

Stimulus reference frame and neural coding precision

Lubomir Kostal*

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

Any particular stimulus intensity, as a physical quantity, can be equivalently described in different unit systems. Researchers automatically expect the methodology and the inference obtained about the neural coding precision to be independent from such a subjective choice. We show, however, that the Fisher information, which is arguably the most popular measure of coding accuracy, may yield incompatible and in fact arbitrary results just by re-evaluating the identical stimulation scenario in transformed units. We consider only regular scale transformations given by strictly increasing and differentiable functions. On one hand, our results point to a potentially problematic aspect of the Fisher information application. On the other hand, we speculate that the unwanted transformation covariance may be removed by considering the psychophysical scale based on the ideal observer paradigm. We show that such scale implies constant Fisher information and that the matching stimulus distribution is given by the Jeffreys prior. The psychophysical perspective thus provides a novel justification for the special role of the Jeffreys prior in neural coding theory.

Keywords: Fisher information, Coding accuracy, Measurement scale, Jeffreys prior

1. INTRODUCTION

The methods of signal estimation and detection theory play an important role in the study of the neuronal coding problem. In particular, the Fisher information and the Cramér-Rao bound are frequently employed to address the theoretical limits on the coding and discrimination precision [1–12]. Comparing this precision for different types of responses (e.g., counts of action potentials or intervals in between them) then yields a potentially important piece of evidence about the principles of neural coding. In this regard, the influential *efficient coding hypothesis* [13] states that neuronal responses are adjusted, through evolutionary and adaptive processes, to optimally encode such stimulus statistics that matches the local sensory environment [14–20]. Consequentially, it is often expected that the high-probability stimulus regions match the peaks of the coding accuracy [21–26].

In this paper we investigate the general consequences of the measurement scale choice on the coding accuracy, as described by the Fisher information. We focus especially on the potential ambiguity of the conclusions that may be drawn from the shape of the Fisher information curve, viewed either as a function of the stimulus intensity (Section 3.1) or as a function of the probability of the stimulus intensity (Section 3.2). Finally, we speculate that the proper scale, on which the coding accuracy should be interpreted, results from the psychophysical function relating the decoding error to the just noticeable difference in perception (Section 3.3). We show, by following the efficient coding principle on the psychophysical scale, that the optimal stimulus distribution is given by the Jeffreys prior. The Jeffreys prior arises also as the capacity-achieving input distribution in the limit of vanishing noise during the information transmission [27–32], hence our results provide yet another interpretation of its optimality.

2. METHODS

The neural coding problem is often approached by virtue of the stimulus-response model, in which the neuronal response r (often given by the firing frequency) is related to the stimulus parameter θ . It is well known that the response varies randomly across trials [21, 33, 34]. The stimulus-response model, denoted also as the encoding model, is therefore fully described in terms of the response probability density (or mass) function $f(r; \theta)$.

The stimulus θ can be equivalently evaluated in different physical or measurement units $[\lambda]$, where $\lambda = \varphi(\theta)$, provided that the function φ is bijective. In this paper we restrict ourselves to *regular* transformations of the measurement scale, defined as follows.

Definition 1. Let the stimulus parameter θ take values in a closed interval of the real line $[\theta_{\min}, \theta_{\max}]$. The new stimulus parameter $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ is given by the regular transformation of θ ,

$$\lambda = \varphi(\theta), \quad (1)$$

if φ is strictly increasing and continuous on $[\theta_{\min}, \theta_{\max}]$ and differentiable on $(\theta_{\min}, \theta_{\max})$.

The regular transformations are appealing since they preserve the relative ordering of the stimulus intensities on various scales, i.e., $\theta_1 > \theta_2$ implies $\varphi(\theta_1) > \varphi(\theta_2)$. An example of such φ is given by the sound loudness parameter: one may use the units of sound intensity, pressure or level, which are mutually related by quadratic and logarithmic transformations [35]. The probability distributions $f(r; \theta)$ and $f(r; \lambda)$ thus describe *the same* encoding model, and there is no persuasive preference for the θ -parameterization over the λ -parameterization, perhaps besides convenience.

The problem of the stimulus coding precision, i.e., the exact evaluation of the smallest achievable error, is generally non-trivial [11, 36]. It is often more practical to evaluate the Cramér-Rao bound on the decoding mean square error (MSE)

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

instead. If there is no bias in the decoding and the model satisfies certain regularity conditions [37, 38], the Cramér-Rao bound states that the MSE for given stimulus θ satisfies

$$\text{MSE}(\theta) \geq \frac{1}{J_\theta(\theta)}. \quad (2)$$

The function $J_\theta(\theta)$ is the *Fisher information*,

$$J_\theta(\theta) = \int \left(\frac{\partial \log f(r; \theta)}{\partial \theta} \right)^2 f(r; \theta) dr, \quad (3)$$

which is strictly positive and continuous on $[\theta_{\min}, \theta_{\max}]$, see, e.g., Brown and Gajek [39] and Ibragimov and Has'minskii [38, Chapter 7] for details.

The Fisher information $J_\lambda(\lambda)$ in the regularly transformed units is related to the original function $J_\theta(\theta)$ as

$$J_\lambda(\lambda) = \left(\frac{d\varphi^{-1}(\lambda)}{d\lambda} \right)^2 J_\theta(\varphi^{-1}(\lambda)), \quad (4)$$

which can be proven by employing the chain rule for derivatives [40], with φ^{-1} denoting the inverse to Eq. (1). In this paper we examine the potential effect of regular transformations on the inference about neural coding precision (given by the Fisher information), hence Eq. (4) is of key importance.

Furthermore, the stimulus parameter itself may be a random variable, distributed according to the probability density function $p_\theta(\theta)$. The corresponding probability density function $p_\lambda(\lambda)$ in the regularly transformed units is then given by [41]

$$p_\lambda(\lambda) = p_\theta(\varphi^{-1}(\lambda)) \left| \frac{d\varphi^{-1}(\lambda)}{d\lambda} \right|. \quad (5)$$

From Eq. (4) follows that the square root of the Fisher information transforms similarly to Eq. (5). Therefore, if the proportionality

$$p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}, \quad (6)$$

holds in the stimulus units of θ , it holds *invariantly* as $p_\lambda(\lambda) \propto \sqrt{J_\lambda(\lambda)}$ in any regularly transformed units. The particular stimulus distribution given by Eq. (6) is known as the *Jeffreys prior*, an important concept in the Bayesian inference due to the invariance property [42, 43].

3. RESULTS AND DISCUSSION

3.1. Fisher information under regular transformations of the stimulus measurement units

The profile of the Fisher information as a function of the stimulus parameter has been investigated extensively in the computational neuroscience literature. For example, Greenwood *et al.* [44] employed $J_\theta(\theta)$ to describe the stochastic resonance effect in a generalized McCulloch-Pitts neuronal model. Wilke and Eurich [45] analyzed the accuracy

with which a neural population encodes a number of stimulus features, and found that the stimulus value yielding maximal Fisher information is below the mode of the mean response curve. More generally, the mode of $J_\theta(\theta)$ was employed to identify the optimal stimulation intensity in a number of stochastic [8, 46] and empirical [47] neuronal models. Optimal coding accuracy in terms of latency (time to first spike after the stimulus onset) was investigated in a similar way [48, 49]. Jenison and Reale [50] used the Fisher information to quantify the estimation precision of multiple sound localization parameters in the auditory system. And finally, profiles of $J_\theta(\theta)$ and certain information-theoretic measures were compared in Yarrow *et al.* [30] and Kostal and Lansky [51]. See also Pilarski and Pokora [12] for additional references.

In a recent study Kostal and Lansky [52] demonstrated that, paradoxically, the Fisher information cannot be used to decide whether loud or quiet tones are encoded more precisely in a classical model of the auditory nerve fiber [53]. The Fisher information turned out to be an increasing function of the sound intensity evaluated in decibels (dB SPL), and a decreasing function of the same intensity in Pascals.

We now prove that the Fisher information of a given encoding model may in fact follow the shape of any desired function, just by changing the stimulus measurement scale.

Theorem 1. *Let $g(x) > 0$, $x \in [0, 1]$ be a Lipschitz continuous function. For any given $J_\theta(\theta)$ there exists a unique regular transformation φ in Eq. (1) and a constant $a > 0$ such that*

$$J_\lambda(\lambda) = ag(\lambda), \quad (7)$$

and $\lambda_{\min} = \varphi(\theta_{\min}) = 0$ and $\lambda_{\max} = \varphi(\theta_{\max}) = 1$. The transformation φ is given by the initial value problem

$$\frac{d\varphi(\theta)}{d\theta} \sqrt{ag(\varphi(\theta))} = \sqrt{J_\theta(\theta)}, \quad (8)$$

with $\varphi(\theta_{\min}) = 0$.

Proof. Eq. (4) can be also stated as

$$J_\theta(\theta) = \left(\frac{d\varphi(\theta)}{d\theta} \right)^2 J_\lambda(\varphi(\theta)). \quad (9)$$

Although the numerical values of Fisher information in different units change accordingly to Eq. (4), the integral of $\sqrt{J_\theta(\theta)}$ is *conserved*. To show this we employ the substitution rule and Eq. (9),

$$\begin{aligned} \int_{\varphi(\theta_{\min})}^{\varphi(\theta_{\max})} \sqrt{J_\lambda(\lambda)} d\lambda &= \int_{\theta_{\min}}^{\theta_{\max}} \sqrt{J_\lambda(\varphi(\theta))} \left| \frac{d\varphi(\theta)}{d\theta} \right| d\theta = \\ &= \int_{\theta_{\min}}^{\theta_{\max}} \sqrt{J_\theta(\theta)} d\theta \equiv C, \end{aligned} \quad (10)$$

Next, the continuity of $g(x)$ guarantees that the integral

$$\int_0^1 \sqrt{g(x)} dx = G \quad (11)$$

exists. Therefore, Eq. (7) is a valid requirement only if

$$a = \frac{C^2}{G^2}, \quad (12)$$

since from Eq. (10) follows that it must hold $\int_0^1 \sqrt{ag(x)} dx = C$. Finally, we substitute Eq. (7) into Eq. (9) and re-arrange the terms to obtain Eq. (8) \square

As a special case we find the regular transformation such that $J_\lambda(\lambda)$ is constant. Let therefore $g(x) = 1$ and $\lambda_{\max} = C$ so that $J_\lambda(\lambda) = 1$. From $\lambda_{\min} = 0$ and Eq. (8) then follows

$$\varphi(\theta) = \int_{\theta_{\min}}^{\theta} \sqrt{J_\theta(z)} dz. \quad (13)$$

3.2. Matching the Fisher information to the stimulus distribution

According to the efficient coding principle, the neuronal coding strategy is adapted to the local stimulus *statistics* (see, e.g., Wark *et al.* [24] and references therein). In other words, the coding accuracy and the stimulus probability distribution are expected to be related. In particular, peaks of the Fisher information are often found near the high-probability regions of the stimulus distribution, as reported in the auditory coding of the sound intensity [23, 25, 26, 54], coding of the interaural level differences [55] and time differences [56]. The situation has also been investigated in other sensory modalities, e.g., in the primary visual cortex [57] and primary somatosensory cortex [58].

In order to proceed, the notion of ‘‘large Fisher information values matching the high-probability stimulus regions’’ must be made more precise. We formalize the requirement by letting the Fisher information vary as a suitable function of the stimulus probability. The following theorem shows, however, that almost any desired match between the Fisher information and the stimulus distribution can be achieved just by choosing the stimulus unit properly.

Theorem 2. *Let the stimulus parameter θ be distributed according to a continuous probability density function $p_\theta(\theta) > 0$ on $[\theta_{\min}, \theta_{\max}]$. Let $h(x) \neq ax^2, a > 0$ be a strictly increasing Lipschitz continuous function on $x > 0$. For any $J_\theta(\theta) \neq bp_\theta^2(\theta), b > 0$ there exists a transformation φ such that*

$$J_\lambda(\lambda) = h(p_\lambda(\lambda)), \quad (14)$$

where $p_\lambda(\lambda)$ is the probability density function of the transformed stimulus parameter, and φ is given by the implicit differential equation

$$\left(\frac{d\varphi(\theta)}{d\theta} \right)^2 h \left(p_\theta(\theta) \left| \frac{d\varphi(\theta)}{d\theta} \right|^{-1} \right) = J_\theta(\theta), \quad (15)$$

with $\lambda_{\min} = \varphi(\theta_{\min}) = 0$.

Proof. Substituting Eq. (14) into Eq. (4) gives

$$\frac{d\varphi^{-1}(\lambda)}{d\lambda} = \sqrt{\frac{h(p_\lambda(\lambda))}{J_\theta(\varphi^{-1}(\lambda))}}. \quad (16)$$

By employing the inverse function theorem,

$$\frac{d\varphi^{-1}(\lambda)}{d\lambda} \Big|_{\lambda=\varphi(\theta)} = \left[\frac{d\varphi(\theta)}{d\theta} \right]^{-1}, \quad (17)$$

and by substituting Eq. (5) into Eq. (16) we obtain Eq. (15). \square

The special case $h(x) \propto x^2$ reduces Eq. (15) to the algebraic form $p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$. This is equivalent to the statement of the Jeffreys prior invariance: it is impossible to obtain $p_\lambda(\lambda) \propto \sqrt{J_\lambda(\lambda)}$ by virtue of a regular transformation unless Eq. (6) holds in the original θ -units already.

The implicit form of Eq. (15) is not immediately useful, hence we particularize the Theorem 2 to $h(x) = x$. We obtain the transformation that produces the *exact* numerical match between any given Fisher information and stimulus distribution.

Corollary 1. *Under the conditions of Theorem 2, the regular transformation yielding*

$$J_\lambda(\lambda) = p_\lambda(\lambda), \quad (18)$$

is given by

$$\varphi(\theta) = \int_{\theta_{\min}}^{\theta} \frac{J_\theta(z)}{p_\theta(z)} dz. \quad (19)$$

We illustrate the stimulus rescaling in Eq. (19) on the model of a rat olfactory receptor neuron [59]. The mean response $\mu_\theta(\theta)$ (average evoked firing rate in action potentials per second) in dependence on the odorant concentration follows the Hill function

$$\mu_\theta(\theta) = \frac{F_M}{1 + 10^{(\log_{10} K - \theta)N}}. \quad (20)$$

The stimulus θ is the decadic logarithm of odorant concentration in mol/L, F_M is the maximum asymptotic firing rate, K is the odorant concentration that evokes response rate equal to $F_M/2$ and N is the Hill coefficient. Let us assume that the response distribution (for each θ) is a Gaussian with mean $\mu_\theta(\theta)$ and a Poisson-like relationship between the mean and variance σ_θ^2 , i.e., $\sigma_\theta(\theta) = \sqrt{\mu_\theta(\theta)}$, hence the encoding model is

$$f(r; \theta) = \frac{1}{\sqrt{2\pi\mu_\theta(\theta)}} \exp \left(- \frac{[r - \mu_\theta(\theta)]^2}{2\mu_\theta(\theta)} \right). \quad (21)$$

The values of the model parameters depend on the odorant type, here we choose the typical values $F_M = 49$ spikes/s,

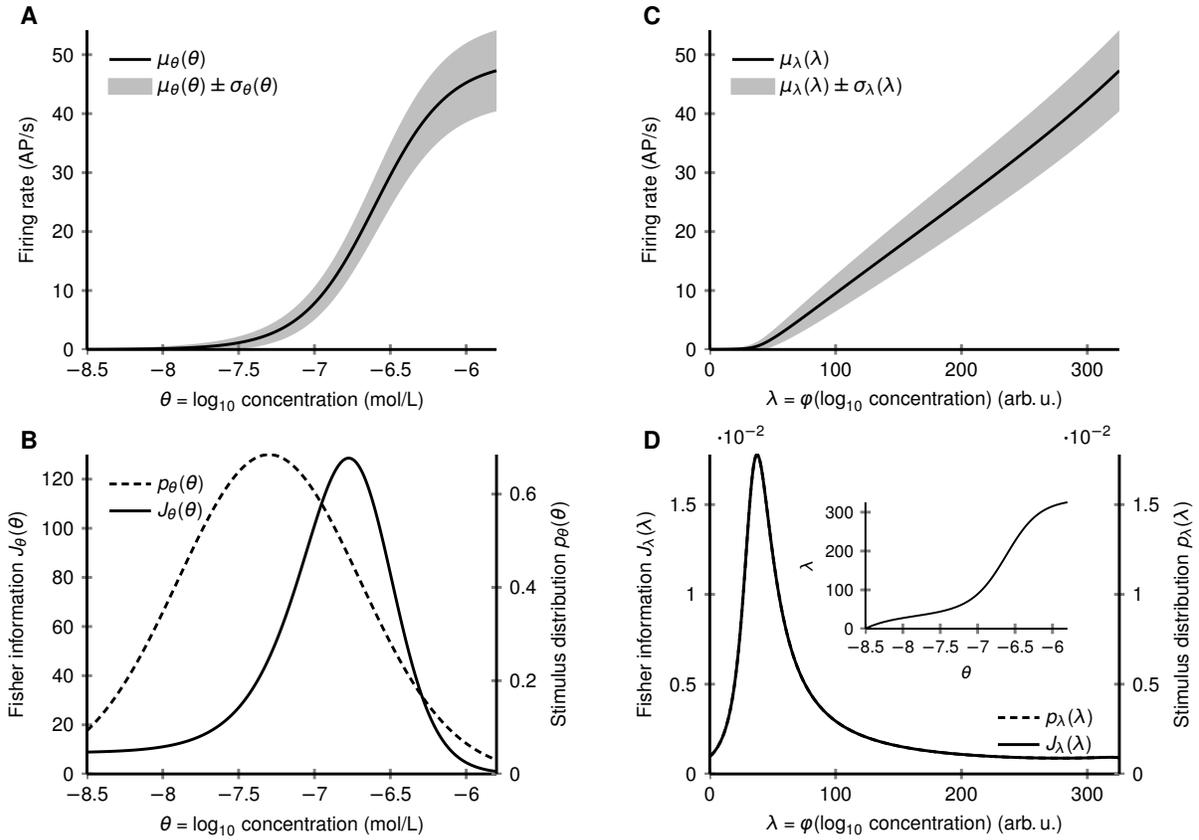


Figure 1. Regular unit transformation guarantees the match between the given Fisher information and an arbitrary stimulus input distribution. (A) Stimulus-response model based on rat olfactory receptor neurons electrophysiological recordings [59]. The mean response $\mu_\theta(\theta)$ (solid) is described by the Hill function and the Gaussian response model with Poisson-like variability is assumed (the standard deviation is indicated by the gray region). (B) The shape of the Fisher information $J_\theta(\theta)$ does not match the shape of the given stimulus distribution $p_\theta(\theta)$, e.g., the region of high $J_\theta(\theta)$ does not coincide with the most frequent stimulus. (C) The same neuronal model and stimulation scenario as in (A); the stimulus (λ) is evaluated in the transformed units given by Eq. (19). (D) Although the only change with respect to (B) is that the stimulus is evaluated in different units, the match between the Fisher information $J_\lambda(\lambda)$ and the input distribution $p_\lambda(\lambda)$ is exact. The conversion function φ between the stimulus evaluated in θ -units and in λ -units is strictly increasing and differentiable.

$N = 1.8$ and $K = 2.5 \cdot 10^{-7}$ mol/L (Fig. 1A), see Rospars *et al.* [59] for details. The Fisher information for the model given by Eq. (21) can be expressed as

$$J_\theta(\theta) = \frac{1 + 2\mu_\theta(\theta)}{2\mu_\theta^2(\theta)} \left(\frac{d\mu_\theta(\theta)}{d\theta} \right)^2. \quad (22)$$

Let the stimulus distribution $p_\theta(\theta)$ be a Gaussian $\Theta \sim N(m, s^2)$, renormalized on the interval $[\theta_{\min}, \theta_{\max}]$, and deliberately such that the overall shapes of $J_\theta(\theta)$ and $p_\theta(\theta)$ do not match, $m = -7.3$ and $\sigma = 0.6$ (Fig. 1B). The transformation relating the θ - and λ -description of the odorant concentration, $\lambda = \varphi(\theta)$, is given by Eq. (19) and cannot be evaluated in a closed form. The stimulus-response model $f(r; \theta)$ is then equivalently described by $f(r; \lambda)$. The mean response, $\mu_\lambda(\lambda) = \mu_\theta(\varphi^{-1}(\lambda))$, and the accompanying standard deviation, $\sigma_\lambda(\lambda) = \sigma_\theta(\varphi^{-1}(\lambda))$, are shown in Fig. 1C. Note that the Poisson-like mean-to-variance ratio is preserved since we re-scale only the stimulus variable and not the response variable. Coincidentally, the mean response in the λ -units is

approximately linear over the entire coding range. The Fisher information $J_\lambda(\lambda)$ and the input distribution $p_\lambda(\lambda)$, given by Eq. (5), match exactly in the transformed units (Fig. 1D). The transformation φ relating the old (θ) and new (λ) stimulus units is nonlinear, strictly increasing and differentiable (Fig. 1D).

3.3. Fisher information and the psychophysical function

The psychophysical function describes the perceptual intensity, S , in dependence on the stimulus parameter [60, 61]. Arguably, the most famous psychophysical function example is the Weber-Fechner law, $S \propto \log \theta$ [21, 62–64]. In his pioneering study, Riesz [65] provided a correction to the Weber-Fechner law for the perception of sound intensity. Kostal and Lansky [66] showed that this seemingly small modification significantly improves the match between the stimulus distribution and the Fisher information in the study of Watkins

and Barbour [25]. Motivated by this example we speculate that, generally, the stimulus measurement scale based on the psychophysical function yields the proper inference about the efficient-coding adaptation.

We propose the *psychophysical scale* of the stimulus parameter λ such that it is *proportional* to the sensation level,

$$S = c\lambda, \quad (23)$$

where $c > 0$ is a “dimension-correcting” constant. Let, numerically, be $c = 1$ for simplicity. The reasoning is that the *just noticeable difference* (JND) in the perception, ΔS [22, 60, 67, 68] is a function of some increment $\Delta\lambda$ but *not* of λ . Equivalently, the JND-inducing $\Delta\lambda$ must be *independent* of λ . Therefore the unique feature of the stimulus unit $[\lambda]$ is that the coding accuracy can be meaningfully compared for different stimuli intensities. Contrast that with an arbitrary unit $[\theta]$: an apparently large variation in $J_\theta(\theta)$ might be immaterial for the neural system, provided that the actual difference in sensation always falls below the JND.

The true psychophysical function is rarely known exactly [60, 64, 65, 69]. However, the JND-inducing $\Delta\lambda$ is often derived from the Cramér-Rao bound approximation to the ideal observer error as [21, 67, 68]

$$\Delta\lambda = D_\alpha \frac{1}{\sqrt{J_\lambda(\lambda)}}, \quad (24)$$

where $D_\alpha > 0$ is the constant discriminability factor based on the separation of two Gaussian distributions. Since $\Delta\lambda$ is independent of λ , it must be that $J_\lambda(\lambda)$ is *constant*. Assuming that the Fisher information “matches” the stimulus distribution according to Eq. (14) we have that $p_\lambda(\lambda)$ must be also constant. However, constant J_λ and uniform p_λ together imply that the Jeffreys prior, $p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$, holds in the original units. We can see that from the reverse form of Eq. (5),

$$p_\theta(\theta) = p_\lambda(\varphi(\theta)) \left| \frac{d\varphi(\theta)}{d\theta} \right|, \quad (25)$$

which implies $|d\varphi(\theta)/d\theta| \propto p_\theta(\theta)$, while Eq. (9) implies $|d\varphi(\theta)/d\theta| \propto \sqrt{J_\theta(\theta)}$. Moreover, the regular transformation from the original (and arbitrary) θ -units to the λ -units is given by (*cf.* Eq. 13)

$$\varphi(\theta) = a \int_{\theta_{\min}}^{\theta} \sqrt{J_\theta(z)} dz, \quad (26)$$

where $a > 0$, and hence $S \propto \varphi(\theta)$ is the psychophysical function.

We illustrate Eq. (26) on the stimulus-response relationship from Eq. (21). Without loss of generality we put $a = 1$. Due to the simplicity of the Gaussian model we can evaluate Eq. (26) in a closed form as

$$\varphi(\theta) = \Phi(\theta) - \Phi(\theta_{\min}), \quad (27)$$

where

$$\Phi(\theta) = \frac{1}{\sqrt{2}} \left(2\sqrt{1 + 2\mu_\theta(\theta)} + \log \frac{1 - \sqrt{1 + 2\mu_\theta(\theta)}}{1 + \sqrt{1 + 2\mu_\theta(\theta)}} \right). \quad (28)$$

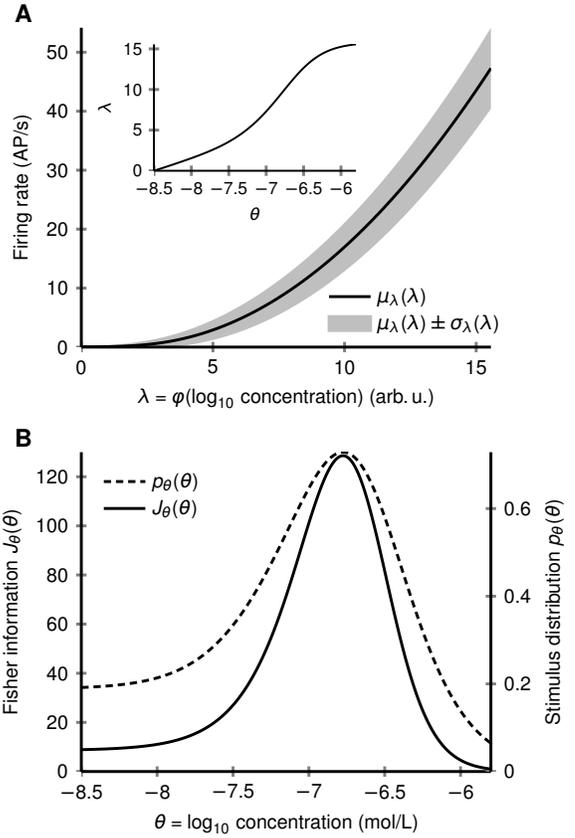


Figure 2. The psychophysical scale implies the Jeffreys prior as the matching stimulus distribution. (A) The stimulus-response model from Fig. 1A viewed on the psychophysical scale given by Eq. (26). The stimulus unit transformation relates the original units of θ (\log_{10} of the odorant concentration) to the arbitrary psychophysical units of λ . The Fisher information $J_\lambda(\lambda)$ is constant (not shown) and hence the matching stimulus distribution $p_\lambda(\lambda)$ is uniform. (B) The stimulus distribution transformed back to the units of θ is given by the Jeffreys prior, $p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$ (dashed). The Fisher information (solid) from Fig. 1B is shown for comparison.

The stimulus-response model on the psychophysical scale is shown in Fig. 2A (*cf.* Fig. 1A, C), together with the transformation $\lambda = \varphi(\theta)$. The mean response $\mu_\lambda(\lambda)$ is nonlinear but $J_\lambda(\lambda) = 1$, since the response variability depends on λ , and $p_\lambda(\lambda) = 1/\lambda_{\max}$. We have $p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$ (Fig. 2B, dashed) on the original stimulus scale.

4. CONCLUSIONS

Taken together, our results may be summarized in three main points.

First, we show that the change of measurement units may not only reverse the monotonicity of the Fisher information profile as a function of the stimulus parameter [52], but may potentially result in an arbitrary (e.g., multimodal) form of dependence. Put more broadly, since the basic principle in

optimal design is the maximization of Fisher information [70], this can lead to ambiguous design choices.

Second, we consider the matching between the Fisher information profile and the probability of the stimulus parameter, motivated by the efficient coding hypothesis. We demonstrate that such matching is only relative, changing with the stimulus reference frame. In this sense, the proper choice of the stimulus scale is essential for the investigation of the neural coding precision.

Third, we hypothesize that the stimulus scale which is proportional to the perception intensity provides the natural fixed point in the continuum of possible different reference frames, relating the decoding precision to the just noticeable differ-

ence. By approximating the ideal observer error with the Cramér-Rao bound we show that the psychophysical stimulus scale implies constant Fisher information, and furthermore, that the matching distribution on this scale is uniform. The relationship between the Fisher information and the input distribution in any other unit systems is then given by the Jeffreys prior.

Acknowledgments. I thank Petr Lansky and Peter Latham for helpful comments on the manuscript. This work was supported by the Institute of Physiology RVO:67985823 and by the Czech Science Foundation project GA15-08066S.

-
- [1] M. A. Paradiso, “A theory for the use of visual orientation information which exploits the columnar structure of striate cortex,” *Biol. Cybern.* **58**, 35–49 (1988).
- [2] M. Stemmler, “A single spike suffices: the simplest form of stochastic resonance in model neurons,” *Network* **7**, 687–716 (1996).
- [3] P. Dayan and L. F. Abbott, “The Effect of Correlated Variability on the Accuracy of a Population Code,” *Neural Comput.* **11**, 91–101 (1999).
- [4] M. Bethge, D. Rotermund, and K. Pawelzik, “Optimal short-term population coding: when Fisher information fails,” *Neural Comput.* **14**, 2317–2351 (2002).
- [5] P. Seriès, P. E. Latham, and A. Pouget, “Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations,” *Nat. Neurosci.* **7**, 1129–1135 (2004).
- [6] K. Zhang and T. J. Sejnowski, “Neuronal tuning: to sharpen or broaden?” *Neural Comput.* **11**, 75–84 (1999).
- [7] P. E. Greenwood, L. M. Ward, D. F. Russell, A. Neiman, and F. Moss, “Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture,” *Phys. Rev. Lett.* **84**, 4773–4776 (2000).
- [8] P. E. Greenwood and P. Lansky, “Optimal signal estimation in neuronal models,” *Neural Comput.* **17**, 2240–2257 (2005).
- [9] K. Zhang, I. Ginzburg, B. L. McNaughton, and T. J. Sejnowski, “Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells,” *J. Neurophysiol.* **79**, 1017–1044 (1998).
- [10] N. S. Harper and D. McAlpine, “Optimal neural population coding of an auditory spatial cue,” *Nature* **430**, 682–686 (2004).
- [11] L. Kostal, P. Lansky, and S. Pilarski, “Performance breakdown in optimal stimulus decoding,” *J. Neural Eng.* **12**, 036012 (2015).
- [12] S. Pilarski and O. Pokora, “On the Cramér-Rao bound applicability and the role of Fisher information in computational neuroscience,” *BioSystems* **136**, 11–22 (2015).
- [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] J. J. Atick, “Could information theory provide an ecological theory of sensory processing?” *Netw. Comput. Neural Syst.* **3**, 213–251 (1992).
- [15] W. Bialek and W. G. Owen, “Temporal filtering in retinal bipolar cells. elements of an optimal computation?” *Biophys. J.* **58**, 1227–1233 (1990).
- [16] E. P. Hornstein, D. C. O’Carroll, J. C. Anderson, and S. B. Laughlin, “Sexual dimorphism matches photoreceptor performance to behavioural requirements,” *Proc. Biol. Sci.* **267**, 2111–2117 (2000).
- [17] E. P. Simoncelli and B. A. Olshausen, “Natural image statistics and neural representation,” *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
- [18] M. S. Lewicki, “Efficient coding of natural sounds,” *Nat. Neurosci.* **5**, 356–363 (2002).
- [19] E. C. Smith and M. S. Lewicki, “Efficient auditory coding,” *Nature* **439**, 978–982 (2006).
- [20] 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).
- [21] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2001).
- [22] S. B. Laughlin, “A simple coding procedure enhances a neuron’s information capacity,” *Z. Naturforsch.* **36**, 910–912 (1981).
- [23] I. Dean, N. S. Harper, and D. McAlpine, “Neural population coding of sound level adapts to stimulus statistics,” *Nat. Neurosci.* **8**, 1684–1689 (2005).
- [24] B. Wark, B. N. Lundstrom, and A. Fairhall, “Sensory adaptation,” *Curr. Opin. Neurobiol.* **17**, 423–429 (2007).
- [25] P. V. Watkins and D. L. Barbour, “Specialized neuronal adaptation for preserving input sensitivity,” *Nat. Neurosci.* **11**, 1259–1261 (2008).
- [26] 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).
- [27] J. M. Bernardo, “Reference posterior distributions for Bayesian inference,” *J. Roy. Stat. Soc. B* **41**, 113–147 (1979).
- [28] N. Brunel and J-P. Nadal, “Mutual information, Fisher information, and population coding,” *Neural Comput.* **10**, 1731–1757 (1998).
- [29] 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).
- [30] S. Yarrow, E. Challis, and P. Seriès, “Fisher and Shannon information in finite neural populations,” *Neural Comput.* **24**, 1740–1780 (2012).
- [31] 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).

- [32] X. Wei and A. A. Stocker, “A Bayesian observer model constrained by efficient coding can explain ‘anti-bayesian’ percepts,” *Nat. Neurosci.* **18**, 1509–1517 (2015).
- [33] 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).
- [34] 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).
- [35] D. R. Raichel, *The Science and Applications of Acoustics* (Springer, New York, 2006).
- [36] H. L. van Trees and K. L. Bell, *Detection, Estimation and Modulation Theory, Part I* (John Wiley and Sons, New York, 2013).
- [37] E. J. G. Pitman, *Some basic theory for statistical inference* (John Wiley and Sons, Inc., New York, 1979).
- [38] L. A. Ibragimov and R. Z. Has’minskii, *Statistical Estimation – Asymptotic Theory* (Springer-Verlag, New York, 1981).
- [39] L. D. Brown and L. Gajek, “Information inequalities for the Bayes risk,” *Ann. Stat.* **18**, 1578–1594 (1990).
- [40] E. L. Lehmann and G. Casella, *Theory of point estimation* (Springer Verlag, New York, 1998).
- [41] A. Papoulis, *Probability, random variables, and stochastic processes* (McGraw-Hill, New York, 1991).
- [42] H. Jeffreys, “An invariant form for the prior probability in estimation problems,” *Proc. Roy. Soc. A*, 453–461 (1946).
- [43] R. E. Kass and L. Wasserman, “The selection of prior distributions by formal rules,” *J. Am. Stat. Assoc.* **91**, 1343–1370 (1996).
- [44] P. E. Greenwood, L. M. Ward, and W. Wefelmeyer, “Statistical analysis of stochastic resonance in a simple setting,” *Phys. Rev. E* **60**, 4687–4695 (1999).
- [45] S. D. Wilke and C. W. Eurich, “Representational accuracy of stochastic neural populations,” *Neural Comput.* **14**, 155–189 (2002).
- [46] P. E. Greenwood and P. Lansky, “Optimum signal in a simple neuronal model with signal-dependent noise,” *Biol. Cybern.* **92**, 199–205 (2005).
- [47] 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).
- [48] G. Wainrib, M. Thioullien, and K. Pakdaman, “Intrinsic variability of latency to first-spike,” *Biol. Cybern.* **103**, 43–56 (2010).
- [49] M. Levakova, “Effect of spontaneous activity on stimulus detection in a simple neuronal model,” *Math. Biosci. Eng.* **13**, 551–568 (2016).
- [50] R. L. Jenison and R. A. Reale, “Likelihood approaches to sensory coding in auditory cortex,” *Netw. Comput. Neural Syst.* **14**, 83–102 (2003).
- [51] L. Kostal and P. Lansky, “Information capacity and its approximations under metabolic cost in a simple homogeneous population of neurons,” *BioSystems* **112**, 265–275 (2013).
- [52] L. Kostal and P. Lansky, “Coding accuracy is not fully determined by the neuronal model,” *Neural Comput.* **27**, 1051–1057 (2015).
- [53] 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).
- [54] P. V. Watkins and D. L. Barbour, “Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds,” *Cereb. Cortex* **21**, 178–190 (2011).
- [55] J. C. Dahmen, P. Keating, F. R. Nodal, A. L. Schulz, and A. J. King, “Adaptation to stimulus statistics in the perception and neural representation of auditory space,” *Neuron* **66**, 937–948 (2010).
- [56] J. K. Maier, P. Hehrmann, N. S. Harper, G. M. Klump, D. Pressnitzer, and D. McAlpine, “Adaptive coding is constrained to midline locations in a spatial listening task,” *J. Neurophysiol.* **108**, 1856–1868 (2012).
- [57] S. Durant, C. W. G. Clifford, N. A. Crowder, N. S. C. Price, and M. R. Ibbotson, “Characterizing contrast adaptation in a population of cat primary visual cortical neurons using Fisher information,” *J. Opt. Soc. Am. A* **24**, 1529–1537 (2007).
- [58] J. A. Garcia-Lazaro, S. S. M. Ho, A. Nair, and J. W. H. Schnupp, “Shifting and scaling adaptation to dynamic stimuli in somatosensory cortex,” *Eur. J. Neurosci.* **26**, 2359–2368 (2007).
- [59] J-P. Rospars, P. Lansky, M. Chaput, and P. Viret, “Competitive and noncompetitive odorant interaction in the early neural coding of odorant mixtures,” *J. Neurosci.* **28**, 2659–2666 (2008).
- [60] G. A. Gescheider, *Psychophysics: The fundamentals* (Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997).
- [61] 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).
- [62] E. H. Weber, *De pulsus, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae* (Koehler, Leipzig, 1834).
- [63] E. N. Dzhafarov and H. Colonius, “The Fechnerian Idea,” *Am. J. Psychol.* **124**, 127–140 (2011).
- [64] S. C. Masin, “The (Weber’s) law that never was,” in *Proceedings of the twenty-fifth annual meeting of the International Society for Psychophysics*, edited by M. A. Elliot, S. Antonijević, S. Berthaud, P. Mulcahy, B. Bargary, C. Martyn, and H. Schmid (International Society for Psychophysics, Galway, Ireland, 2009) pp. 441–446.
- [65] R. R. Riesz, “Differential intensity sensitivity of the ear for pure tones,” *Phys. Rev.* **31**, 867–875 (1928).
- [66] L. Kostal and P. Lansky, “Coding accuracy on the psychophysical scale,” *Sci. Rep.* **6**, 23810 (2016).
- [67] T. Tzvetanov and T. Womelsdorf, “Predicting human perceptual decisions by decoding neuronal information profiles,” *Biol. Cybern.* **98**, 397–411 (2008).
- [68] P. Seriès, A. A. Stocker, and E. P. Simoncelli, “Is the homunculus ‘aware’ of sensory adaptation?” *Neural Comput.* **21**, 3271–3304 (2009).
- [69] S. S. Stevens, J. Volkman, and E. B. Newman, “A scale for the measurement of the psychological magnitude pitch,” *J. Acoustic. Soc. Am.* **8**, 185–190 (1937).
- [70] V. V. Fedorov, *Theory of Optimal Experiments* (Academic Press, London, UK, 1972).