Frontiers in Neuroinformatics

Research Article
20 January 2014

1

Local active information storage as a tool to understand distributed neural information processing

Michael Wibral [1,*], **Joseph T. Lizier,**[2], **Sebastian Vögler** [3], **Viola Priesemann,**[4], **and Ralf Galuske** [3]

[1] *MEG Unit, Brain Imaging Center, Goethe University, Frankfurt am Main, Germany*
[2] *CSIRO Computational Informatics, Marsfield, NSW, Australia*
[3] *Fakultät für Biologie, Technische Universtät, Darmstadt, Germany*
[4] *Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany*

Correspondence*:
Michael Wibral
MEG Unit, Brain Imaging Center, Goethe University, Heinrich-Hoffmann Strasse 10, Frankfurt am Main, D-602528, Germany, wibral@em.uni-frankfurt.de

Information-based methods for neuroimaging: analyzing structure, function and dynamics

2 **ABSTRACT**

3 Every act of information processing can in principle be decomposed into the component
4 operations of information storage, transfer, and modification. Yet, while this is easily done for
5 today's digital computers, the application of these concepts to neural information processing
6 was hampered by the lack of proper mathematical definitions of these operations on information.
7 Recently, definitions were given for the dynamics of these information processing operations on
8 a local scale in space and time in a distributed system, and the specific concept of local active
9 information storage was successfully applied to the analysis and optimization of artificial neural
10 systems. However, no attempt to measure the space-time dynamics of local active information
11 storage in neural data has been made to date. Here we measure local active information storage
12 on a local scale in time and space in voltage sensitive dye imaging data from area 18 of the cat.
13 We show that storage reflects neural properties such as stimulus preferences and surprise upon
14 unexpected stimulus change, and in area 18 reflects the abstract concept of an ongoing stimulus

despite the locally random nature of this stimulus. We suggest that LAIS will be a useful quantity to test theories of cortical function, such as predictive coding.

Keywords: Visual System, Neural Dynamics, Predictive Coding, Local Information Dynamics, Voltage Sensitive Dye Imaging, Distributed Computation, Complex Systems, Information Storage

1 INTRODUCTION

It is commonplace to state that brains exist to 'process information'. Curiously enough, however, it is much more difficult to exactly quantify this putative processing of information. In contrast, we have no difficulties to quantify information processing in a digital computer, e.g. in terms of the information stored on its hard disk, or the amount of information transfered per second from its hard disk to its random access memory, and then on to the CPU. Why then is it so difficult to perform a similar quantification for biological, and especially neural information processing?

One answer to this question is the conceptual difference between a digital computer and a neural system: In a digital computer all components are laid out such that they only perform specific operations on information: a hard disk should store information, and not modify it, while the CPU should quickly modify the incoming information and then immediately forget about it, and system buses exist solely to transfer information. In contrast, in neural systems it is safe to assume that each element of the system (each neuron) *simultaneously* stores, transfers and modifies information in variable amounts, and the component processes are hard to separate quantitatively. Thus, while in digital computers the distinction between information storage, transfer and modification comes practically for free, in neural systems separating the components of distributed information processing requires thorough mathematical definitions of information storage, transfer and modification. Such definitions, let alone a conceptual understanding of what the terms meant in distributed information processing, were unavailable until very recently (**Langton**, 1990; **Mitchell**, 1998; **Lizier**, 2013).

These necessary mathematical definitions were recently derived building on Turing's old idea that every act of information processing can be decomposed into the component processes of information storage, transfer and modification (**Turing**, 1936) – very much in line with our everyday view of the subject. Later, Langton and others expanded Turing's concepts to describe the emergence of the capacity to perform arbitrary information processing algorithms, or 'universal computation', in complex systems, such as cellular automata (**Langton**, 1990; **Mitchell et al.**, 1993), or neural systems. The definitions of information transfer and storage were then given by **Schreiber** (2000), **Crutchfield and Feldman** (2003) and **Lizier et al.** (2012b). However, the definition of information modification is still a matter of debate (**Lizier et al.**, 2013).

Of these three component processes above – information transfer, storage, and modification – information storage in particular has been used with great success to analyze cerebro-vascular dynamics (**Faes et al.**, 2013), information processing in swarms (**Wang et al.**, 2012), and most importantly, to evolve (**Prokopenko et al.**, 2006), and optimize (**Dasgupta et al.**, 2013) artificial information processing systems. This suggests that the analysis of information storage could also be very useful for the analysis of neural systems.

Yet, while neuroscientists have given much attention to considering how information is stored structurally in the brain, e.g. via synaptic plasticity, the same attention has not been given to information storage in neural dynamics, and its quantification. As an exception **Zipser et al.** (1993) clearly contrasted two different ways of storing information: *passive storage*, where information is stored "in modified values of physiological parameters such as synaptic strength", and *active storage* where "information is preserved by maintaining neural activity throughout the time it must be remembered". In the same paper, the authors go on to point out that there is evidence for the use of both storage strategies in higher animals, and link the relatively short time scale for active storage (at maximum in the tens of seconds) with short-term or working memory and, therefore, refer to it as "active information storage".

Despite the importance of information storage for neural information processing, information theoretic measures of active information storage have not yet been used to quantify information processing in neural systems, and in particular not to measure spatiotemporal patterns of information storage dynamics. Therefore, it is the aim of this article to introduce measures of information storage as analysis tools for the investigation of neural systems, and to demonstrate how cortical information storage in visual cortex unfolds in space and time. We will also demonstrate how neural activity may be misinformative about its own future and thereby generates 'surprise'.

To this end, we first give a rigorous mathematical definition of information storage in dynamic activity in the form of local active information storage (LAIS). We then show how to apply this measure to voltage sensitive dye imaging data from cat visual cortex. In these data , we found sustained increases in dynamic information storage during visual stimulation, organized in clear spatiotemporal patterns of storage across the cortex, including stimulus-specific spatial patterns, and negative storage, or surprise, upon a change of the stimulus. Finally, we discuss the implications of the LAIS measure for neurophysiological theories of predictive coding (see **Bastos et al.** (2012), and references therein), that have been suggested to explain general operating principles of the cortex and other hierarchical neural systems.

2 MATERIAL & METHODS

The use of the stored information for information processing inevitably requires its re-expression in neural activity and its interaction with ongoing neural activity and incoming information. Hence, information storage *actively in use for information processing* will inevitably be reflected in the dynamics of neural activity, and is therefore accessible in recordings of neural activity alone. To quantify this stored information that is present in neural time series we will now introduce a measure of information storage called *local active information storage* (**Lizier et al.**, 2012b). In brief, this measure quantifies the amount of information in a sample from a neural time series that is predictable from its past – and thereby has been stored in this past. This is done by simply computing the local mutual information between the past of a neural signal and its next sample at each point in time, and for each channel of a recording. As the following material is necessarily formal, the reader may consider skipping ahead to section 2.2.3 at first reading to gain an intuitive understanding of mechanisms that serve active information storage.

2.1 NOTATION AND INFORMATION THEORETIC PRELIMINARIES

To avoid confusion, we first have to state how we formalize observations from neural systems mathematically. We define that a neural (sub-)system of interest (e.g. a neuron, or brain area) $\mathcal{X}$ produces an observed time series $\{x_1, \ldots, x_t, \ldots, x_N\}$, sampled at time intervals δ. For simplicity we choose our temporal units such that $\delta = 1$, and hence index our measurements by $t \in \{1...N\} \subseteq \mathbb{N}$, i.e. we index in terms of samples. The full time series is understood as a realization of a *random process* $\mathtt{X}$. This random processes is nothing but a collection of random variables X_t, sorted by an integer index (t in our case). Each random variable X_t, at a specific time t, is described by the set of all its J possible outcomes $\mathcal{A}_{X_t} = \{a_1, \ldots, a_j, \ldots, a_J\}$, and their associated probabilities $p_t(x_t = a_j)$. The probabilities of a specific outcome $p_t(x_t = a)$ may change with t, i.e. when going from one random variable to the next. In this case, we will indicate the specific random variable X_t the probability distribution belongs to – hence the subscript in $p_t(\cdot)$. For practical estimation of $p_t(\cdot)$ then, multiple time-series realizations or trials would be required. For stationary processes, where $p_t(x_t = a)$ does not change with t, we simply write $p(x_t)$, and practical estimation may be done from a single time-series realization. In sum, in this notation the individual random variables X_t produce realizations x_t, and the time-point index of a random variable X_t is necessary when the random process is nonstationary. When using more than one system, the notation is generalized to multiple systems $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$.

As we will see below, active information storage is nothing but a specific mutual information between collections of random variables in the process in question. We therefore start by giving the definition of *mutual information* (MI) $I(X;Y)$ as the amount of information held in common by two random variables U, V on average (**Cover and Thomas**, 1991):

$$I(U;V) = \sum_{u\in\mathcal{A}_U, v\in\mathcal{A}_V} p(u,v)\log\frac{p(u,v)}{p(u)p(v)}\,, \tag{1}$$

$$= \sum_{u\in\mathcal{A}_U, v\in\mathcal{A}_V} p(u,v)\log\frac{p(v\mid u)}{p(v)}\,, \tag{2}$$

where the log can be taken to an arbitrary base, and choosing base 2 yields the mutual information in bits. Note that the mutual information $I(U;V)$ is symmetric in U and V. As shown more explicitly in equation 2, the MI $I(U;V)$ measures the amount of information provided (or the amount that uncertainty is reduced) by an observation of a specific outcome u of the variable U about the occurrence of another specific outcome v of V - on average over all possible values of u and v. As originally pointed out by **Fano** (1961), the summands $\log\frac{p(v|u)}{p(v)}$ have a proper interpretation even without the weighted averaging – as the information that observation of a specific u provides about the occurrence of a specific v. The *pointwise* or *local mutual information* is therefore defined as:

$$i(u;v) = \log\frac{p(v\mid u)}{p(v)}\,. \tag{3}$$

It is important to note the distinction of the local mutual information measure $i(x;y)$ considered here from partial localization expressions, i.e. the partial mutual information or specific information $I(u;V)$ which are better known in neuroscience (**DeWeese and Meister**, 1999; **Butts**, 2003; **Butts and Goldman**, 2006). Partial MI expressions consider information contained in specific values u of one variable U about the other (unknown) variable V. Crucially, there are two valid approaches to measuring partial mutual information, one which preserves the additivity of information and one which retains non-negativity (**DeWeese and Meister**, 1999). In contrast, the fully local mutual information $i(x;y)$ that is used here is uniquely defined as shown by **Fano** (1961).

2.2 LOCAL ACTIVE INFORMATION STORAGE

Using the definition in equation 3, we can immediately quantify how much of the information in in the outcome x_t of the random variable X_t at time t was predictable from the observed past *state* $\mathbf{x}_{t-1}^{k-}$ of the process at time $t-1$:

$$a(x_t) = i(\mathbf{x}_{t-1}^{k-}; x_t) \tag{4}$$

$$= \log \frac{p_t(x_t \mid \mathbf{x}_{t-1}^{k-})}{p_t(x_t)} \, . \tag{5}$$

This quantity was introduced by **Lizier et al.** (2012b) and called *local active information storage* (LAIS). Here, $\mathbf{x}_{t-1}^{k-}$ is an outcome of the collection of previous random variables $\mathbf{X}_{t-1}^{k-} = \{X_{t-1}, X_{t-t_1}, \ldots, X_{t-t_{kmax}}\}$, called a *state* (see below). The corresponding expectation value over all possible observations of x_t and $\mathbf{x}_{t-1}^{k-}$, $A(X_t) = I(\mathbf{X}_{t-1}^{k-}; X_t)$, is known simply as the *active information storage*. The naming of this measure aligns well with the concept of active storage in neuroscience by **Zipser et al.** (1993), but is more general than capturing only sustained firing patterns. In the following subsections, we comment on practical issues involved in estimating the LAIS, and discuss its interpretation.

2.2.1 Interpretation and construction of the past state As indicated above, the joint variable $\mathbf{x}_{t-1}^{k-}$ in (equation 4) is an outcome of the collection of previous random variables: $\mathbf{X}_{t-1}^{k-} = \{X_{t-1}, X_{t-t_1}, \ldots, X_{t-t_{kmax}}\}$. This collection should be constructed such, that it captures the *state* of the underlying dynamical system $\mathcal{X}$, and can be viewed as a state-space reconstruction of this system. In this sense, $\mathbf{X}_{t-1}^{k-}$ must be chosen such that X_t is conditionally independent of all X_{t-t_l} with $t_l > t_{k_{max}}$, i.e. of all variables that are observed earlier in the process $\mathtt{X}$ than the variables in the state at $t-1$. The choice must be made carefully, since using too few variables X_{t-t_l} from the history can result in an underestimate of $a(x_t)$, while using too many (given the amount of data used to estimate the probability density functions (PDFs) in (equation 4)) will artificially inflate it. Typically, the state can be captured via Takens delay embedding (**Takens**, 1981), using d variables X_{t-t_l} with the t_l delays equally spaced by some $\tau \geq 1$, with d and τ selected using the Ragwitz criteria (**Ragwitz and Kantz**, 2002) – as recommended by **Vicente et al.** (2011) for the related transfer entropy measure (**Schreiber**, 2000). Alternatively, non-uniform embeddings may be used (e.g. see **Faes et al.** (2012)).

If the process has infinite memory, and k_{max} does not exist, then the local active information storage is defined as the limit $\lim_{k\to\infty}$ of equation 4:

$$a(x_t) = \lim_{k\to\infty} i(\mathbf{x}_{t-1}^{k-}; x_t) \tag{6}$$

$$= \lim_{k\to\infty} \log \frac{p_t(x_t \mid \mathbf{x}_{t-1}^{k-})}{p_t(x_t)} \, . \tag{7}$$

2.2.2 Relation to other measures and dynamic state updates The average active information storage (AIS), is related to two measures introduced previously. On the one hand, a similar measure called 'regularity' had been introduced by **Porta et al.** (2000). On the other hand, AIS is closely related to the excess entropy (**Crutchfield and Feldman**, 2003), as observed in (**Lizier et al.**, 2012b). The excess entropy $E(X_t) = I(\mathbf{X}_{t-1}^{k-}; \mathbf{X}_t^{k+})$, with $\mathbf{X}_t^{k+} = \{X_t, X_{t+t_1}, \ldots, X_{t+t_{k_{max}}}\}$ being a similar collection of future random variables from the process, measures the amount of information (on average) in the future outcomes $\mathbf{x}_t^{k+}$ of the process this is predictable from the observed past state $\mathbf{x}_{t-1}^{k-}$ at time $t-1$. As such, the excess entropy captures all of the information in the future of the process that is predictable from its past. In measuring the subset of that information in only the next outcome of the process, the AIS is focused on the dynamic state updates of the process.

From the point of view of dynamic state updates, the AIS is *complementary* to a well-known measure of uncertainty of the next outcome of the process which cannot be resolved by its past state. Following **Crutchfield and Feldman** (2003) we refer to this quantity as the "entropy rate", the conditional entropy of the next outcome given the past state: $H_\mu(X_t) = H(X_t \mid \mathbf{X}_{t-1}^{k-}) = \left\langle -\log_2 p_t(x_t \mid \mathbf{x}_{t-1}^{k-}) \right\rangle$. The complementarity of the entropy rate and AIS was shown by **Lizier et al.** (2012b): $H(X_t) = A(X_t) + H_\mu(X_t)$, where $H(X_t)$ is the Shannon entropy of the next measurement X_t. $H_\mu(X_t)$ is approximated by measures known as the Approximate Entropy (**Pincus**, 1991), Sample Entropy (**Richman and Moorman**, 2000), and Corrected Conditional Entropy (**Porta et al.**, 1998), which have been well studied in neuroscience (see e.g. the work by **Vakorin et al.** (2011); **Gómez and Hornero** (2010), and references therein). Many such studies refer to $H_\mu(X_t)$ as a measure of complexity, however modern complex systems perspectives focus on complexity as being captured in how much structure can be resolved rather than how much cannot (**Crutchfield and Feldman**, 2003).

Furthermore, given that the most appropriate measure of complexity of a process is a matter of open debate (**Prokopenko et al.**, 2009), we take the perspective that complexity of a system is best approached as arising out of the interaction of the component operations of information processing: information storage, transfer and modification (**Lizier**, 2013), and focus on measuring these quantities since they are rigorously defined and well-understood. Crucially, in comparison to the excess entropy discussed above, the focus of AIS in measuring the information storage in use in dynamic state updates of the process make it directly comparable with measures of information storage and modification. Of particular importance here is the relationship of AIS to the transfer entropy (**Schreiber**, 2000), where the two measures together reveal the sources of information (either being the past of that process itself – storage, or of other processes – transfer) which contribute to prediction of the process' next outcome.

The formulation of the transfer entropy specifically eliminates information storage in the past of the target process from being mistakenly considered as having been transferred (**Lizier**, 2013; **Lizier and**

Prokopenko, 2010; **Wibral et al.**, 2013). An interesting example is where a periodic target process is in fact causally driven by another periodic process – after any initial entrainment period, our information processing view concludes that we have information storage here in the target but no transfer from the driver (**Lizier and Prokopenko**, 2010). While causally there is a different conclusion, our observational information processing perspective is simply focussed on decomposing apparent information sources of the process, regardless of underlying causality (which in practise cannot often be determined anyway). In this view, a causal interaction can computationally subserve both information storage or transfer (as discussed further in the next section). Information transfer is necessarily linked to a causal interaction, but the reverse is not true. It has previously been demonstrated that the information processing perspective is more relevant to emergent information processing structure in complex systems, e.g. coherent information cascades, in contrast to causal interactions being more relevant to the micro-scale physical structure of a system, e.g. axons in a neural system (**Lizier and Prokopenko**, 2010).

2.2.3 Mechanisms producing active information storage In contrast to passive storage in terms of modifications to system structure (e.g. synaptic gain changes), the mechanisms underlying active information storage are not immediately obvious. The mechanisms that subserve this task have been formally established however, and can be grouped as follows:

1. *Physical mechanisms in the system*. This could incorporate some internal memory mechanism in the individual physical element giving rise to the process X (e.g. some decay function, or the stereotypical processes during the refractory period after a neural spike). More generally, it may involve network structures which offload or distribute the memory function onto edges or other nodes. In particular, **Zipser et al.** (1993) reported that networks with fixed, *recurrent connections* were sufficient to account for such active storage patterns, which is in line with earlier proposals. Furthermore, **Lizier et al.** (2012a) quantified the AIS contribution from self-loops, feedback and feedforward loops (as the only network structures contributing to active information storage).
2. *Input-driven storage*. This describes situations where the apparent memory in the process is caused by information storage structure which lies in another element which is driving that process, e.g. a periodically spiking neuron that may cause a downstream neuron to spike with the same period (**Obst et al.**, 2013). As described in Section 2.2.2 above, an observer of the process attributes these dynamics to information storage, regardless of the (unobserved) underlying causal mechanism.

Of these mechanisms of active information storage the case of circular causal interactions in a loop motif, and the causal, but repetitive influence from another part of the system may seem counterintuitive at first, as we might think that in these cases there should be information transfer rather than active information

storage. To see why these interactions serve storage rather than transfer, it may help to consider that *all* components of information processing, i.e. transfer, active storage and modification, ultimately have to rely on causal interactions in physical systems. Hence, the presence of a causal interaction cannot be linked in a one-to-one fashion to information transfer, as otherwise there would be no possibility for physical causes of active information storage and of information modification left, and no consistent decomposition of information processing would be possible. Therefore, the notion of storage that is measurable in a part of the system but that can be related to external influences onto that part is to be prefered for the sake of mathematical consistency and ultimately, usefulness. We acknowledge that information transfer has often been used as a proxy for a causal influence, dating back to suggestions by **Wiener** (1956) and **Granger** (1969). However, now that causal interventional measures and measures of information transfer can be clearly distinguished (**Lizier and Prokopenko**, 2010; **Ay and Polani**, 2008) it seems no longer warranted to map causal interactions to information transfer in a one-to-one manner.

2.2.4 Interpretation of LAIS values Measurements of the LAIS tells us the amount to which observing the past *state* $\mathbf{x}_{t-1}^{k-}$ reduced our uncertainty about the specific next outcome x_t that was observed. We can interpret this in terms of *encoding* the outcome x_t in bits: encoding x_t using an optimal encoding scheme for the distribution $p_t(x_t)$ takes $-\log_2 p_t(x_t)$ bits, whereas encoding x_t if we know $\mathbf{x}_{t-1}^{k-}$ using an optimal encoding scheme for the distribution $p_t(x_t \mid \mathbf{x}_{t-1}^{k-})$ takes $-\log_2 p_t(x_t \mid \mathbf{x}_{t-1}^{k-})$ bits, and the LAIS is the number of bits saved via the latter approach.

At first glance we may assume that the LAIS is a positive quantity. Indeed, as a mutual information, the *average* AIS will always be non-negative. However, the LAIS can be negative as well as positive. It is positive where $p_t(x_t \mid \mathbf{x}_{t-1}^{k-}) > p_t(x_t)$, i.e. where the observed past state $\mathbf{x}_{t-1}^{k-}$ made the following observation x_t more likely to occur than we would have guessed without the knowledge of the past state. In this case, we state that $\mathbf{x}_{t-1}^{k-}$ was informative. In contrast, the LAIS is negative where $p_t(x_t \mid \mathbf{x}_{t-1}^{k-}) < p_t(x_t)$; i.e. where the observed past state $\mathbf{x}_{t-1}^{k-}$ made the following observation x_t less likely to occur than we would have guessed without the knowledge of the past state (but it occurred nevertheless, making the cue given by $\mathbf{x}_{t-1}^{k-}$ misleading). In this case, we state that $\mathbf{x}_{t-1}^{k-}$ was *misinformative* about x_t. To better understand negative LAIS also see the further discussion in (**Lizier et al.**, 2012a), including examples in cellular automata where the past state of a variable was misinformative about the next observation due to the strong influence of an unobserved other source variable at that time point.

2.2.5 Choice of the overall time window for constructing probability densities from data As already pointed out above, active information storage is tightly related to predictability of a given brain area's output as seen by the receiving brain area. This predictability hinges on the ability of the receiver to see the past states in the output of a brain area (see previous section) and to interpret the past states in

the received time series in order to make a prediction about the next value. In other words, the receiver needs to guess $p_t(x_t, \mathbf{x}_{t-1}^{k-})$ correctly in order to exploit the active information storage. If the guess of the receiving neuron (n) or brain area, i.e. $\widetilde{p_n}(x_t, \mathbf{x}_{t-1}^{k-})$, is incorrect, then only a fraction of the information storage can be used for successfully predicting future events. The losses could be quantified as the extra coding cost for the receiving area, when assuming $\widetilde{p_n}(\cdot)$ instead of $p_t(\cdot)$. This loss would simply be the Kullback-Leibler divergence $D_{KL}(p_t||\widetilde{p_n})$. This scenario sees the receiving brain area mostly as an optimal encoder or compressor. In contrast, the cost occurring in the framework of predictive coding theories would arise because the receiving brain area could not predict the incoming signal well, and thereby inhibit it via feedback to the sending brain area (**Rao and Ballard**, 1999). In this scenario, the cost of imperfect predictions resulting from using $\widetilde{p_n}$ instead of p_t, would be reduced inhibition and a more frequent signaling of prediction errors by the sending system, leading to a metabolic cost.

To see the storage that the receiving brain area can exploit, the time interval used for the practical estimation of the probability density functions (PDFs) from neural recordings should best match the expected sampling strategy of the receiving brain area. For example, if we think that probabilities are evaluated over long time frames, then it might make sense to pool all available data in the experiment, as even a mis-estimation of the true probability densities $p_t(\cdot)$ (due to potential nonstationarities) then will better reflect the internal estimate $\widetilde{p_n}(x_t, \mathbf{x}_{t-1}^{k-})$, and thus the *internally* predictable information. However, if we think that probabilities are only estimated instantaneously by pooling over all available inputs to a brain area at any time point, then we should construct the necessary PDFs only from all simultaneously acquired data from all measurement channels, but not pool over time. The latter view could also be described as assuming that the brain area receiving the signals in question computes the PDF instantaneously by pooling over all its inputs, without keeping any longer term memory of the observed probabilities. This construction of a PDF would be linked closely to an instantaneous physical ensemble approach, considering that all incoming channels are physically equivalent, but are only assessed at a single instant in time. In contrast, if we assume that learning of the relevant PDFs takes place on a lifelong timescale, then PDFs should be acquired from very long recordings of a freely behaving subject or animal in a natural environment, and the outcomes of a specific experiment should be interpreted using this 'lifelong' PDF. Here we lean towards this latter approach and pool all available data to estimate the internally available $\widetilde{p}_n$.

Note that while we indeed pool over all the available data to obtain the distribution $\widetilde{p}_n$, the interpretation of the data in terms of the active information storage is *local per agent and time step*. This is exactly the meaning of 'local' in local active information storage as introduced in (**Lizier et al.**, 2012b) (this is also akin to the relation of the local mutual information introduced by **Fano** (1961) and the corresponding global PDF). The local active information storage values are thus obtained by interpreting realizations for a single agent and a single time step in the light of a probability distribtion that is obtained over a more

global view of the system in space and time. This is also indicated by the use of $\widetilde{p}_n$ instead of p_t. Also see the discussion for potential other choices of obtaining p.

2.3 ACQUISITION OF NEURAL DATA

2.3.1 Animal preparation Data were obtained from an anesthetized cat. The animal had been anesthetized and artificially ventilated with a mixture of O_2 and N_2O (30/70 %) supplemented with Halothane (0.7 %). All procedures were along the guidelines of the Society for Neuroscience, in accordance with the German law for the protection of laboratory animals, permitted by the local authorities and overseen by a designated veterinarian.

2.3.2 Voltage sensitive dye imaging For optical imaging the visual cortex (area 18) was exposed and an imaging chamber was implanted over the craniotomy. The chamber was filled with silicone oil and sealed with a glass plate. A voltage sensitive dye (RH1691, Optical Imaging Ltd, Rehovot, Israel) was applied to the cortex for about 2 hours and subsequently the excess of the dye was washed out. For imaging we used a CMOS camera system (Imager 3001, Optical Imaging Ltd, Rehovot, Israel, Camera: Photon Focus MV1 D1312, chip size 1312x1082 pixel) fitted with a lens system consisting of two 50mm Nikon objectives providing a field of view of 8.7x10.5mm and an epifluorescence illumination system (excitation: 630+/-10nm, emission high pass 665nm). In order to optimize the signal-to-noise ratio raw camera signals were spatially binned to 32x32 camera pixels allowing for a spatial resolution of 30x32μm^2 per data pixel. Camera frames were collected at a rate 150Hz, resulting in a temporal resolution of 6.7ms.

2.3.3 Visual stimulation Stimuli were presented triggered to the heartbeat of the animal for 2s and camera frames were collected during the entire stimulation period. We will denote such a single stimulation period and the corresponding data acquisition as a trial here. Each trial consisted of 1s stimulation with an isoluminant grey screen followed by stimulation with fields of randomly positioned dots (dot size: 0.23° visual angle; 384 dots distributed over an area of 30° (vertical) by 40° (horizontal) visual angle) moving coherently in one of 8 different directions at 16 degree/s. Stimuli were presented in blocks of 16 trials, consisting of 8 trials using the stimuli described before and an additional 8 trials which consisted only of the presentation of the isoluminant grey screen for 2s ('blank trials'). Each motion direction condition was presented 8 times in total (8 trials), resulting in the presentation on 64 stimulus trials and 64 blank trials in total. Of the presented set of 8 stimulus types, 7 were used for the final analysis, as the computational process for one condition did not finish on time before local compute clusters were taken down for service.

2.3.4 VSD data post-processing After spatial binning of 32 x 32 camera pixels into one data pixel, VSD data were averaged over all presentations of blank trials and this average was subtracted from the raw data to remove the effects of dye-bleaching and heartbeat. Finally, the data were denoised using a median filter of 3x3 data pixels.

2.4 MEASUREMENT OF LAIS ON VSD NEURAL DATA

Estimation of LAIS was performed using the open source *Java information dynamics toolkit* (JIDT) (**Lizier**, 2012), with a history parameter k_{max} of ten time points, spaced 2 samples, or $(2/150Hz) = 13.\overline{3}$ ms, apart. The total history length thus covered 133 ms, or or roughly one cycle of a neural theta oscillation, which seems to be a reasonable time horizon for a downstream neural population that ultimately must assess these states. To enable LAIS estimation from a sufficient amount of samples, we considered the data pixels as homogeneous variables executing comparable state transitions, such that the pixels form a physical ensemble in terms of information storage dynamics. Pooling data over pixels thus enables an ensemble estimate of the PDFs in question. This approach seems justified as all pixels reported activity from a single brain area (area 18 of cat visual cortex, see below). Mutual information was estimated using a box kernel-estimator (**Kantz and Schreiber**, 2003) with a kernel width of 0.5 standard deviations of the data.

Here we assume that the neural system is at least capable of exploiting the statistics arising from the stimulation given throughout the experiment and thus construct PDFs from all data (time points and pixels) for a given condition. Therefore, we pool data over the full time course from -1 to 1 second of the experiment. Thus, each image of the VSD data had a spatial configuration of 67x137 spatial data pixels after removal of the 2 rows/columns on each side of an image because of the median filter that was applied. Each trial (of a total of 8 trials per condition) resulted in 288 LAIS values, based on an original data length of 298 samples and a history length (state dimension) of 10 pixels. The product of final image size and LAIS samples resulted in $2.64 \cdot 10^6$ data points per trial for the estimation of the PDF for each of the 8 motion direction conditions. Due to computational limitations, LAIS estimates were performed on two blocks of four trials separately, resulting in $1.06 \cdot 10^7$ data points entering the estimation in JIDT.

2.5 CORRELATION ANALYSIS OF LAIS AND VSD DATA

For each of the 7 analyzed motion direction conditions, VSD data and LAIS were initially organized separately per condition into 5 dimensional data structures, with dimensions: blocks (1,2), trials (1-4), time (-1 to 1 s), and pixel row (67) and columns (137). For correlation analysis, these arrays were linearized and entered into a Spearman rank correlation analysis to obtain correlation coefficients ρ(VSD,LAIS) and significance values.

Table 1. Correlation of LAIS and local VSD activity (*=$p < 0.05/7$)

Motion direction	Corr. coeff full epoch	Corr. coeff. -1s to 0	Correlation coefficient 0.04 to 0.14 s	0.2 to 1 s
0°	0.05*	-0.33*	-0.09*	0.45*
4 °	0.09*	-0.50*	-0.20*	0.65*
90°	0.12*	-0.30*	-0.13*	0.48*
180°	0.07*	-0.27*	-0.22*	0.44*
225°	0.07*	-0.58*	-0.22*	0.71*
270°	0.17*	-0.39*	-0.33*	0.68*
315°	0.03*	-0.37*	-0.17*	0.40*

Correlation coefficients are Spearman rank correlations.

3 RESULTS

LAIS values exhibited a clear spatial and temporal pattern. The temporal pattern exhibited higher LAIS values during stimulation with a moving random dot pattern than under baseline stimulation with an isoluminant grey screen, with effects being largest in spatially clearly segregated regions (Figure 1, 2, 3). The spatial pattern of LAIS under stimulation was dependent on the motion direction of the drifting random dots in the stimulus (Figure 2).

In contrast to this spatially highly selective elevation of LAIS values under stimulation, there was a sharp drop in LAIS values at approximately 40 ms after stimulus onset, with negative LAIS values measured at many pixels (Figure 1, 40 ms window; Figure 2, middle column; Figure 3, lower row). This indicates that the baseline activity was misinformative about the following stimulus related activity (since an observer would expect the baseline activity to continue). This transient, stimulus induced drop in LAIS was more evenly distributed throughout the imaging window than the elevated LAIS in the later stimulus period post 200 ms (Figure 2, middle column). The transient drop in LAIS had a recovery time of approximately 34 ms, also giving an estimate of the dominant intrinsic storage duration of the neural processes.

In all conditions we observed a positive, but weak correlation between the local VSD activity values and LAIS values over time and space (Table 1). Looking at individual time intervals, we found stronger, and negative, correlation coefficients both, for the baseline interval (-1 to 0 s), and for the initial interval after the onset of the moving dot stimulus (0.04 to 0.14 s). In contrast, we observed a strong positive correlation at the late stimulus interval (0.2 to 1 s). This means that the increased dynamic range observed in the VSD signals during stimulation with the moving stimuli led to an increased amount of predictable information, rather than to a decrease. This correlation also means that storage was generally higher in neurons that were preferentially activated by the respective moving stimulus (also compare left and right columns in Figure 2 for each motion direction).

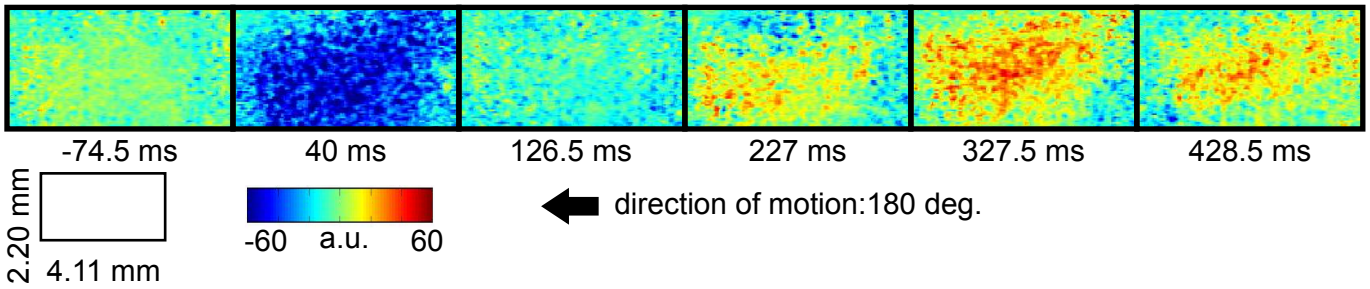

Figure 1. Local active information storage (LAIS) allows to trace neural information processing in space and time. Spatio-temporal structure of LAIS in cat area 18 – seven frames from the spatio-temporal LAIS data, taken at the times indicated below each frame. Stimulation onset was at time 0. Baseline activity (-74.5 ms) is around zero and mostly uniform. At 40 ms after stimulus onset, LAIS is negative in a region that correlates to the region that later exhibits high LAIS. Around 227 ms increased LAIS sets in and lasts until the end of the data epoch, albeit with slow fluctuations (up to 1 s, see Figure 3). Also see the post-stimulus time-average in Figure 2.

4 DISCUSSION

Our results demonstrate increased local active information storage in the primary visual cortex of the cat under sustained stimulation, compared to baseline. The spatial pattern of the LAIS increase was clustered spatially and stimulus-specific (Figure 2). The temporal pattern of LAIS consisted of a first sharp drop in LAIS from 0.04 to 0.14 s after onset of the moving stimulus and a sustained rise in LAIS up to the end of the stimulation epoch ((Figure 3). The sharp drop at stimulus onset for many pixels is important because it indicates the past activity of the pixels was surprising or misinformative about the next outcomes near that onset. This has the potential to be used in detecting changes of processing regimes directly from neural activity.

The subsequent sustained rise in LAIS is particularly notable because of the *random spatial* structure of each stimulus on a local scale; this random spatial structure translates into a random temporal stimulation sequence in the receptive field of each neuron because of the stimulus motion. The increased LAIS despite random stimulation of the neurons suggests that our observation is not due to input-driven storage, i.e. memory or storage contained already in the spatio-temporal stimulus features that drive the observed LAIS (as discussed in Section 2.2.3 and by **Obst et al.** (2013)). Nevertheless, as revealed by correlation analysis, storage was highest in regions preferentially activated by the stimulus, suggesting a representational nature of LAIS in these data with respect to the motion features of the stimulus. In sum, the changes of LAIS with stimulation onset, stimulation duration, and stimulus type clearly demonstrate that LAIS reflects neural processing, rather than mere physiological or instrumentation-dependent noise regularities. This leads us to believe that LAIS is a promising tool for the analysis of neural data in general, and of VSD data in particular.

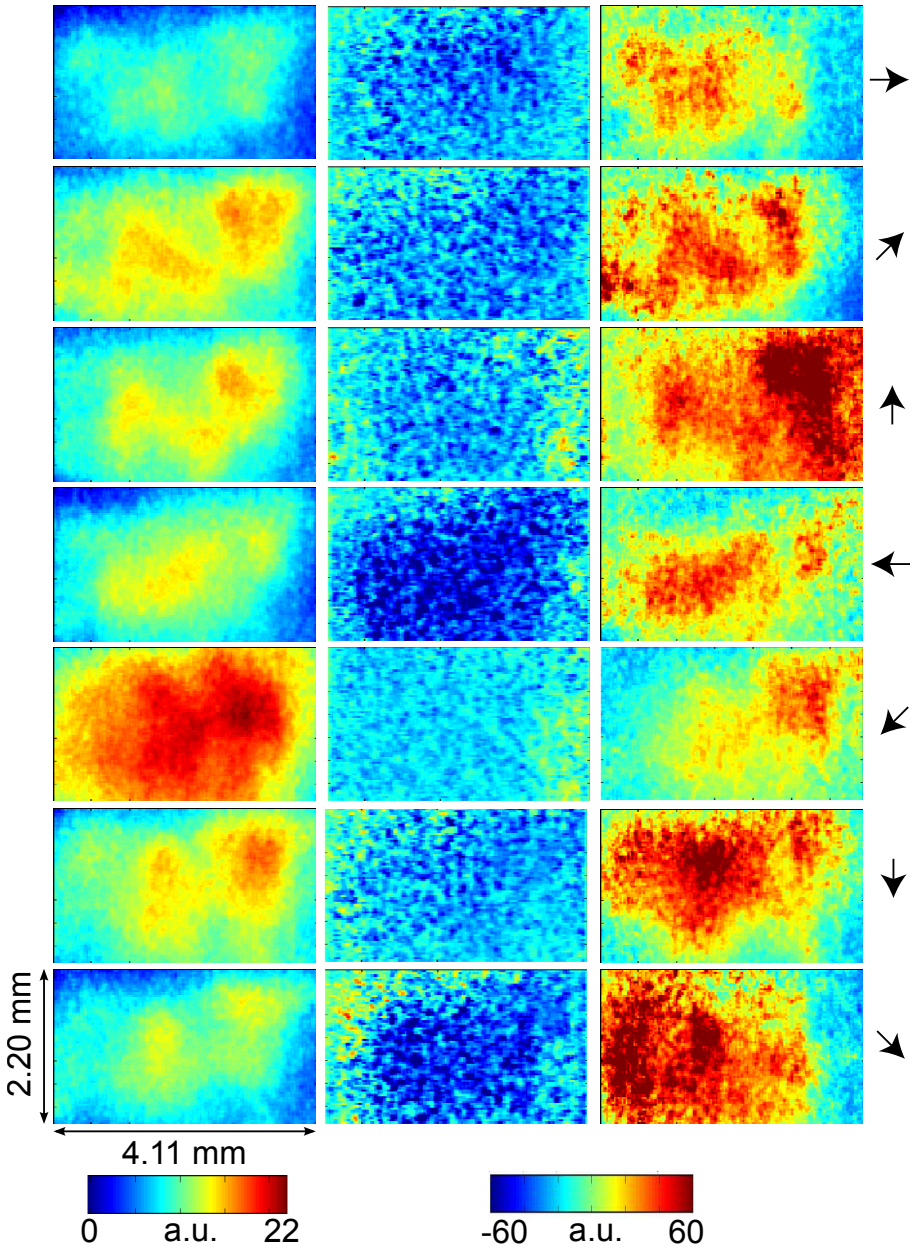

Figure 2. VSD-activity and local active information storage (LAIS) maps. VSD activity averaged over stimulation epochs and time after stimulus onset after the initial transient (0.2-1s) (left column). LAIS map immediately after stimulus onset – negative values (blue) indicate surprise of the system (middle column). Time-average LAIS maps from the stimulus period after the initial transient (0.2-1s) (right column). Rows 1-7 present different stimulus motion directions: 0, 45, 90, 180, 225, 270, 315 (in degrees, indicated by arrows on the right, arrow colors match time-trace colors in Figure 3). 67x137 data pixel per image, pixel dimension 30 x 32 μm^2. Left-right image direction is anterior-posterior direction.

4.1 LOCAL ACTIVE INFORMATION STORAGE AND NEURAL ACTIVITY LEVELS

Any increase in LAIS may in principle arise from two sources: First, a richer dynamics with a larger amplitude range – increasing overall information content, while maintaining the predictability of the time

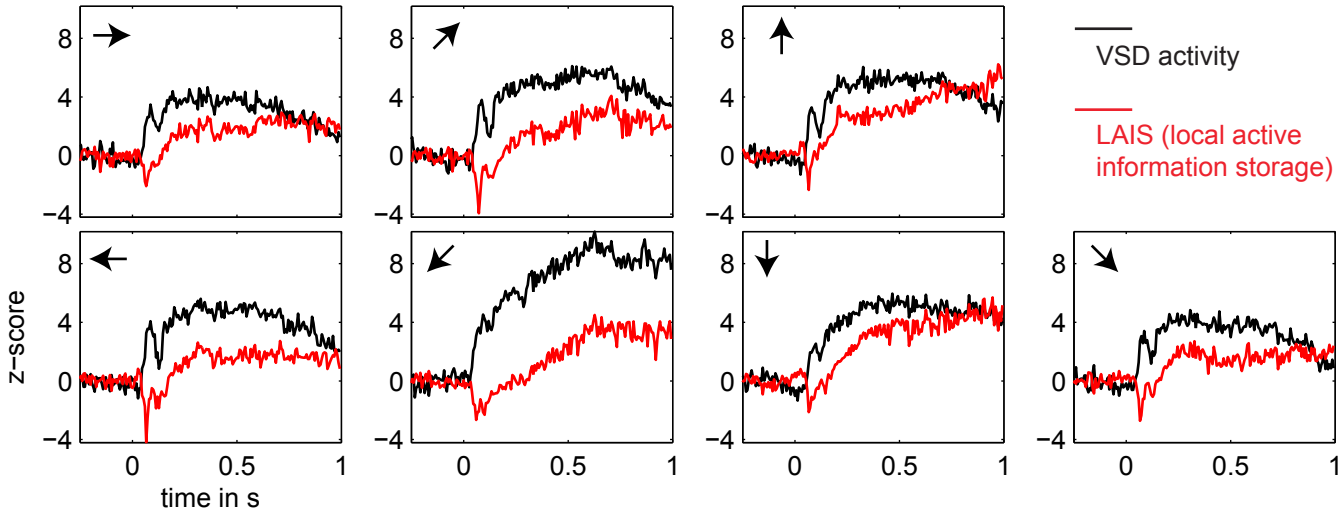

Figure 3. Temporal evolution of VSD activity and local active information storage. Spatial averages over the 67x137 data pixels for VSD activity (black traces), and the LAIS (red traces) versus time. Motion directions are indicated by arrows for each panel. Note that LAIS for the vertical, the right, and the downward-right motion directions continues to rise towards the end of the stimulus interval, despite declining activity levels. Also note that the unexpected onset response at approximately 40 ms leads to negative active information storage. For an explanation see the Methods section.

series (e.g. quantified as the inverse of the signal prediction error, or the entropy-normalized LAIS), may increase LAIS. Alternatively, increased LAIS may be based on increased predictability under essentially unchanged dynamics. The significant positive correlation between LAIS and VSD activity after stimulus onset suggests that a richer, but still predictable, dynamics of VSD activity is at the core of the stimulus-dependent effects observed here. As a caveat we have to note that the use of a kernel estimator for LAIS measurement, coupled with pooling of observations over the whole ensemble of pixels and time points may also have introduced a slight bias in favor of a positive correlation between high VSD activity and LAIS, as it allows storage to be more easily measured in pixels with larger amplitude here. The negative correlation observed in the baseline interval, however, demonstrates that this bias is not a dominant effect in our data. This is because a dominant effect of the kernel-based bias would also assign higher storage values to high amplitude data in the baseline interval, and thereby result in a positive correlation in the baseline. This was not the case. The relatively low correlation coefficients across the complete time-interval, which are between 0.02 and 0.13, further suggest that LAIS increases due not follow higher VSD tightly. Therefore, LAIS extracts additional useful information about neural processing. This point is further supported by the stimulus-dependent changes that seem more pronounced in LAIS maps than in the VSD activity maps (compare left and right columns in Figure 2).

For future studies the amplitude-bias problem introduced by the fixed-width kernel estimator should easily be overcome using a Kraskov-type variable width kernel estimator – see the original work of

Kraskov et al. (2004), and (**Vicente et al.**, 2011; **Lindner et al.**, 2011; **Wibral et al.**, 2011, 2013; **Lizier**, 2012) for implementation details of Kraskov-type estimators. Another possibility would be to condition the analysis on the activity level, as for example done for the transfer entropy measure by **Stetter et al.** (2012).

4.2 TIMESCALES OF LAIS

The recovery time of the stimulus-induced, transient drop in LAIS was 34 ms. A drop of this kind means that the activity before the drop (baseline activity) was not useful to predict the activity during the drop (the onset response). This is expected as the stimulus is presented in an unpredictable way to the neural system. However, the recovery time of this drop of approximately 34 ms yields an insight into the intrinsic storage time scales of the neural processes. We note that the observed time-scale corresponds to the high beta frequency band around 29 Hz ($1/34$ ms). In how far this is an incidental finding or bears significance must be clarified in future studies.

4.3 ON THE INTERPRETATION OF LOCAL ACTIVE INFORMATION STORAGE MEASURES IN NEUROSCIENCE

When working with measures from information theory, it is important to keep in mind that the basic definition of information as given by Shannon revolves around the probabilities of events and the possibility to encode something using these events. To separate Shannon information content from information about something (new) in a more colloquial sense, one often also speaks about *potential* or *syntactic* information, when referring to Shannon information content, of *semantic* information when referring to human interpretable information, and last of pragmatic information for our everyday notion of information as in 'news' (for details see for example the treatment of **Deacon** (2010) on this topic). In the same way, LAIS does not directly describe information that the neural system stores about things in the outside world – rather, it quantifies how much of the future (Shannon) information in the activity can be predicted from its past.

In fact, information in the neural system *about* something in the outside world would have to be quantified by some kind of mutual information between aspects of the outside world and neural activity, while information in the classic sense of semantic information represented symbolically (e.g. in books, and other media) would be even more complicated: theoretically it should be quantified as a mutual information between the medium containing the symbols and activity in the neural system, while additionally satisfying the constraint that this mutual information should vanish when conditioning on the states of the world variables represented by the symbols.

While this lack of a more semantic interpretation of LAIS may seem disappointing at first, the quantification of the predictable amount of information makes this measure highly useful in understanding information processing at a more abstract level. This is important wherever we have not yet gained insights into what (if anything) may be explicitly represented by a neural system. Moreover, the focus on predictability provides a non-trivial link between LAIS and current theories of brain function as pointed out below. Nevertheless, a use of the concept in neuroscience may have to take the properties of the receiving neuron or brain area into account to consider how much of the mathematical storage in a signal is accessible to neural information processing. To address this concern, we used a pooling over all available data in space and time here as it seems to represent a way by which a receiving brain area could construct its (implicit) guesses of the underlying probability densities. However, also other strategies are possible and need to be explored in the future. As one example for another strategy of probability-density estimation, we have investigated a construction of probability densities via pooling over all data pixels but separately for each point in time. This approach avoids any potential issues with nonstationarities, but obscures the view of the 'typical transitions' in the system over time to a point that no interpretable results were obtained (data not shown).

4.4 LOCAL ACTIVE INFORMATION STORAGE AND PREDICTIVE CODING THEORIES

Information storage in neural activity means that information from the past of a neural process will predict some non-zero fraction of information in the future of this process. It is via this predictability improvement that information storage is also tightly connected with predictive coding, an important family of theories of cortical function. Predictive coding theories propose that a neural system is constantly generating predictions about the incoming sensory input (**Rao and Ballard**, 1999; **Bastos et al.**, 2012; **Friston**, 2005; **Knill and Pouget**, 2004) to adapt internal behavior and processing accordingly. These predictions of incoming information must be implemented in neural activity, and they typically need to be maintained for a certain duration – as it will typically be unknown to the system when the predictive information will be needed. Hence, the neural activity subserving prediction must itself have a predictable character, i.e. non-zero information storage *in activity*. Analysis of active information storage may thereby enable us to test central assumptions of predictive coding theories rather directly. This is important because tests of predictive coding theories so far mostly relied on the predictions being explicitly known and then violated – a condition not given for most brain areas beyond early sensory cortices, and for most situations beyond simple experimental designs. Here, the quantification of the predictability of brain signals themselves via LAIS may open a second approach to testing these important theories. To this end we may scan brain signals for negative LAIS, as negative LAIS values indicate the past states of the neural signals in question were not informative about the future, i.e. negative LAIS signals a breakdown of predictions. In our example dataset this was brought about by the sudden, unexpected onset of the stimulus. However, the

same analyses may be applied in situations that are not a under external control – for example to analyze internally driven changes in information processing regimes.

In relation to predictive coding theories it is also encouraging that the predictive information was found on timescales related to the beta band. This is because this frequency band has been implied in the intra-cortical transfer of predictions (**Bastos et al.**, 2012).

4.5 SUB-SAMPLING AND COARSE GRAINING, AND NON-LOCALITY OF PDF ESTIMATION

When interpreting LAIS values it should be kept in mind that in neural recordings we typically do not observe the system fully or at the relevant scales – in contrast to artificial systems, such as cellular automata and robots, where the full system is accessible. More precisely, in neural data one of two types of sub-sampling is typically present – either coarse graining with local averaging of activity indices (as in VSD) or sub-sampling proper, where neural activity is recorded faithfully (e.g. via intracellular recordings) but with incomplete coverage of the full system. This sub-sampling may have non-trivial effects on the probability distributions of neural events (see for example (**Priesemann et al.**, 2009, 2013)). Hence, LAIS values obtained under sub-sampling should be interpreted as *relative* rather than absolute measures and should only be compared to other experiments, or experimental conditions, when obtained under identical sampling conditions.

In addition there is necessarily temporal subsampling in the form of finite data; we therefore note again the potential for bias in the actual MI values returned via the use of kernel estimation here, particularly for large embedding dimensions and small kernel widths. Alternatives to kernelestimators are known to be more effective in bias compensation (e.g. Kraskov-Grassberger-Stögbauer estimation (**Kraskov et al.**, 2004)); or use of use kernel estimation is solely motivated by practical computational reasons. Effects of temporal subsampling also mandates to focus on relative rather than absolute values within this experiment.

Even within the experiment though, the bias may not be evenly distributed amongst the local MI values, which tend to exhibit larger bias for low frequency events. With that said, our experiment did use a large amount of data (by pooling observations over pixels and time), which counteracts such concerns to a large degree, and many of the key results (e.g. Figure 3) involve averaging or correlating over many local values, which further ameliorates this. There are techniques suggested to alleviate bias in local or pointwise MI, e.g. (**Turney and Pantel**, 2010), and while none were applied here, we do not believe this alters the general conclusions of our experiment for the aforementioned reasons. As a particular example, the surprise caused by the onset of stimulus is still clearly visible as negative LAIS, despite any propensity for such low frequency events to have been biased strongly towards positive values.

4.6 ON THE LOCALITY OF INFORMATION VALUES

As a concluding remark, we would like to point out again that various 'levels of locality' have to be carefully chosen in the analysis of neural data. One important level is the spatial extent (ensemble of agents) and the time span over which data are pooled to obtain the PDF. However, even pooling over a large spatial extent, i.e. many agents and a long time span, may still allow to interpret the information value of the data agent-by-agent and time step-by-time step, if agents i are *identical* and samples at subsequent time points t come from a *stationary* random process (see the book of **Lizier** (2013) for several examples). This is because one may pool data to estimate a PDF as long as these data can be considered 'replications', i.e. as coming from the same random variable. Pooling data under these conditions will obviously not bias the PDF estimate away from the ground truth for any agent or time step. Irrespective of how many data points are pooled this way, it is then still possible to interpret each data point $(x_{i,t}, \mathbf{x}_{i,t-1}^{k-})$ individually in terms of its LAIS, $a(x_t, \mathbf{x}_{\mathbf{t}t-1}^{k-})$. This locality of information values is identical to the local interpretation of the (Shannon) information terms $h(x_i) = -\log(p(x_i))$ that together, as a weighted average over all possible outcomes x_i, yield the (Shannon) entropy $H(X) = \sum_i p(x_i)h(x_i)$ of a random variable X. As explained for example by **MacKay** (chapter 4, 2003), each and every outcome x_i of a random variable X has its own meaningful Shannon information value $h(x_i)$, that may be very different from that of another outcome x_j, although repeated draws from this random variable can be considered stationary. It is this sense of 'local' that gives *local* active information storage its name. In contrast, how locally in space and time we obtain the PDF is more important for the precision of the LAIS estimates.

In the analysis of LAIS from neural data three issues will necessarily blur locality, and impair the precision of the LAIS estimate to some extent:

1. If a pool of identical agents i, all running identical stationary random processes $\mathtt{X}_i$, is available, the only blurring of locality arises due to the intrinsic temporal extent of the state variables. However, the while the stored information may be encoded in a temporally non-local state $\mathbf{x}_{\mathbf{t}t-1}^{k-}$, this information is used to predict the next value of the process x_t at a *single* point in time.
2. If agents are non-identical, but their data are pooled nonetheless, then the overall empirical PDF obtained across these agents is no longer fully representative of each single agent and the local information storage values per agent are biased due to the use of this non-optimal PDF. This effect may be present to some extent in our analysis, as we cannot guarantee that all parts of area 18 behave strictly identical.
3. If the random process in question is not stationary, then a PDF obtained via pooling samples across time is also not representative of what happens at single points in time, and again a bias in the LAIS values for each agent and time step arises. This bias is potentially more severe. Nevertheless, we

pooled data across all available time samples here, as this seems to be closer to the strategy available to a neuron in a downstream brain area (also see section 2.2.5), when trying to estimate, or adapt to, its input distribution. This is because a neuron may more easily estimate approximate PDFs of its inputs across time than across all possible neurons in an upstream brain area, to most of which it simply doesn't interface.

4.7 CONCLUSION

Distributed information processing in neural systems can be decomposed into component processes of information transfer, storage and modification. Information storage can be quantified locally in space and time using an information theoretic measure termed local active information storage (LAIS). Here we present for the first time the application of this measure to neural data. We show that storage reflects neural properties such as stimulus preferences and surprise, and reflects the abstract concept of an ongoing stimulus despite the locally random nature of this stimulus. We suggest that LAIS will be a useful quantity to test theories of cortical function, such as predictive coding.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENT

The authors thank Matthias Kaschube from the Frankfurt Institute for Advanced Studies (FIAS) for fruitful discussions on active information storage.

Funding: MW, VP were supported by LOEWE Grant "Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF)". MW thanks the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for supporting a visit in Sydney which contributed to this work. SV was supported by the Bernstein Focus: Neurotechnology (BFNT) Frankfurt/M.

REFERENCES

Ay, N. and Polani, D. (2008), Information flows in causal networks, *Adv. Complex Syst.*, 11, 17

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012), Canonical microcircuits for predictive coding, *Neuron*, 76, 4, 695–711

549 Butts, D. A. (2003), How much information is associated with a particular stimulus?, *Network*, 14, 2,
550 177–187
551 Butts, D. A. and Goldman, M. S. (2006), Tuning curves, neuronal variability, and sensory coding., *PLoS*
552 *Biol*, 4, 4, e92, doi:10.1371/journal.pbio.0040092
553 Cover, T. M. and Thomas, J. A. (1991), Elements of information theory (Wiley-Interscience, New York,
554 NY, USA)
555 Crutchfield, J. P. and Feldman, D. P. (2003), Regularities unseen, randomness observed: Levels of entropy
556 convergence, *Chaos*, 13, 1, 25–54, doi:10.1063/1.1530990
557 Dasgupta, S., Wörgötter, F., and Manoonpong, P. (2013), Information dynamics based self-adaptive
558 reservoir for delay temporal memory tasks, *Evolving Systems*, 1–15
559 Deacon, T. W. (2010), Information and the Nature of Reality (Cambridge University Press), chapter 8
560 What is missing from theories of information?, 146
561 DeWeese, M. R. and Meister, M. (1999), How to measure the information gained from one symbol.,
562 *Network*, 10, 4, 325–340
563 Faes, L., Nollo, G., and Porta, A. (2012), Non-uniform multivariate embedding to assess the information
564 transfer in cardiovascular and cardiorespiratory variability series., *Comput Biol Med*, 42, 3, 290–297,
565 doi:10.1016/j.compbiomed.2011.02.007
566 Faes, L., Porta, A., Rossato, G., Adami, A., Tonon, D., Corica, A., et al. (2013), Investigating
567 the mechanisms of cardiovascular and cerebrovascular regulation in orthostatic syncope through an
568 information decomposition strategy, *Autonomic Neuroscience*
569 Fano, R. (1961), Transmission of information
570 Friston, K. (2005), A theory of cortical responses., *Philos Trans R Soc Lond B Biol Sci*, 360, 1456,
571 815–836, doi:10.1098/rstb.2005.1622
572 Gómez, C. and Hornero, R. (2010), Entropy and complexity analyses in alzheimers disease: An MEG
573 study, *The open biomedical engineering journal*, 4, 223
574 Granger, C. W. J. (1969), Investigating causal relations by econometric models and cross-spectral
575 methods., *Econometrica*, 37, 424–438
576 Kantz, H. and Schreiber, T. (2003), Nonlinear Time Series Analysis (Cambridge University Press), 2
577 edition
578 Knill, D. C. and Pouget, A. (2004), The bayesian brain: the role of uncertainty in neural coding and
579 computation., *Trends Neurosci*, 27, 12, 712–719, doi:10.1016/j.tins.2004.10.007
580 Kraskov, A., Stoegbauer, H., and Grassberger, P. (2004), Estimating mutual information., *Phys Rev E Stat*
581 *Nonlin Soft Matter Phys*, 69, 6 Pt 2, 066138
582 Langton, C. G. (1990), Computation at the edge of chaos: Phase transitions and emergent computation,
583 *Physica D: Nonlinear Phenomena*, 42, 1, 12–37

584 Lindner, M., Vicente, R., Priesemann, V., and Wibral, M. (2011), Trentool: A Matlab open source toolbox
585 to analyse information flow in time series data with transfer entropy., *BMC Neurosci*, 12(119), 1, 1–22,
586 doi:10.1186/1471-2202-12-119

587 Lizier, J. T. (2012), JIDT: An information-theoretic toolkit for studying the dynamics of complex systems,
588 `http://code.google.com/p/information-dynamics-toolkit/`

589 Lizier, J. T. (2013), The Local Information Dynamics of Distributed Computation in Complex Systems.,
590 Springer theses (Springer)

591 Lizier, J. T., Atay, F. M., and Jost, J. (2012a), Information storage, loop motifs, and clustered structure in
592 complex networks., *Phys Rev E Stat Nonlin Soft Matter Phys*, 86, 2 Pt 2, 026110

593 Lizier, J. T., Flecker, B., and Williams, P. L. (2013), Towards a synergy-based approach to measuring
594 information modification, in Proceedings of the 2013 IEEE Symposium on Artificial Life (ALIFE)
595 (IEEE), 43–51, doi:10.1109/alife.2013.6602430

596 Lizier, J. T. and Prokopenko, M. (2010), Differentiating information transfer and causal effect, *Eur. Phys.*
597 *J. B*, 73, 605–615

598 Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2012b), Local measures of information storage in
599 complex distributed computation, *Information Sciences*, 208, 39–54, doi:10.1016/j.ins.2012.04.016

600 MacKay, D. J. (2003), Information theory, inference and learning algorithms (Cambridge university press)

601 Mitchell, M. (1998), Computation in cellular automata: A selected review, in T. Gramß, S. Bornholdt,
602 M. Groß, M. Mitchell, and T. Pellizzari, eds., Non-Standard Computation (Wiley-VCH Verlag GmbH
603 & Co. KGaA, Weinheim), 95–140, doi:10.1002/3527602968.ch4

604 Mitchell, M., Hraber, P., and Crutchfield, J. P. (1993), Revisiting the edge of chaos: Evolving cellular
605 automata to perform computations, *arXiv preprint adap-org/9303003*

606 Obst, O., Boedecker, J., Schmidt, B., and Asada, M. (2013), On active information storage in input-driven
607 systems, arXiv:1303.5526

608 Pincus, S. M. (1991), Approximate entropy as a measure of system complexity., *Proceedings of the*
609 *National Academy of Sciences*, 88, 6, 2297–2301, doi:10.1073/pnas.88.6.2297

610 Porta, A., Baselli, G., Liberati, D., Montano, N., Cogliati, C., Gnecchi-Ruscone, T., et al. (1998),
611 Measuring regularity by means of a corrected conditional entropy in sympathetic outflow., *Biological*
612 *Cybernetics*, 78, 1, 71–78

613 Porta, A., Guzzetti, S., Montano, N., Pagani, M., Somers, V., Malliani, A., et al. (2000), Information
614 domain analysis of cardiovascular variability signals: evaluation of regularity, synchronisation and co-
615 ordination, *Medical and Biological Engineering and Computing*, 38, 2, 180–188

616 Priesemann, V., Munk, M. H. J., and Wibral, M. (2009), Subsampling effects in neuronal avalanche
617 distributions recorded in vivo., *BMC Neurosci*, 10, 40, doi:10.1186/1471-2202-10-40

Priesemann, V., Valderrama, M., Wibral, M., and Le Van Quyen, M. (2013), Neuronal avalanches differ from wakefulness to deep sleep–evidence from intracranial depth recordings in humans., *PLoS Comput Biol*, 9, 3, e1002985, doi:10.1371/journal.pcbi.1002985

Prokopenko, M., Boschietti, F., and Ryan, A. J. (2009), An Information-Theoretic primer on complexity, Self-Organization, and emergence, *Complexity*, 15, 1, 11–28

Prokopenko, M., Gerasimov, V., and Tanev, I. (2006), Evolving spatiotemporal coordination in a modular robotic system, in S. Nolfi, G. Baldassarre, R. Calabretta, J. C. T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, eds., From Animals to Animats 9: Proceedings of the Ninth International Conference on the Simulation of Adaptive Behavior (SAB'06), volume 4095 of *Lecture Notes in Computer Science* (Springer, Berlin Heidelberg), 558–569, doi:10.1007/11840541_46

Ragwitz, M. and Kantz, H. (2002), Markov models from data by simple nonlinear time series predictors in delay embedding spaces., *Phys Rev E Stat Nonlin Soft Matter Phys*, 65, 5 Pt 2, 056201

Rao, R. P. and Ballard, D. H. (1999), Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects., *Nat Neurosci*, 2, 1, 79–87, doi:10.1038/4580

Richman, J. S. and Moorman, J. R. (2000), Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology - Heart and Circulatory Physiology*, 278, 6, H2039–H2049

Schreiber (2000), Measuring information transfer, *Phys Rev Lett*, 85, 2, 461–464

Stetter, O., Battaglia, D., Soriano, J., and Geisel, T. (2012), Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals., *PLoS Comput Biol*, 8, 8, e1002653, doi:10.1371/journal.pcbi.1002653

Takens, F. (1981), Detecting strange attractors in turbulence, in D. Rand and L.-S. Young, eds., Dynamical Systems and Turbulence, Warwick 1980, volume 898 of *Lecture Notes in Mathematics* (Springer, Berlin / Heidelberg), chapter 21, 366–381

Turing, A. M. (1936), On computable numbers, with an application to the Entscheidungsproblem, *Proceedings of the London mathematical society*, 42, 2, 230–265

Turney, P. D. and Pantel, P. (2010), From frequency to meaning: Vector space models of semantics, *Journal of Artificial Intelligence Research*, 37, 141–188+, doi:10.1613/jair.2934

Vakorin, V. A., Mii, B., Krakovska, O., and McIntosh, A. R. (2011), Empirical and theoretical aspects of generation and transfer of information in a neuromagnetic source network., *Front Syst Neurosci*, 5, 96, doi:10.3389/fnsys.2011.00096

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011), Transfer entropy – a model-free measure of effective connectivity for the neurosciences., *J Comput Neurosci*, 30, 1, 45–67

Wang, X. R., Miller, J. M., Lizier, J. T., Prokopenko, M., and Rossi, L. F. (2012), Quantifying and tracing information cascades in swarms., *PLoS One*, 7, 7, e40084, doi:10.1371/journal.pone.0040084

653 Wibral, M., Pampu, N., Priesemann, V., Siebenhhner, F., Seiwert, H., Lindner, M., et al. (2013),
654 Measuring information-transfer delays., *PLoS One*, 8, 2, e55809, doi:10.1371/journal.pone.0055809
655 Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., and Kaiser, J. (2011), Transfer entropy
656 in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks.,
657 *Prog Biophys Mol Biol*, 105, 1-2, 80–97
658 Wiener, N. (1956), The theory of prediction., in E. F. Beckmann, ed., In Modern Mathematics for the
659 Engineer (McGraw-Hill, New York)
660 Zipser, D., Kehoe, B., Littlewort, G., and Fuster, J. (1993), A spiking network model of short-term active
661 memory, *The Journal of Neuroscience*, 13, 8, 3406–3420