

# Coordinate invariance as a fundamental constraint on the form of stimulus-specific information measures

Lubomir Kostal\* and Giuseppe D’Onofrio

*Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic*

The value of Shannon’s mutual information is commonly used to describe the total amount of information that the neural code transfers between the ensemble of stimuli and the ensemble of neural responses. In addition, it is often desirable to know which stimulus features or which response features are most informative. The literature offers several different decompositions of the mutual information into its stimulus or response-specific components, such as the specific surprise or the uncertainty reduction, but the number of mutually distinct measures is in fact infinite. We resolve this ambiguity by requiring the specific information measures to be invariant under invertible coordinate transformations of the stimulus and the response ensembles. We prove that the Kullback-Leibler divergence is then the only suitable measure of the specific information. On a more general level we discuss the necessity and the fundamental aspects of the coordinate invariance as a selection principle. We believe that our results will encourage further research into invariant statistical methods for the analysis of neural coding.

Keywords: Coordinate invariance, Mutual information, Specific information, Kullback-Leibler divergence

## 1. INTRODUCTION

An information theoretic framework has been applied successfully to a wide range of neuroscientific problems, as reviewed for example in Borst and Theunissen [1], Dimitrov *et al.* [2], McDonnell *et al.* [3], including a quantification of different aspects of information processing in neural systems [4–10]. Most frequently, the studies investigate the *mutual information* between stimuli and responses in order to quantify the amount of information the neural systems transfer. Formally, information-theoretic methods rely on the probabilistic description of both the stimulus ensemble and the neuronal system, which relates the stimulus feature, or intensity, to some particular response [1, 11, 12]. Since the amount of information transmitted depends on the stimulus probability distribution, the mutual information represents an indispensable mathematical tool for the description of the efficient coding hypothesis [13–17].

Mutual information, however, is an average quantity, which describes the total information transfer between the stimulus and response ensembles as a single number. It is often important to know *which* stimulus intensities or which response features are most informative. Several decompositions of the mutual information into its stimulus and response-specific components are being used in the literature [18–21]. These *specific information* measures generally provide mutually incompatible values, and there are various arguments favoring one measure over the others, such as in DeWeese and Meister [18], Bezzi [21], Yarrow *et al.* [22], also reviewed by Thomson and Kristan [23] and more recently by Wibral *et al.* [10]. Nevertheless, most of the studies employ multiple measures of the specific information, see, e.g., Bezzi [21], Olypher *et al.* [24], Butts and Goldman [25], Lansky *et al.* [26], Kastner and Baccus [27].

It has been demonstrated in a series of recent papers [28–30] that the choice of stimulus units has a deep impact on the inference about the neural decoding accuracy. Reference frame *invariance* is one of the cornerstones of modern physics [31], however, it is rarely considered in the field of computational neuroscience. For example, the amount of transmitted information is *a priori* expected to be independent from the model parameterization or the choice of measurement units. In fact, as shown in this paper, the individual specific information expressions may depend on the frame of reference even though the mutual information itself is an invariant quantity. We argue that the invariance principle is a logical necessity, which helps us to resolve the ambiguity in the choice of possible specific information measures.

The paper is organized as follows. After presenting the basic methodology, we employ the classical model of an auditory nerve fiber [32] responding to a natural-like sound intensity distribution [33, 34], to illustrate the dramatic impact of an apparently benign and frequently used information-preserving transformation of the response. We prove our central result in Theorem 1, which states that the requirement for transformation invariance significantly restricts the form of possible specific information measures. The Kullback-Leibler divergence [35] is shown to hold a unique position under mild and reasonable technical assumptions. We also identify special stimulus and response reference frames, in which the other popular specific information measures become equal to the Kullback-Leibler divergence.

Finally in Section 4, we discuss the implications of our findings from a broader point of view. We provide some alternative and indirect reasons to justify the prominent place of the Kullback-Leibler divergence within the specific decomposition of the mutual information. Ultimately, we speculate that the reference frame invariance should be considered a general and fundamental selection criterion whenever several possible methods are at hand. In particular, if the subjective choice of measurement units leads to contradictory inference, invariant

---

\* E-mail: kostal@biomed.cas.cz

method should be used instead. We hope that our conclusions will stimulate prospective research within this area, focusing not only on the decoding precision [29] or information-like aspects, but also on more general statistical methods used in computational neuroscience.

## 2. METHODS

Let  $s$  be the *stimulus* intensity, or feature, taking values in a continuous set of possibilities denoted as  $\mathcal{S}$ . The naturally unpredictable character of stimulation is formally described in terms of the random variable  $S$ , which is defined by the probability distribution (or density) function  $\pi_S(s)$ ,  $\int_{\mathcal{S}} \pi_S(s) ds = 1$  [12]. The function  $\pi_S$  assigns the probability that  $S$  falls within the interval  $[s, s + ds)$ ,  $\Pr(s \leq S \leq s + ds) = \pi_S(s) ds$ , to all elements of the stimulus sample space  $\mathcal{S}$ . The sample space  $\mathcal{S}$  and the probability distribution  $\pi_S$  taken together are denoted as the *stimulus ensemble* [36, p. 13].

The *response*,  $r$ , of the neural system to some stimulus,  $s$ , is typically modeled stochastically [37–40]. Whether  $r$  is the firing frequency, spike timing precision or some other characteristics [11, 41–44] is largely immaterial for information-theoretic methodology. In the following we assume that  $r$  takes values in a continuous set  $\mathcal{R}$  and that the neural model, i.e., the information channel, is generally defined by the conditional distribution function  $f_{R|S}(r|s)$  of the response random variable  $R$ . The marginal (or unconditional) distribution function,  $p_R(r)$  is given by

$$p_R(r) = \int_{\mathcal{S}} f_{R|S}(r|s) \pi_S(s) ds. \quad (1)$$

The critical quantity generally required for the implementation of the information-theoretic framework is the *joint* probability distribution of both stimuli and responses,

$$f_{S,R}(s, r) = f_{R|S}(r|s) \pi_S(s). \quad (2)$$

The *mutual information*,  $I(S; R)$ , between stimuli and responses is then

$$I(S; R) = \int_{\mathcal{S}} \int_{\mathcal{R}} f_{S,R}(s, r) \log \frac{f_{S,R}(s, r)}{\pi_S(s) p_R(r)} dr ds. \quad (3)$$

The value of  $I(S; R)$  is the maximum amount of information that can be communicated reliably, under as low a probability of decoding error as desired, through the information channel  $f_{R|S}$  subject to the input statistics  $\pi_S$  [36]. Traditionally,  $I(S; R)$  is used in neuroscience to describe how much information the neural code transfers about the stimulus ensemble, e.g., how much the sequence of postsynaptic currents tells us about the input train of action potentials [1, 4, 7, 9, 45–47]. The mutual information is not causal (or directional), it generally holds  $I(S; R) = I(R; S)$  [36].

In order to gain insight into how much information the responses of a neuron convey about a particular stimulus, the *stimulus-specific information*,  $i(s; R)$ , is implicitly defined as

$$I(S; R) = \int_{\mathcal{S}} \pi_S(s) i(s; R) ds, \quad (4)$$

see, e.g., DeWeese and Meister [18] or Bezzi [21]. The units of  $i(s; R)$  are therefore the same as that of  $I(S; R)$ , that is 'nats' for a natural logarithm in Eq. (3) or 'bits' if the logarithm base is 2. This quantity is typically computed for sensory neurons that selectively respond to sensory stimuli [48]. Analogously to the stimulus-specific information in Eq. (4) one defines the *response-specific* information,  $i(r; S)$ , by

$$I(S; R) = \int_{\mathcal{R}} p_R(r) i(r; S) dr. \quad (5)$$

Neither the decomposition in Eq. (4) nor the one in Eq. (5) determines the specific information uniquely.

The most frequent expression for the specific information, denoted in the following as  $i_h$ , is based on the difference of two entropy (uncertainty) values [10, 21, 26, 49]. The average response uncertainty  $h(R)$  across the whole stimulus ensemble is given by the marginal response distribution from Eq. (1),

$$h(R) = - \int_{\mathcal{R}} p_R(r) \log p_R(r) dr, \quad (6)$$

while the uncertainty in the response given some particular stimulus value is

$$h(R|s) = - \int_{\mathcal{R}} f_{R|S}(r|s) \log f_{R|S}(r|s) dr. \quad (7)$$

The stimulus-specific information  $i_h(s; R)$  is then the difference

$$i_h(s; R) = h(R) - h(R|s). \quad (8)$$

An analogous expression to Eq. (8) it also employed outside the neuroscientific context, e.g., in Dubuis *et al.* [50].

The second popular expression for the stimulus-specific information,  $i_{KL}(s; R)$ , is based on the concept of the Kullback-Leibler divergence [35, 51], and is often denoted as the 'specific surprise' [18, 20, 21, 49],

$$i_{KL}(s; R) = \int_{\mathcal{R}} f_{R|S}(r|s) \log \frac{f_{R|S}(r|s)}{p_R(r)} dr. \quad (9)$$

Finally, Butts [19] argued in favor of the stimulus-specific information  $i_{ssi}(s; R)$ , which is based on the analogue of Eq. (8) for the response-specific case,

$$\begin{aligned} i_{ssi}(s; R) &= \int_{\mathcal{R}} i_h(r; S) f_{R|S}(r|s) dr \\ &= \int_{\mathcal{R}} f_{R|S}(r|s) \int_{\mathcal{S}} \left[ f_{S|R}(\tilde{s}|r) \log f_{S|R}(\tilde{s}|r) \right. \\ &\quad \left. - \pi_S(\tilde{s}) \log \pi_S(\tilde{s}) \right] d\tilde{s} dr. \end{aligned} \quad (10)$$

Note that any weighted combination of  $i_h$ ,  $i_{KL}$  and  $i_{ssi}$  yields a "new" valid formula for the specific information.

The main result of this paper relies on the description of the *same* neuronal information transfer in *different* coordinate systems. That is, we investigate the impact of the choice of measurement units for the stimulus and the response. Such choice is formally described in terms of functions

$$x = \varphi(s), \quad (11)$$

$$y = \xi(r), \quad (12)$$

which map the stimuli and response ensembles as  $\varphi : \mathcal{S} \rightarrow \mathcal{X} = \varphi(\mathcal{S})$  and  $\xi : \mathcal{R} \rightarrow \mathcal{Y} = \xi(\mathcal{R})$ . We restrict ourselves to *regular* transformations of the reference frame, so that  $\varphi, \xi$  are *strictly monotonic* and *differentiable* within the interior of their domains of definition [29]. The corresponding probability density functions  $\pi_X(x), f_{Y|X}(y|x)$  in the regularly transformed units are given by [52]

$$\pi_X(x) = \pi_S(\varphi^{-1}(x)) |\dot{\varphi}^{-1}(x)|, \quad (13)$$

$$f_{Y|X}(y|x) = f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) |\dot{\xi}^{-1}(y)|, \quad (14)$$

where  $\dot{\varphi}^{-1}(x) = d\varphi^{-1}(x)/dx$  and  $\varphi^{-1}$  denotes the inverse function to  $\varphi$  (and similarly for  $\dot{\xi}^{-1}$ ).

The *reference frame invariance* means that the values of the stimulus-specific information evaluated in the  $(s, r)$ -frame and in the transformed  $(x, y)$ -frame are equivalent,

$$i(s; R) = i(x; Y) \Big|_{\substack{x=\varphi(s) \\ Y=\xi(R)}}. \quad (15)$$

The mutual information itself is well known to be invariant with respect to invertible transformations of the reference frame,  $I(S; R) = I(X; Y)$  [36, 53]. The same arguments for proving the invariance apply also to the generalized conditional mutual information, i.e., mutual information conditioned on the  $\sigma$ -algebra generated by the random variable [54]. Therefore, non-symmetric (causal) information measures based on the conditional mutual information, such as transfer entropy [55] and directed information [56], are also invariant.

### 3. RESULTS

#### 3.1. Example: Non-invariance leads to contradictory inference

We illustrate the key problem of non-invariant information measures, which yields incongruent inference about the specific information from equal stimulus intensities. As an example we choose the joint stimulus-response probability distribution  $f_{S,R}$  constructed as follows.

The conditional distribution  $f_{R|S}$  is derived from the classical model of a cat auditory nerve fiber responding to the characteristic 8 kHz sound frequency [32]. The mean rate-level function,  $\mu(s)$ , of the model is

$$\mu(s) = \frac{10^{cs/20} r_m}{10^{t_E/20} (1 + 10^{-t_I/20} 10^{s/10})^{c/3} + 10^{cs/20}} + r_{sp}, \quad (16)$$

where the stimulus,  $s$ , is the sound pressure level of pure tones and  $\mu(s)$  is the average firing frequency (rate), i.e., the number of action potentials per second (Fig. 1, solid white line). The parameter values corresponding to the nerve fiber with low spontaneous activity [32] are given as:  $t_E = 89.4$  dB SPL is the response threshold,  $t_I = 100$  dB SPL is a parameter related to the two-tone suppression model [57],  $r_m = 135.1$  is the maximum rate change that can be observed in response to a pure tone stimulus,  $r_{sp} = 0.5$  is the spontaneous discharge rate, and  $c = 1.77$  is a model constant. The response spike count distribution in auditory nerve fibers is approximately

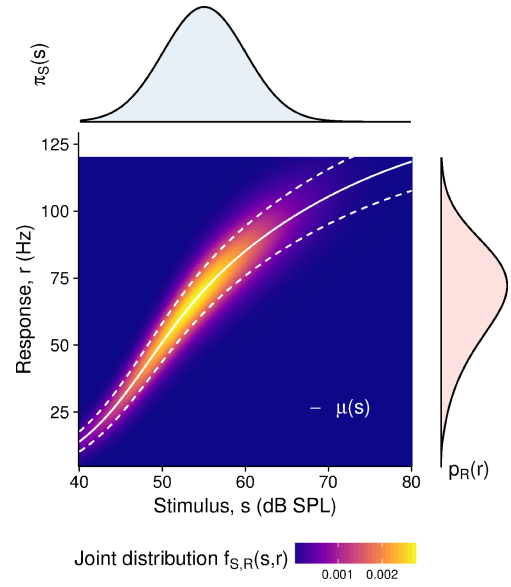


Figure 1. Probabilistic description of the stimulus-response model. The average firing frequency as a function of the sound intensity (solid white line) of a cat auditory nerve fiber best responding to 8 kHz pure tones [32]. The spread of standard deviation from the mean response (dashed white line) is Poisson-like, the conditional response distribution is Gaussian (Eq. 17). The stimulus distribution  $\pi_S(s)$  (top) approximates the sound intensity ensemble in a natural environment. The joint stimulus-response distribution  $f_{S,R}(s, r)$  (color) provides the complete description from the information-theoretic point of view (Eq. 3). The marginal response distribution  $p_R(r)$  (right) is calculated from Eq. (1), and the mutual information is  $I(S; R) \doteq 1.19$  bit (Eq. 3).

Poisson [32, 58]. For simplicity we assume that the response distribution is Gaussian with mean  $\mu(s)$  and standard deviation  $\sigma(s) = \sqrt{\mu(s)}$  (Fig. 1, dashed lines),

$$f_{R|S}(r|s) = \frac{1}{\sqrt{2\pi\mu(s)}} \exp\left(-\frac{[r - \mu(s)]^2}{2\mu(s)}\right). \quad (17)$$

The stimulus distribution  $\pi_S(s)$  describes the typical ensemble of sound intensities, such as animal vocalizations in the natural environment, which is approximately modeled as a Gaussian on the sound pressure level scale [34]. We set  $\pi_S$  to be normal with mean 55 dB SPL and standard deviation 5 dB SPL (Fig. 1), in accord with Wen *et al.* [33], Sun *et al.* [34].

The complete joint probability distribution,  $f_{S,R}(s, r)$ , then follows from Eq. (2). The amplitude of  $f_{S,R}(s, r)$  is visualised by the 'heatmap' plot in Fig. 1, together with the marginal response distribution,  $p_R(r)$ , given by Eq. (1). The value of mutual information for this system is  $I(S; R) \doteq 1.19$  bit (Eq. 3).

Next we examine the impact of regular coordinate transformations in Eqs. (11) and (12) on the stimulus-specific measures of information. Since regular transformations are bijective, they are also intuitively expected to be information-preserving. However, it turns out that out of Eqs. (8)–(10) only  $i_{KL}(s; R)$  is invariant in the sense of Eq. (15),

$$i_{KL}(\varphi(s); \xi(R)) = i_{KL}(s; R), \quad (18)$$

while for the other measures it holds

$$i_h(\varphi(s); \xi(R)) = i_h(s; R) + \int_{\mathcal{R}} \log |\dot{\xi}(r)| (p_R(r) - f_{R|S}(r|s)) dr, \quad (19)$$

$$i_{\text{ssi}}(\varphi(s); \xi(R)) = i_{\text{ssi}}(s; R) + \int_{\mathcal{R}} \int_{\mathcal{S}} \log |\dot{\varphi}(\tilde{s})| \times (\pi_S(\tilde{s}) - f_{S|R}(\tilde{s}|r)) f_{R|S}(r|s) d\tilde{s} dr. \quad (20)$$

Indeed, from the definition of  $i_{\text{KL}}$  in Eq. (9), and by using Eqs. (13) and (14), one has

$$i_{\text{KL}}(x; Y) = \int_{\mathcal{Y}} f_{Y|X}(y|x) \log \frac{f_{Y|X}(y|x)}{p_Y(y)} dy = \int_{\mathcal{Y}} f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) |\dot{\xi}^{-1}(y)| \times \log \frac{f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) |\dot{\xi}^{-1}(y)|}{p_R(\xi^{-1}(y)) |\dot{\xi}^{-1}(y)|} dy. \quad (21)$$

Eq. (18) then follows from  $dr = |\dot{\xi}^{-1}(y)| dy$  and Eqs. (11) and (12). Similarly, from the definition in Eq. (8),

$$i_h(x; Y) = - \int_{\mathcal{Y}} p_R(\xi^{-1}(y)) |\dot{\xi}^{-1}(y)| \log [p_R(\xi^{-1}(y)) \times |\dot{\xi}^{-1}(y)|] dy + \int_{\mathcal{Y}} f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) \times |\dot{\xi}^{-1}(y)| \log [f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) |\dot{\xi}^{-1}(y)|] dy = i_h(s; R) \Big|_{\substack{s=\varphi^{-1}(x) \\ R=\xi^{-1}(Y)}} + \int_{\mathcal{Y}} |\dot{\xi}^{-1}(y)| \log |\dot{\xi}^{-1}(y)| \times [f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) - p_R(\xi^{-1}(y))] dy, \quad (22)$$

and Eq. (19) follows since  $\dot{\xi}(\xi^{-1}(y)) = 1/\dot{\xi}^{-1}(y)$ . By the same argument, Eq. (20) follows from the expression

$$i_{\text{ssi}}(x; Y) = \int_{\mathcal{Y}} f_{R|S}(\xi^{-1}(y)|\varphi^{-1}(x)) |\dot{\xi}^{-1}(y)| \times \int_{\mathcal{X}} \left\{ f_{S|R}(\varphi^{-1}(\tilde{x})|\xi^{-1}(y)) \times \log [f_{S|R}(\varphi^{-1}(\tilde{x})|\xi^{-1}(y)) |\dot{\varphi}^{-1}(\tilde{x})|] - \pi_S(\varphi^{-1}(\tilde{x})) \times \log [\pi_S(\varphi^{-1}(\tilde{x})) |\dot{\varphi}^{-1}(\tilde{x})|] \right\} |\dot{\varphi}^{-1}(\tilde{x})| d\tilde{x} dy. \quad (23)$$

Note that for every regular transformation  $\varphi(s)$  of the stimulus and for every *linear* transformation of the response  $\xi(r) = ar + b$ , with  $a \neq 0$  and  $b$  real, the integral on the right-hand side of Eq. (19) is zero, i.e.,  $i_h(s; R)$  is invariant with respect to linear transformations of the response. By similar arguments,  $i_{\text{ssi}}(s; R)$  is invariant with respect to linear transformations of the stimulus.

In order to highlight the (apparently paradoxical) situation, in which the information content within the same neuronal system

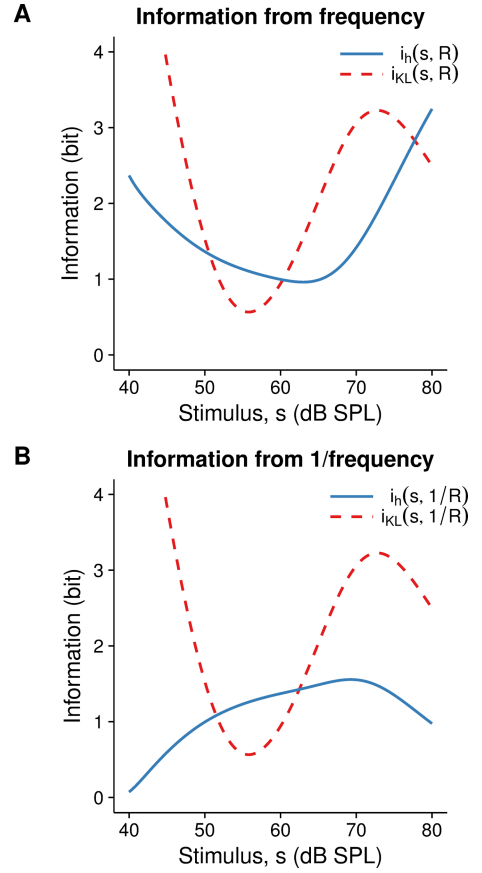


Figure 2. Non-invariant stimulus-specific information measures yield incompatible results under invertible, information-preserving transformations. **(A)** Information gained from sound intensities (dB SPL) in the auditory nerve fiber model (Fig. 1), where neuronal response is the firing frequency. Two classic information measures are shown: the uncertainty reduction ( $i_h$ , Eq. 8, solid line) and the specific surprise ( $i_{\text{KL}}$ , Eq. 9, dashed line). The quantities  $i_h$  and  $i_{\text{KL}}$  are known to have distinct properties, therefore the dependence of their profiles on the stimulus  $s$  differ. **(B)** Neuronal response is the inverse value of the firing frequency, which is invertible, hence no information loss is expected. However, the profile of  $i_h$  changes dramatically between (A) and (B), e.g., the least informative stimulus in (A) becomes approximately the most informative in (B). The explanation is that  $i_h$  depends not only on the stimulus-response model *per se* but also on the arbitrarily chosen frame of reference (Eq. 19). On the other hand, the coordinate invariance of  $i_{\text{KL}}$  guarantees consistent results across different reference frames (Eq. 18).

depends on the arbitrarily chosen reference frame, we choose to inversely rescale the response value,

$$\xi(r) = \frac{1}{r}, \quad (24)$$

while keeping the stimulus scale unaffected,  $\varphi(s) = s$ . The transformation in Eq. (24) is strictly monotonic and differentiable for  $r > 0$ . Theoretically, no information is being lost if we use  $y = 1/r$  instead of  $r$  as the response value. Heuristically, the particular choice for the inverse scaling in Eq. (24) can be motivated by the general relationship between the mean

firing frequency and the mean interspike interval. Let  $\langle N(\Delta) \rangle$  denote the expected number of spikes within some time window  $\Delta$ , and let  $\langle T \rangle$  be the mean interval between two spikes. By describing the neuronal firing as a stationary point process in equilibrium it holds [37, 59]

$$\frac{\langle N(\Delta) \rangle}{\Delta} = \frac{1}{\langle T \rangle}. \quad (25)$$

Therefore,  $1/\mu(s)$  from Eq. (17) can be thought of as being the mean interspike interval. Note though that  $1/r$  itself cannot be interpreted as the actual interspike interval [42].

The information measures  $i_h(s; R)$  and  $i_{\text{KL}}(s; R)$ , for the model in Fig. 1, are compared in Fig. 2. The profiles of the two measures as a function of sound intensity differ, consistent with their distinct properties [10, 18, 21, 25]. Since the response transformation in Eq. (24) is invertible, no change in information gain is expected when using the  $(s, 1/r)$ -frame instead of the  $(s, r)$ -frame of reference. However, the profile of  $i_h$  changes significantly between Fig. 2A and Fig. 2B. Particularly, within the interval [40, 60] dB SPL, the function  $i_h(s, R)$  is monotonically decreasing while  $i_h(s, 1/R)$  is monotonically increasing, and the least informative stimulus in Fig. 2A becomes the most informative one in Fig. 2B, approximately. The explanation of the 'paradox' relies on Eq. (19), which shows that  $i_h$  depends not only on the stimulus-response model, as a physical object, but also on the arbitrary frame of reference for the response. On the other hand the coordinate invariance of  $i_{\text{KL}}$  proves to be crucial, guaranteeing consistent results across different coordinate systems.

We also note that the logarithmic rescaling  $\varphi(s) \propto \log(s)$ , is frequently employed whenever the coding range of the stimulus spans several orders of magnitude [60–62]. The impact of such stimulus rescaling on the decoding accuracy in the model given by Eq. (17) was analyzed in Kostal and Lansky [28], and more generally in Kostal [29]. Analogously to Eq. (24), we may consider the transformation  $\xi(r) = \log(r)$  of the response. The measures of stimulus-specific information  $i_h(s, R)$  and  $i_h(s, \log(R))$  again exhibit different behaviors, qualitatively similar to Fig. 2 (not shown).

### 3.2. Invariant specific information measures

As demonstrated by Eqs. (18)–(20),  $i_{\text{KL}}(s; R)$  is the only invariant stimulus-specific information measure among the three possibilities defined by Eqs. (8)–(10). In the following we prove a stronger statement, however, that the reference frame invariance determines  $i_{\text{KL}}$  uniquely, under certain mild assumptions. To the best of our knowledge, this claim constitutes a novel finding.

Before proceeding to our main result, some preliminary considerations are appropriate. The mutual information  $I(S; R)$  has the same value if one replaces, in Eq. (4),  $i(s; R)$  with the specific information

$$i'(s; R) = \alpha i(s; R) - (\alpha - 1)I(S; R) + \Lambda(s), \quad (26)$$

where  $\alpha$  is a real number and the function  $\Lambda(s)$  is arbitrary, except for the constraint

$$\int_S \pi_S(s) \Lambda(s) ds = 0. \quad (27)$$

If  $i(s; R)$  is invariant under coordinate transformations, so is  $i'(s; R)$  given by Eq. (26). However,  $\Lambda(s)$  implicitly depends on  $\pi_S$ , making  $i'(s; R)$  not universal, therefore it is reasonable to set  $\Lambda(s) = 0$ . The class of possible measures defined by Eq. (26) then simplifies:  $i'(s; R)$  is either constant (equal to  $I(S; R)$  for  $\alpha = 0$ ) or proportional to the original  $i(s; R)$  plus a constant. For convenience, we will consider only such  $i(s; R)$  that do not contain any constant term.

Our main result can be formulated as follows.

**Theorem 1.** *Assume that  $i(s; R)$  can be expressed as a functional*

$$i(s; R) = \int_{\mathcal{R}} g(\pi_S(s), f_{R|S}(r|s), p_R(r), f_{S|R}(s|r)) dr, \quad (28)$$

where  $g$  is a suitable function. Then the only coordinate-invariant measure of specific information, i.e., the function  $i(s; R)$  for which Eqs. (4) and (15) hold, is given by Eq. (9),

$$i(s; R) = i_{\text{KL}}(s; R) = \int_{\mathcal{R}} f_{R|S}(r|s) \log \frac{f_{R|S}(r|s)}{p_R(r)} dr. \quad (29)$$

*Proof.* It generally holds

$$f_{S|R}(s|r) = \frac{f_{R|S}(r|s)\pi_S(s)}{p_R(r)}, \quad (30)$$

therefore it is sufficient to consider  $g$  as a function of  $\pi_S(s)$ ,  $f_{R|S}(r|s)$  and  $p_R(r)$  only. Due to the substitution rule it holds

$$i(s; R) = \int_{\mathcal{R}} g[\pi_X(\varphi(s))\dot{\varphi}(s), f_{Y|X}(\xi(r)|\varphi(s))\dot{\xi}(r), p_Y(\xi(r))\dot{\xi}(r)] dr, \quad (31)$$

and therefore, by the inverse function theorem  $\dot{\xi}(\xi^{-1}(y)) = 1/\dot{\xi}^{-1}(y)$ , we obtain

$$i(s; R) \Big|_{\substack{s=\varphi^{-1}(x) \\ R=\xi^{-1}(y)}} = \int_{\mathcal{Y}} \dot{\xi}^{-1}(y) \times g \left[ \frac{\pi_X(x)}{\dot{\varphi}^{-1}(x)}, \frac{f_{Y|X}(y|x)}{\dot{\xi}^{-1}(y)}, \frac{p_Y(y)}{\dot{\xi}^{-1}(y)} \right] dy \quad (32)$$

By definition, it also holds

$$i(x; Y) = \int_{\mathcal{Y}} g[\pi_X(x), f_{Y|X}(y|x), p_Y(y)] dy, \quad (33)$$

and due to the invariance property in Eq. (15) the right-hand sides in Eqs. (32) and (33) must be equal. Since the transformation  $\varphi$  is arbitrary, it must be that the function  $g$  does not depend on the distribution  $\pi_X$ .

We introduce the function  $\tilde{g}$  as

$$\tilde{g}\left(f_{R|S}(r|s), \frac{f_{R|S}(r|s)}{p_R(r)}\right) = g(f_{R|S}(r|s), p_R(r)), \quad (34)$$

noting that  $p_R(r) = 0$  also implies  $f_{R|S}(r|s) = 0$  for all  $s$  so that the ratio is well defined everywhere. By using Eq. (34) in Eqs. (32) and (33), and by comparing their right-hand sides we see that it must hold

$$\begin{aligned} \tilde{g}\left(\frac{f_{Y|X}(y|x)}{\dot{\xi}^{-1}(y)}, \frac{f_{Y|X}(y|x)}{p_Y(Y)}\right) \\ = \frac{1}{\dot{\xi}^{-1}(y)} \tilde{g}\left(f_{Y|X}(y|x), \frac{f_{Y|X}(y|x)}{p_Y(Y)}\right), \end{aligned} \quad (35)$$

almost everywhere. However, Eq. (35) is satisfied only if  $\tilde{g}$  is such that

$$\tilde{g}\left(f_{R|S}(r|s), \frac{f_{R|S}(r|s)}{p_R(r)}\right) = f_{R|S}(r|s) q\left(\frac{f_{R|S}(r|s)}{p_R(r)}\right), \quad (36)$$

where the  $q$  is a function of the ratio  $f_{R|S}(r|s)/p_R(r)$ . Substituting Eq. (36) into Eq. (28), and comparing Eq. (4) with the definition of mutual information in Eq. (3), yields  $q(\cdot) = \log(\cdot)$  as the only possible function, hence proving Eq. (29).  $\square$

The information measures similar to  $i_{\text{ssi}}$  in Eq. (10) contain double integration, which is not covered by Eq. (28). However, a small extension of Theorem 1 allows us to treat them as well. The following corollary is a direct consequence of Theorem 1 applied to the case of response-specific information defined by Eq. (5).

**Corollary 1.** *Let  $i(r; S)$  satisfy Eq. (5) and*

$$i(r; S) = \int_S g(\pi_S(s), f_{R|S}(r|s), p_R(r), f_{S|R}(s|r)) ds, \quad (37)$$

where  $g$  is a suitable function. Then the only coordinate-invariant measure of the response-specific information is

$$i(r; S) = i_{\text{KL}}(r; S) = \int_S f_{S|R}(s|r) \log \frac{f_{S|R}(s|r)}{\pi_S(s)} ds. \quad (38)$$

The invariance conditions for the class of stimulus-specific information measures similar to Eq. (10) then follow from the Lemma 1.

**Lemma 1.** *The stimulus-specific information*

$$i(s; R) = \int_{\mathcal{R}} i(r; S) f_{R|S}(r|s) dr \quad (39)$$

is invariant if and only if the response-specific information  $i(r; S)$  is given by Eq. (38).

*Proof.* The invariance condition together with the substitution rule require

$$\begin{aligned} i(s; R) &= \int_{\mathcal{R}} i(r; S) f_{Y|X}(\xi(r)|\varphi(s)) |\dot{\xi}(r)| dr \\ &= \int_{\mathcal{Y}} i(y; X) f_{Y|X}(y|x) \Big|_{x=\varphi(s)} dy \\ &= i(x; Y) \Big|_{\substack{x=\varphi(s) \\ Y=\xi(R)}} \end{aligned} \quad (40)$$

However, Eq. (40) holds only if

$$i(r; S) = i(y; X) \Big|_{\substack{x=\varphi(s) \\ y=\xi(r)}}, \quad (41)$$

that is, if the response-specific information is invariant, given by Eq. (38).  $\square$

Finally, we remark that is possible to introduce stimulus-specific measures that are defined by multiple integrals over  $\mathcal{S}$  and  $\mathcal{R}$ . Consider the following equality, based on  $\pi_S(s) = \int_{\mathcal{R}} f_{S|R}(s|r) p_R(r) dr$  and on the re-labeling of integration variables,

$$\begin{aligned} I(S; R) &= \int_S \pi_S(s) i(s; R) ds = \\ &= \int_{\mathcal{R}} p_R(r) \int_S f_{S|R}(s|r) i(s; R) ds dr \\ &= \int_S \pi_S(s) \int_{\mathcal{R}} f_{R|S}(r|s) \\ &\quad \times \int_S f_{S|R}(\tilde{s}|r) i(\tilde{s}; R) d\tilde{s} dr ds. \end{aligned} \quad (42)$$

Thereupon Eq. (42) can be used to define a new specific information,  $i'(s; R)$ , as a functional of the original  $i(s; R)$ . By employing Bayes rule,

$$i'(s; R) = \int_{\mathcal{R}} f_{R|S}(r|s) \int_S \frac{f_{R|S}(r|\tilde{s}) \pi_S(\tilde{s})}{p_R(r)} i(\tilde{s}; R) d\tilde{s} dr. \quad (43)$$

The process outlined in Eq. (42) can be iterated indefinitely, producing a sequence of mutually different stimulus-specific information measures  $i_n(s; R)$ ,  $n \geq 1$ , where

$$i_n(s_n; R) = \int i_0(s_0; R) \prod_{i=0}^{n-1} f_{R|S}(r_i|s_{i+1}) f_{S|R}(s_i|r_i) ds_i dr_i. \quad (44)$$

The quantity  $i_n(s; R)$  is mentioned for completeness only as its practical relevance seems limited. By repeated application of Lemma 1 it follows that  $i_n(s; R)$  is invariant only if  $i_0(s; R)$  is given by Eq. (29).

Consequently, it is generally impossible to find transformations  $\varphi, \xi$  such that any two different stimulus- or response-specific measures of information would become numerically equal. However, it is possible to find these coordinate transformations for the simplest case. It follows from Eq. (19) that there exists the response transformation  $y = \xi(r)$  such that

$$i_{\text{KL}}(s; Y) = i_h(s; Y) \quad (45)$$

holds for all stimuli in the transformed coordinates, even if  $i_{\text{KL}}(s; R) \neq i_h(s; R)$  in the original coordinates. By requiring Eq. (45) and substituting from Eqs. (8) and (9) into Eq. (19) we obtain that it must be

$$\xi(r) = \int_{-\infty}^r p_R(\tilde{r}) d\tilde{r}. \quad (46)$$

In other words, Eq. (45) holds if the transformed marginal response variable  $Y$  is uniform on  $[0, 1]$ . By symmetry, the stimulus transformation  $x = \varphi(s)$  of the form

$$\varphi(s) = \int_{-\infty}^s \pi_S(\tilde{s}) d\tilde{s} \quad (47)$$

yields

$$i_{\text{KL}}(r; X) = i_h(r; X) \quad (48)$$

for the transformed response-specific information. Applying the same process to the requirement, e.g.,  $i_{\text{KL}}(s; Y) = i_{\text{ssi}}(s; Y)$  reveals that generally no such  $\xi$  exists. The right-hand side of Eq. (20) can be written as  $\int_{\mathcal{R}} f_{R|S}(r|s)g(r) dr$ , where  $g(r)$  is a function of  $r$  only, while Eq. (9) cannot be written in this way.

#### 4. DISCUSSION

The requirement for the invariance of physical laws under invertible and differentiable coordinate transformations has a long history, and represents one of the foundations of modern physical theories [31, 63]. Generally, inferences and predictions should be independent from the arbitrarily and freely chosen frame of reference. In the context of this paper, the probability distributions  $f_{R|S}(r|s)$  and  $f_{Y|X}(y|x)$  describe *the same* neuronal model, just as  $\pi_S(s)$  and  $\pi_X(x)$  describe the same stimulus ensemble, as a physical object. Therefore, there is no persuasive preference for the  $(s, r)$ -parameterization over the  $(x, y)$ -parameterization, perhaps besides convenience. Furthermore, the value of specific information is generally given in bits (or nats), which is by definition independent from either of the measurement units of the stimulus or the response. The transformed value,  $i(x; Y)$ , should therefore not depend on  $\varphi$ ,  $\xi$  or their derivatives, in principle. Therefore the reference frame independence of  $i(s; R)$  or  $i(r; S)$  is not only natural, but in fact a necessary requirement.

Although the invariance property of  $i_{\text{KL}}$ , proven in Theorem 1, is sufficient to justify its prominence among the possible measures of specific information, there are other, perhaps indirect, reasons why  $i_{\text{KL}}$  measures the informative importance of individual stimuli intensities uniquely.

The measure  $i_{\text{KL}}(s; R)$  describes not only how informative are the individual stimulus values in the ensemble, but also how *sensitive* is the value of  $I(S; R)$  to a small change in the probability of  $S \in [s, s + ds]$ . To see this we calculate the functional derivative of  $I(S; R)$  with respect to the stimulus distribution  $\pi_S$ . Let  $\mathcal{F}$  be the convex set of all possible stimulus distribution functions, and denote  $I[\pi_S] \equiv I(S; R)$  to emphasise the functional dependence. The directional derivative  $\delta_g I[\pi_S]$  of  $I(S; R)$  at a given  $S \sim \pi_S$  in the 'direction' of the distribution function  $g \in \mathcal{F}$  is defined as

$$\delta_g I[\pi_S] = \lim_{\varepsilon \downarrow 0} \frac{I[(1 - \varepsilon)\pi_S + \varepsilon g] - I[\pi_S]}{\varepsilon}, \quad (49)$$

where  $\varepsilon \in [0, 1]$  [7, 64, 65]. We evaluate Eq. (49) by substituting Eqs. (1) and (2) into Eq. (3), and after changing the order

of the limit and the integration, the result can be manipulated into the form

$$\delta_g I[\pi_S] = \int_S (g(s) - \pi_S(s)) i_{\text{KL}}(s; R) ds. \quad (50)$$

That is, the directional perturbation of the stimulus distribution function changes the total information to be

$$I[(1 - \varepsilon)\pi_S + \varepsilon g] \doteq (1 - \varepsilon)I[\pi_S] + \varepsilon \int_S g(s) i_{\text{KL}}(s; R) ds, \quad (51)$$

for  $\varepsilon$  small enough.

Second, assuming that the set  $\mathcal{F}$  of all possible stimulus distributions is both convex and compact, and  $I[\pi_S]$  a strictly convex functional [7, 36, 64], there exists a unique  $\pi_0 \in \mathcal{F}$  such that the value of mutual information attains its maximum, denoted as

$$C = I[\pi_0] = \max_{\pi_S \in \mathcal{F}} I[\pi_S]. \quad (52)$$

The necessary and sufficient condition for such  $\pi_0$  is that  $\delta_g I[\pi_0] \leq 0$  for all  $g \in \mathcal{F}$  [64]. The value of  $i_{\text{KL}}(s; R)$  can thus be used to check whether the maximum of  $I(S; R)$  has been attained. Due to Eq. (50) the maximum is attained if and only if (see also Gallager [36, p. 91])

$$i_{\text{KL}}(s; R) \leq C, \quad (53)$$

with equality for all  $s$  in the support of  $\pi_0(s)$ , assuming that Eq. (1) is evaluated at  $\pi_S = \pi_0$ . In other words, the maximum value of  $I(S; R)$  is achieved only if all stimulus values with nonzero probability are *equally informative* in terms of  $i_{\text{KL}}$ . The condition in Eq. (53) can be generalized to the optimization of  $I(S; R)$  subject to average values of certain constraining functions, e.g., to the metabolic cost of neuronal activity [8, 9, 47, 66, 67]. If  $w(s)$  is the metabolic cost associated with stimulus intensity  $s$ , and the goal is to maximize  $I(S; R)$  subject to  $\int \pi_S(s)w(s) ds \leq W$ , then, due to Kuhn-Tucker conditions [68], Eq. (53) becomes

$$i_{\text{KL}}(s; R) \leq c + \lambda w(s), \quad (54)$$

where  $c$  is a constant and  $\lambda \geq 0$  is the Lagrange multiplier [69, p. 147].

Finally, we remark that  $i_{\text{KL}}(s, R)$  also exhibits the desirable property of *additivity*, i.e., the information accumulates over subsequent response observations to a given stimulus. If  $R_1$  and  $R_2$  are two generally non-independent response random variables (given some stimulus intensity  $S = s$ ) it holds [35, Theorem 2.2]

$$i_{\text{KL}}(s; \{R_1, R_2\}) = i_{\text{KL}}(s; R_1) + i_{\text{KL}}(s; R_2|R_1), \quad (55)$$

by using the factorization of the joint probability distributions,  $f_{R_1, R_2|S}(r_1, r_2|s) = f_{R_2|S, R_1}(r_2|s, r_1)f_{R_1|S}(r_1|s)$  and  $p_{R_1, R_2}(r_1, r_2) = p_{R_2|R_1}(r_2|r_1)p_{R_1}(r_1)$ , in Eq. (9). A symmetric relationship holds for  $i_{\text{KL}}(r; S)$ . As shown in DeWeese and Meister [18],  $i_{\text{KL}}(s, R)$  is non additive when conditioned on multiple stimulus values,

$$i_{\text{KL}}(\{s_1, s_2\}; R) \neq i_{\text{KL}}(s_1; R) + i_{\text{KL}}(s_2|s_1; R), \quad (56)$$

unlike  $i_h(s; R)$ , which is additive in the sense of both Eqs. (55) and (56) [10, 21]. However, Eq. (56) should *not discourage* us from choosing  $i_{KL}$  as the proper specific information measure. It is the additivity of the form of Eq. (55), which stands as a prerequisite in the classical axiomatic approach to information theory, see Kullback [35, p. 12] and references therein. On a more fundamental level, the mutual information in Eq. (3) is essentially the expected value (with respect to the joint distribution  $f_{S,R}$ ) of the likelihood ratio [36]

$$I(s; r) = \log \frac{f_{R|S}(r|s)}{p_R(r)} = \log \frac{f_{S|R}(s|r)}{\pi_S(s)}. \quad (57)$$

The quantity  $I(s; r)$  is symmetric, additive and invariant, and plays a key role in several modern developments within the field of information theory [70, 71].

## 5. CONCLUSIONS

Taken together, our results can be summarized into the following three main points.

First, by employing the classical model of the auditory nerve fiber [32], we demonstrate that just by evaluating the neuronal response in two *equivalent* parameterizations we obtain absolutely incongruent results about the information content of the

corresponding sound intensities.

Second, the requirement for reference frame invariance helps us to resolve the ambiguity in the choice of possible specific information measures, and identify the Kullback-Leibler divergence  $i_{KL}$  (specific surprise) as the unique solution to the problem. We also point out that a change in mutual information, due to a perturbation in the stimulus distribution, is expressed in terms of the Kullback-Leibler divergence and not in terms of the other specific information measures. Similarly, the necessary and sufficient conditions for achieving the maximal information transfer (capacity or capacity-cost), lead to  $i_{KL}$ . The sensitivity to perturbations and the capacity-achieving conditions are not related to the coordinate invariance problem *per se*, nonetheless, they confirm the prominent role of the Kullback-Leibler divergence from additional points of view.

Third, we argue that the transformation invariance is a fundamental and necessary criterion, which should be imposed on any framework or methodology used in computational neuroscience. If no invariant method exists for the given task then the proper coordinate system should be identified from other principles [29, 30]. We hope that our conclusions will stimulate further research focusing on special reference frames and suitable invariant methods for the analysis of neural code.

*Acknowledgments.* We thank Sidney Lehky for helpful comments on the manuscript. This work was supported by the Institute of Physiology RVO:67985823 and by the Czech Science Foundation project 17-06943S.

- 
- [1] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nature Neurosci.* **2**, 947–958 (1999).
  - [2] A. G. Dimitrov, A. L. Lazar, and J. D. Victor, "Information theory in neuroscience," *J. Comput. Neurosci.* **30**, 1–5 (2011).
  - [3] M. D. McDonnell, S. Ikeda, and J. H. Manton, "An introductory review of information theory in the context of computational neuroscience," *Biol. Cybern.* **105**, 55–70 (2011).
  - [4] R. B. Stein, "The information capacity of nerve cells using a frequency code," *Biophys. J.* **7**, 797–826 (1967).
  - [5] A. G. Dimitrov and J. P. Miller, "Neural coding and decoding: communication channels and quantization," *Netw. Comput. Neural Syst.* **12**, 441–472 (2001).
  - [6] M. D. McDonnell and N. G. Stocks, "Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations," *Phys. Rev. Lett.* **101**, 058103 (2008).
  - [7] S. Ikeda and J. H. Manton, "Capacity of a single spiking neuron channel," *Neural Comput.* **21**, 1714–1748 (2009).
  - [8] L. Kostal, P. Lansky, and M. D. McDonnell, "Metabolic cost of neuronal information in an empirical stimulus-response model," *Biol. Cybern.* **107**, 355–365 (2013).
  - [9] L. Kostal and R. Kobayashi, "Optimal decoding and information transmission in Hodgkin-Huxley neurons under metabolic cost constraints," *BioSystems* **136**, 3–10 (2015).
  - [10] M. Wibral, J. T. Lizier, and V. Priesemann, "Bits from brains for biologically inspired computing," *Front. Robot. AI* **2**, 5 (2015).
  - [11] D. H. Perkel and T. H. Bullock, "Neural coding," *Neurosci. Res. Prog. Sum.* **3**, 405–527 (1968).
  - [12] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, 2001).
  - [13] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, edited by W. Rosenblith (MIT Press, Cambridge, 1961) pp. 217–234.
  - [14] S. B. Laughlin, "A simple coding procedure enhances a neuron's information capacity," *Z. Naturforsch.* **36**, 910–912 (1981).
  - [15] J. J. Atick, "Could information theory provide an ecological theory of sensory processing?" *Netw. Comput. Neural Syst.* **3**, 213–251 (1992).
  - [16] L. Kostal, P. Lansky, and J-P. Rospars, "Efficient olfactory coding in the pheromone receptor neuron of a moth," *PLoS Comput. Biol.* **4**, e1000053 (2008).
  - [17] T. Berger and W. B. Levy, "A Mathematical Theory of Energy Efficient Neural Computation and Communication," *IEEE Trans. Inf. Theory* **56**, 852–874 (2010).
  - [18] M. R. DeWeese and M. Meister, "How to measure the information gained from one symbol," *Netw. Comput. Neural Syst.* **10**, 325–340 (1999).
  - [19] D. A. Butts, "How much information is associated with a particular stimulus?" *Netw. Comput. Neural Syst.* **14**, 177–187 (2003).
  - [20] T. Lu and X. Wang, "Information content of auditory cortical responses to time-varying acoustic stimuli," *J. Neurophysiol.* **91**, 301–313 (2004).
  - [21] M. Bezzi, "Quantifying the information transmitted in a single stimulus," *BioSystems* **89**, 4–9 (2007).
  - [22] S. Yarrow, E. Challis, and P. Seriès, "Fisher and Shannon information in finite neural populations," *Neural Comput.* **24**, 1740–1780 (2012).
  - [23] E. E. Thomson and W. B. Kristan, "Quantifying stimulus discriminability: A comparison of information theory and ideal observer analysis," *Neural Comput.* **17**, 741–778 (2005).



- [24] A. V. Olypher, P. Lansky, R. U. Muller, and A. A. Fenton, "Quantifying location-specific information in the discharge of rat hippocampal place cells," *J. Neurosci. Meth.* **127**, 123–135 (2003).
- [25] D. A. Butts and M. S. Goldman, "Tuning curves, neuronal variability, and sensory coding," *PLoS Biol.* **4**, e92 (2006).
- [26] P. Lansky, O. Pokora, and J-P. Rospars, "Classification of stimuli based on stimulus-response curves and their variability," *Brain Res.* **1225**, 57–66 (2008).
- [27] D. B. Kastner and S. A. Baccus, "Coordinated dynamic encoding in the retina using opposing forms of plasticity," *Nat. Neurosci.* **14**, 1317–1322 (2011).
- [28] L. Kostal and P. Lansky, "Coding accuracy is not fully determined by the neuronal model," *Neural Comput.* **27**, 1051–1057 (2015).
- [29] L. Kostal, "Stimulus reference frame and neural coding precision," *J. Math. Psychol.* **71**, 22–27 (2016).
- [30] L. Kostal and P. Lansky, "Coding accuracy on the psychophysical scale," *Sci. Rep.* **6**, 23810 (2016).
- [31] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).
- [32] R. L. Winslow and M. B. Sachs, "Single-tone intensity discrimination based on auditory-nerve rate responses in background of quiet, noise, and with stimulation of the crossed olivocochlear bundle," *Hearing Res.* **35**, 165–190 (1988).
- [33] B. Wen, G. I. Wang, I. Dean, and B. Delgutte, "Dynamic range adaptation to sound level statistics in the auditory nerve," *J. Neurosci.* **29**, 13797–13808 (2009).
- [34] J. Z. Sun, G. I. Wang, V. K. Goyal, and L. R. Varshney, "A framework for Bayesian optimality of psychophysical laws," *J. Math. Psychol.* **56**, 495–501 (2012).
- [35] S. Kullback, *Information theory and statistics* (Dover, New York, 1968).
- [36] R. G. Gallager, *Information Theory and Reliable Communication* (John Wiley and Sons, Inc., New York, USA, 1968).
- [37] H. C. Tuckwell, *Introduction to Theoretical Neurobiology, Vol. 2* (Cambridge University Press, New York, 1988).
- [38] M. N. Shadlen and W. T. Newsome, "The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding," *J. Neurosci.* **18**, 3870–3896 (1998).
- [39] R. B. Stein, E. R. Gossen, and K. E. Jones, "Neuronal variability: noise or part of the signal?" *Nat. Rev. Neurosci.* **6**, 389–397 (2005).
- [40] P. Lansky, L. Sacerdote, and C. Zucca, "The Gamma renewal process as an output of the diffusion leaky integrate-and-fire neuronal model," *Biol. Cybern.* **110**, 193–200 (2016).
- [41] F. Theunissen and J. P. Miller, "Temporal encoding in nervous systems: A rigorous definition," *J. Comput. Neurosci.* **2**, 149–162 (1995).
- [42] P. Lansky, R. Rodriguez, and L. Sacerdote, "Mean instantaneous firing frequency is always higher than the firing rate," *Neural Comput.* **16**, 477–489 (2004).
- [43] R. Kobayashi, Y. Tsubo, P. Lansky, and S. Shinomoto, "Estimating time-varying input signals and ion channel states from a single voltage trace of a neuron," in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 24, edited by J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira, and K. Q. Weinberger (MIT Press, Cambridge, Massachusetts, 2011) pp. 217–225.
- [44] M. Levakova, M. Tamborrino, L. Kostal, and P. Lansky, "Presynaptic spontaneous activity enhances the accuracy of latency coding," *Neural Comput.* **28**, 2162–2180 (2016).
- [45] D. H. Johnson, "Information theory and neural information processing," *IEEE Trans. Inf. Theory* **56**, 653–666 (2010).
- [46] P. Suksompong and T. Berger, "Capacity analysis for integrate-and-fire neurons with descending action potential thresholds," *IEEE Trans. Inf. Theory* **56**, 838–851 (2010).
- [47] J. J. Harris, R. Jolivet, and D. Attwell, "Synaptic energy use and supply," *Neuron* **75**, 762–777 (2012).
- [48] M. Wehr, "Stimulus-specific information," in *Encyclopedia of Computational Neuroscience*, edited by D. Jaeger and R. Jung (Springer, New York, 2013) pp. 1–4.
- [49] E. T. Rolls, A. Treves, M. J. Tovee, and S. Panzeri, "Information in the neuronal representation of individual stimuli in the primate temporal visual cortex," *J. Comput. Neurosci.* **4**, 309–333 (1997).
- [50] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, "Positional information, in bits," *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16301–16308 (2013).
- [51] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications* (MIT Press, New York, 1961).
- [52] A. Papoulis, *Probability, random variables, and stochastic processes* (McGraw-Hill, New York, 1991).
- [53] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E* **69**, 66138–16 (2004).
- [54] A. D. Wyner, "A definition of conditional mutual information for arbitrary ensembles," *Inform. Control* **38**, 51–59 (1978).
- [55] T. Schreiber, "Measuring Information Transfer," *Phys. Rev. Lett.* **85**, 461–464 (2000).
- [56] J. L. Massey, "Causality, feedback and directed information," in *Proceedings of the 1990 International Symposium on Information Theory and its Applications, Waikiki, Hawaii* (1990) pp. 27–30.
- [57] M. B. Sachs and P. J. Abbas, "Phenomenological model for two-tone suppression," *J. Acoust. Soc. Am.* **60**, 1157–1163 (1976).
- [58] M. C. Teich and S. M. Khanna, "Pulse-number distribution for the neural spike train in the cat's auditory nerve," *J. Acoustic. Soc. Am.* **77**, 1110–1128 (1985).
- [59] D. R. Cox and P. A. W. Lewis, *The statistical analysis of series of events* (Latimer Trend and Co. Ltd., Whistable, 1966).
- [60] K. E. Kaissling and J-P. Rospars, "Dose-response relationships in an olfactory flux detector model revisited," *Chem. Senses.* **29**, 529–531 (2004).
- [61] D. R. Raichel, *The Science and Applications of Acoustics* (Springer, New York, 2006).
- [62] S. A. Frank, "Input-output relations in biological systems: measurement, information and the Hill equation," *Biol. Direct* **8**, 13 (2013).
- [63] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, New York, 1958).
- [64] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar gaussian channels," *Inform. Control* **18**, 203–219 (1971).
- [65] M. S. Berger, *Nonlinearity and Functional Analysis* (Academic Press, New York, 1977).
- [66] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain," *J. Cereb. Blood Flow Metab.* **21**, 1133–1145 (2001).
- [67] J. J. Harris, R. Jolivet, E. Engl, and D. Attwell, "Energy-efficient information transfer by visual pathway synapses," *Curr. Biol.* **25**, 3151–3160 (2015).
- [68] D. G. Luenberger, *Optimization by Vector Space Methods* (Wiley, New York, 1969).
- [69] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems* (Academic Press, New York, 1981).
- [70] Y. M. Han, Y. S. Chan, K. S. Lo, and T. M. Wong, "Spontaneous activity and barosensitivity of the barosensitive neurons in the rostral ventrolateral medulla of hypertensive rats induced by transection of aortic depressor nerves," *Brain Res.* **813**, 262–7 (1998).
- [71] Y. Polyanskiy, V. H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory* **56**, 2307–2359 (2010).