration contains 4 distinct dots in the  $8 \times 8$  grid, it follows that the number of potential input patterns is  $635,376 = \binom{8 \times 8}{4}$ .

**Effective information for the vertical rectangle:  $n_2$ .** The number of configurations fitting inside a vertical rectangle of width 7 is computed as follows. 4 dots can fit inside a rectangle of width 7 in  $\binom{8 \times 7}{4}$  ways. Excluding configurations that fit inside a smaller rectangle of width 6, we find that there are  $\binom{8 \times 7}{4} - 2\binom{8 \times 6}{4} + \binom{8 \times 5}{4}$  configurations that fit inside a rectangle of width 7, but *not* a rectangle of width 6 (we add  $\binom{8 \times 5}{4}$  to compensate for double counting). Finally, there are two ways a rectangle of width 7 can fit inside the grid, so there are  $139,040 = 2 \left[ \binom{8 \times 7}{4} - 2\binom{8 \times 6}{4} + \binom{8 \times 5}{4} \right]$  configurations that fit only inside a 7 pixel wide rectangle. Effective information is 2.2 bits.

**Effective information for the horizontal rectangle:  $n_1$ .** The computation is similar to that above. There are  $\binom{8 \times 2}{4} - 2\binom{8}{4}$  ways a configuration of 4 dots can fit inside a rectangle of height 2, *without* fitting inside a smaller rectangle of height 1. There are 7 different ways a rectangle of height 2 can be placed inside the grid, so there are  $11,760 = 7 \left[ \binom{8 \times 2}{4} - 2\binom{8}{4} \right]$  configurations only fit a 2 pixel high rectangle. Effective information is 5.8 bits.

## A3. APPLYING THEOREM 2 TO FIGURES 4 AND 5

We apply equation

$$ei(X \xrightarrow{n_2 \circ n_1} z) \leq \sum_{y \in Y} c(y|z) \cdot ei(X \xrightarrow{n_1} y), \text{ where } c(y|z) := p_2(z|y) \cdot \frac{p_1(y)}{p_{12}(z)}$$

to the three cases in turn.

**AND-gates.**



$$\begin{aligned}
4 = \log_2 \frac{16}{1} &= ei(X \xrightarrow{AND^2 \circ AND} 1) \leq \sum_{y \in \{11\}} 1 \cdot ei(X \xrightarrow{AND^2} 11) \\
&= \sum_{y \in \{11\}} \left[ ei(X \xrightarrow{AND} 1) + ei(X \xrightarrow{AND} 1) \right] \\
&= 2 + 2
\end{aligned}$$

Distribution  $c(y|z)$  is concentrated on  $y = 11$ . The channel  $n_1 = AND^2$  decomposes into two AND-gates, generating 2 bits of information each. The only input on the bottom layer that causes a spike on the top layer is 1111.

### OR-gates.

$$\begin{aligned}
0.1 = \log_2 \frac{16}{15} &= ei(X \xrightarrow{OR^2 \circ OR} 1) \leq \sum_{y \in \{01, 10, 11\}} c(y|z) \cdot ei(X \xrightarrow{OR^2} y) \\
&= \frac{16}{15} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \left[ ei(X \xrightarrow{OR} 0) + ei(X \xrightarrow{OR} 1) \right] \\
&+ \frac{16}{15} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \left[ ei(X \xrightarrow{OR} 1) + ei(X \xrightarrow{OR} 0) \right] \\
&+ \frac{16}{15} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \left[ ei(X \xrightarrow{OR} 1) + ei(X \xrightarrow{OR} 1) \right] \\
&= \frac{3}{15} (2 + 0.415) + \frac{3}{15} (0.415 + 2) + \frac{9}{15} (0.415 + 0.415) \\
&= 1.02
\end{aligned}$$

The distribution  $c(y|z)$  points to three potential causes: 01, 10, and 11, and is therefore less concentrated than in the case of an AND-gate. Moreover, each of these three potential inputs is less informative than in the preceding case.

The set of potential inputs on the bottom layer that cause a spike on the top layer includes everything *except* 0000.

Finally, note that in this case the upper-bound is very loose.

### NOR-gates.

$$\begin{aligned}
0.83 = \log_2 \frac{16}{9} &= ei(X \xrightarrow{NOR^2 \circ NOR} 1) \leq \sum_{y \in \{00\}} 1 \cdot ei(X \xrightarrow{NOR^2} 00) \\
&= \sum_{y \in \{00\}} \left[ ei(X \xrightarrow{NOR} 0) + ei(X \xrightarrow{NOR} 0) \right] \\
&= 0.415 + 0.415
\end{aligned}$$

As for the AND-hierarchy, distribution  $c(y|z)$  is concentrated. However, it points to input 00, for which NOR-gates are uninformative.

The nine potential inputs causing output 1 on the top layer are  $\{0101, 1001, 0110, 1010, 1011, 0111, 1101, 1110, 1111\}$ .

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS, TÜBINGEN, GERMANY  
*E-mail address:* [balduzzi@tuebingen.mpg.de](mailto:balduzzi@tuebingen.mpg.de)

DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF WISCONSIN-MADISON, MADISON, WI, USA  
*E-mail address:* [gtononi@wisc.edu](mailto:gtononi@wisc.edu)