Perceptual processes as charting operators

Peter Neri[1, 2]
[1]Istituto Italiano di Tecnologia, Via Melen 83, 16152 Genova (Italy)
[2]Laboratoire des Systèmes Perceptifs (CNRS UMR 8248), Département d'études cognitives, École Normale Supérieure, PSL University, CNRS, Paris (France)

Keywords: neurogeometry, Volterra expansion, psychophysical kernels, neural system identification

Abstract

Sensory operators are classically modelled using small circuits involving canonical computations, such as energy extraction and gain control. Notwithstanding their utility, circuit models do not provide a unified framework encompassing the variety of effects observed experimentally. We develop a novel, alternative framework that recasts sensory operators in the language of intrinsic geometry. We start from a plausible representation of perceptual processes that is akin to measuring distances over a sensory manifold. We show that this representation is sufficiently expressive to capture a wide range of empirical effects associated with elementary sensory computations. The resulting geometrical framework offers a new perspective on state-of-the-art empirical descriptors of sensory behavior, such as first-order and second-order perceptual kernels. For example, it relates these descriptors to notions of flatness and curvature in perceptual space.

1 Introduction

In order for quantitative accounts of perception to be practically viable, relevant empirical characterizations must be restricted to settings in which animals respond to sensory stimuli by producing measurable motor outputs. In this study, we further restrict our characterization to the simplified scenario in which there are only two possible classes of stimuli, such as "predator" or "prey", and two motor outputs, such as "fight" or "flight". We will also restrict our characterization to conditions under which the choice between these two alternatives is only driven by the nature of the input stimulus (i.e. it does not depend, for example, on whether the animal is hungry or thirsty). Under this simplified scenario, we can describe the underlying process in the form of a functional $\mathcal{H}$ that maps input vector $\mathbf{s}$ (stimulus) onto binary response r (motor output): $r = \mathcal{H}(\mathbf{s})$.

There are many different ways in which we can study this process experimentally. However, if we are to provide a quantitative characterization of its properties, there is essentially only one way to go about it: psychophysics (Green and Swets, 1966). In psychophysical experiments, different instances of $\mathbf{s}$ are presented to the animal, and each instance is associated with a binary output r. By collecting many such instances of stimulus-response associations, it is hoped that enough information can be gathered to constrain a representation of $\mathcal{H}$ that adequately and parsimoniously predicts r in response to arbitrary $\mathbf{s}$. The term "adequately" means: to saturation of the boundary imposed on prediction by stimulus-independent variability that is intrinsic to the animal (Faisal et al., 2008; Neri, 2010a).

$\mathcal{H}$ can be represented using a diverse range of frameworks and tools (Barack and Krakauer, 2021; Langdon et al., 2023). For example, one may use analytical expressions incorporating a summary characteristic of the stimulus (Zhou et al., 2024) (e.g. contrast), or one may specify a biologically plausible circuit that operates directly on the input image (Schütt and Wichmann, 2017; Lyu et al., 2021). More recently, a common strategy has been to learn the stimulus-response mapping via neural networks that, in some cases, spontaneously acquire properties resembling those associated with specific biological structures (Yamins and DiCarlo, 2016; Neri, 2022). In this article, we adopt an altogether different approach that relies on geometric constructs (Petitot, 2017; Chung and Abbott, 2021; Neri, 2024). Although not transparently relatable to physiological structures, this approach offers a novel unified perspective that is able to account for various empirical observations. Furthermore, this perspective carries the potential to seamlessly integrate those observations into a wider-scale account of sensory computation encompassing a much larger class of perceptual phenomena.

2 Theoretical framework

2.1 From stimulus set to Riemannian manifold structure

The most fundamental assumption underlying our theoretical framework is that the set of all possible stimuli that can be generated for a given experiment can be equipped with Riemannian manifold structure. To establish a topology on this set, we inherit the standard topology from the real-valued vector space associated with the manner in which we store and manipulate stimuli on the computer. We therefore start by formalizing how the latter process is structured in the laboratory.

When we operate on the computer for the purpose of generating stimuli and analyzing them, each stimulus is associated with an element in $\mathbb{R}^n$ (up to floating point precision), and we can perform addition and multiplication on a collection of such elements. The resulting vector space is not closed under those two operations: a physically realizable generator cannot produce any arbitrary stimulus due to hardware limitations, however we assume that vector space structure applies to a satisfactory degree of approximation within the operating range of the experiments. In other words, we assume judicious specification of stimulus parameterization such that hardware capability is not exceeded.

The atlas is established by the perceptual process, and so is a metric structure with

an inner product. When perception is queried for the purpose of obtaining a perceptal judgment, it selects a specific chart and performs measurements on that chart that are then used to generate a behavioral decision. To simplify notation (see section 2.2), we start by adopting chart coordinates that are trivially aligned with the stimulus values stored on the computer, however our framework can accommodate other choices of chart coordinates (we discuss this issue further in Section 5.3).

We make one last assumption: the human observer stores two templates in memory, one associated with a noiseless stimulus containing the target signal to be detected, the other one associated with a noiseless stimulus *not* containing the target signal. These templates map to two elements of the manifold, which essentially amounts to stating that the perceptual templates stored in memory by the perceptual process can be realized, at least in principle, on stimulus hardware.

2.2 Stimulus/template specification with associated notation

With the above framework in mind, each stimulus vector $\mathbf{s}$ is specified by a collection of real coordinate values s^i on the chart instantiated by the perceptual process, and is constructed by combining two component vectors: a signal component and a noise component (see **Figure 1D** for a specific example). The signal component can only take one of two pre-specified configurations: the "target-present" configuration $\mathbf{t}$, and the "target-absent" configuration $\tilde{\mathbf{t}}$. In a typical detection task such as those considered here, $\mathbf{t}$ may correspond to a bright bar in the middle of the monitor (**Figure 1C** shows an example embedded in visual noise), while $\tilde{\mathbf{t}}$ may correspond to the featureless background.

The noise component $\mathbf{n}$ is different for every instance of $\mathbf{s}$. More specifically, each noise coordinate n^i is independently drawn from a zero-mean Gaussian distribution. Locally, each stimulus coordinate is specified by either $s^i{=}t^i{+}n^i$ when the stimulus is presented in target-present configuration, or $\tilde{s}^i{=}\tilde{t}^i{+}\tilde{n}^i$ when the stimulus is presented in target-absent configuration. In this study, every component of $\tilde{\mathbf{t}}$ is (without loss of generality) specified to be zero (hence the term "target-absent"), so that $\tilde{s}^i{=}\tilde{n}^i$.

The two memory templates stored by the human observer, one for the target-present configuration and another one for the target-absent configuration, are denoted by vectors $\mathbf{w}$ and $\tilde{\mathbf{w}}$, respectively. In general, $\mathbf{w}{\neq}\mathbf{t}$ and $\tilde{\mathbf{w}}{\neq}\tilde{\mathbf{t}}$: the memorized representations of target-present and target-absent stimuli, which exist in the head of the human observer, are not necessarily aligned with the physical specifications of those stimuli, which exist on the hardware that is used to display them.

We draw attention to a slight abuse of the $\tilde{\cdot}$ notation above. It is always true that $\mathbf{w}{\neq}\tilde{\mathbf{w}}$ (the system otherwise carries no discriminative power). In general, it is also the case that $\mathbf{t}{\neq}\tilde{\mathbf{t}}$ (except when specified otherwise by the experimenter, see for example Section 3.2 and **Figure 2C**,**F**). However, $\langle\mathbf{n}\rangle{=}\langle\tilde{\mathbf{n}}\rangle$ where $\langle.\rangle$ is used to denote average over many trials (and $=$ is intended asymptotically). In other words, the statistical specification of the noise component $\mathbf{n}$ is identical between target-present and target-absent stimuli: when applied to the noise component, the $\tilde{\cdot}$ notation is used to refer to noise samples $\tilde{\mathbf{n}}$ associated with the target-absent stimulus ($\tilde{\mathbf{s}}$) to distinguish them from those ($\mathbf{n}$) associated with the target-present stimulus ($\mathbf{s}$), however these two different types of noise samples conform to identical statistical specifications. In particular, $\langle n^i\rangle{=}\langle\tilde{n}^i\rangle{=}0$

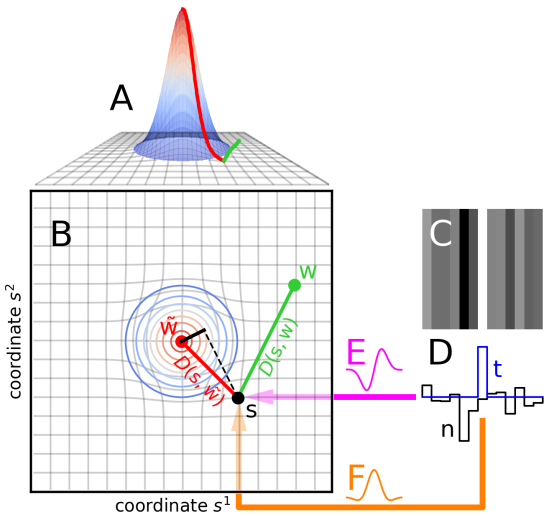

Figure 1: Stimulus **s** maps to a vector (black dot) in metric perceptual space (**B**). Memory templates map to vectors **w** (green dot) for target-to-be-detected, and $\tilde{\mathbf{w}}$ (red dot) for absence of said target. The perceptual system measures distance $D(\mathbf{s}, \mathbf{w})$ between stimulus **s** and template **w** (green segment), and distance $D(\mathbf{s}, \tilde{\mathbf{w}})$ between **s** and $\tilde{\mathbf{w}}$ (red segment). Based on these measurements, the system can establish whether **s** is closer to **w** or to $\tilde{\mathbf{w}}$, and classify it accordingly. In flat space (**B**), taking the difference between $D(\mathbf{s}, \mathbf{w})$ and $D(\mathbf{s}, \tilde{\mathbf{w}})$ is equivalent (i.e. returns same perceptual decision) to measuring the projection of **s** onto **w** (black segment). For the example shown here, **s** is closer to $\tilde{\mathbf{w}}$ than it is to **w** (red segment is shorter than green segment). In curved space (**A**), distance measurements can lead to stimulus classifications that differ from those obtained in flat space: **s** is now closer to **w** than it is to $\tilde{\mathbf{w}}$ (green path is shorter than red path). To visualize higher-dimensional stimuli (**C**) on the plane (**B**), we project them onto vectors **E**,**F**. Each stimulus consists of two components: target and noise (**t**,**n** respectively in **D**).

(zero-mean), $\langle n^i n^j \rangle = \langle \tilde{n}^i \tilde{n}^j \rangle = \delta_{ij}$ (uncorrelated coordinate values), and $\langle n^i \tilde{n}^j \rangle = 0$ (uncorrelated samples).

2.3 Perceptual distance

Perceptual judgments are based on distance measurements on the stimulus manifold. To perform such measurements, we posit the existence of a metric with associated inner product. For a given choice of chart and at a given point on the manifold, we can express this product as $g_{ij} x^i y^j$ (Einstein notation), where x^i / y^i are the chart coordinates of vectors **x**/**y**. Using the inner product, we can readily define kinetic energy along a curve γ running between stimulus **s** and memory template **w** over the manifold. With $\gamma(0) = \mathbf{w}$ and $\gamma(1) = \mathbf{s}$, this quantity can be expressed as:

$$E(\mathbf{s}, \mathbf{w}) = \int_0^1 g_{ij}\big(\gamma(t)\big)\Big(\frac{d}{dt} x^i \circ \gamma(t)\Big)\Big(\frac{d}{dt} x^j \circ \gamma(t)\Big) dt$$

We assume that the perceptual process can measure this quantity along a geodesic path to obtain an energy-based measure of distance D:

$$D(\mathbf{s}, \mathbf{w}) = \min_{\gamma \in \mathcal{P}(\mathbf{s}, \mathbf{w})} \big(E\big)$$

where $\mathcal{P}(\mathbf{s}, \mathbf{w})$ is the collection of all curves running between **s** and **w**. The issue of whether and how the perceptual process may be able to identify a geodesic path is complex, and we return to this point in Section 5.3.

Without detailed knowledge of the metric specification (which we cannot acquire from isolated behavioral experiments, see Section 5.1), the above expressions are useless for applied purposes. To take a first step in the direction of deriving useful expressions (to be further developed in Section 2.6), we notice that, because D is a scalar value, we can always specify a matrix g such that the same value is returned by the simpler expression:

$$D(\mathbf{s}, \mathbf{w}) = g_{ij}\big(\mathbf{s}, \mathbf{w}\big)(s^i - w^i)(s^j - w^j) \tag{1}$$

where $g_{ij}\big(\mathbf{s}, \mathbf{w}\big)$ can, in principle, differ for every choice of **s** and **w**. We refer to g as the "ambient metric", because g no longer corresponds to the metric at a given point on the manifold, but rather summarizes metric structure along a geodesic path between **s** and **w**.

The perceptual process only computes two types of distance D: between stimulus **s** (or $\tilde{\mathbf{s}}$) and target-present template **w**, and between **s** (or $\tilde{\mathbf{s}}$) and target-absent template $\tilde{\mathbf{w}}$. We reserve the symbol g for the former, and $\tilde{g}$ for the latter. We can therefore write $D(\mathbf{s}, \mathbf{w}) = g_{ij}(\mathbf{s})(s^i - w^i)(s^j - w^j)$ with the dependence of g on **w** implicit, and similarly $D(\mathbf{s}, \tilde{\mathbf{w}}) = \tilde{g}_{ij}(\mathbf{s})(s^i - \tilde{w}^i)(s^j - \tilde{w}^j)$ with the dependence of $\tilde{g}$ on $\tilde{\mathbf{w}}$ implicit.

To visualize the above procedure, we imagine that the perceptual system "places itself" at the element corresponding to **s** on the manifold (black point in **Figure 1B**). From this vantage point, it traces out a geodesic path to **w** (green path/segment in **Figure 1A/B**) along a Cartesian coordinate chart (i.e. a chart aligned with the geodesic), and measures $D(\mathbf{s}, \mathbf{w})$ using metric components that are collectively summarized by g_{ij}. It then repeats this procedure with respect to $\tilde{\mathbf{w}}$ (red path/segment in **Figure 1A/B**) to obtain $D(\mathbf{s}, \tilde{\mathbf{w}})$ using metric components summarized by $\tilde{g}_{ij}$. Clearly, it is meaningless to speak of the perceptual system in the manner verbalized above (i.e. as a homunculus operating within itself). This description is adopted here for intuitive purposes only, with no implication of any resemblance to the natural phenomenon.

2.4 Decision variables

We restrict our analysis to the following experimental scenario (two alternative forced choice protocol (Green and Swets, 1966)): on each trial, the human observer is presented with both **s** and $\tilde{\mathbf{s}}$, and is asked to select **s**. In the mind of the human observer, selecting **s** must be understood as selecting, between **s** and $\tilde{\mathbf{s}}$, the stimulus most closely resembling **w**.

In line with established accounts of perceptual decision-making (Pritchett and Murray, 2015), we posit that the human observer assigns a scalar value to each stimulus reflecting the likelihood that said stimulus is **s**. We refer to this value with the term "decision variable". The decision variable associated with the target-present stimulus is $R(\mathbf{s})$, while the decision variable associated with the target-absent stimulus is $R(\tilde{\mathbf{s}})$. The observer selects the stimulus associated with largest decision variable as corresponding to **s**. Therefore, if $R(\mathbf{s}) > R(\tilde{\mathbf{s}})$ the observer will respond correctly, otherwise he/she will respond incorrectly.

Because the human observer is intrinsically noisy (Neri, 2010a), we must model human decision-making probabilistically (Green and Swets, 1966; Neri, 2010b; Caran-

dini, 2024). We write the probability that a given response is correct as:

$$\Psi\big(R(\mathbf{s}) - R(\tilde{\mathbf{s}})\big) \tag{2}$$

where we assume that the static nonlinear transducer Ψ only depends on the difference between $R(\mathbf{s})$ and $R(\tilde{\mathbf{s}})$. A typical choice for Ψ is the cumulative Gaussian distribution function (Green and Swets, 1966), however there is empirical evidence that other distributions may capture perceptual noise more accurately (Neri, 2013) (see also Section A.2).

We consider three rules for generating the decision variable R. The first rule reflects the most commonly adopted model in computational neuroscience (Heeger et al., 1996; Dayan and Abbott, 2001; Ostojic and Brunel, 2011): the linear-nonlinear (LN) cascade, also termed template-matching in the behavioral literature (Brunelli and Poggio, 1997). The corresponding decision variable is:

$$\underset{\perp}{R}(\mathbf{s}) = \delta_{ij} w^i s^j \tag{3}$$

The above operation, which amounts to projecting stimulus $\mathbf{s}$ onto the target-present template $\mathbf{w}$ stored in memory by the observer (hence symbol $\perp$ standing for "projection rule"), is entirely linear, however it is (inevitably) followed by the nonlinear operator Ψ (expression 2). This rule does not refer back to the metric g, in fact it is not even sensibly defined in the context of the geometrical framework we introduced earlier, except for the specific case of a flat geometry, for which g does not vary across the manifold. In that case, we could write $\underset{\perp}{R}(\mathbf{s}) = g_{ij} w^i s^j$, however this is equivalent to equation 3 with g absorbed into $\mathbf{w}$.

The second rule reflects the naive strategy of simply measuring the distance between stimulus $\mathbf{s}$ and $\mathbf{w}$. The corresponding decision variable is:

$$\underset{\vdash}{R}(\mathbf{s}) = -D(\mathbf{s}, \mathbf{w}) \tag{4}$$

where $\vdash$ stands for "absolute distance rule". The minus sign is introduced because the observer must now select the stimulus associated with smallest, not largest, $D(\mathbf{s}, \mathbf{w})$ (the smaller this quantity, the closer the stimulus to the target-present template).

The third rule is similar to the second one, except it explicitly takes into account the memory template $\tilde{\mathbf{w}}$ associated with the target-absent signal. This element is often neglected in perceptual models of detection tasks, because the target-absent stimulus typically consists of a blank screen, and is therefore regarded as a non-stimulus that has no impact on the final outcome. For example, if $\underset{\perp}{R}(\mathbf{s})$ were reformulated to incorporate the term $\delta_{ij}\tilde{w}^i s^j$, this term would have no impact because equal to zero for $\tilde{w}^i{=}0$. In the present framework, $\tilde{\mathbf{w}}$ can always be set to 0 by redefining $\mathbf{w}$ as $\mathbf{w}{-}\tilde{\mathbf{w}}$. However, particularly when observers engage in two-way classification judgments, it is reasonable to posit that the stimulus-to-be-classified should be compared against both target-present and target-absent memory templates, to adjucate whether it is more similar to the former or to the latter. We can formulate the corresponding decision variable as:

$$\underset{\Vdash}{R}(\mathbf{s}) = D(\mathbf{s}, \tilde{\mathbf{w}}) - D(\mathbf{s}, \mathbf{w}) \tag{5}$$

where $\Vdash$ stands for "relative distance rule". Notice that, in the expression above, the term $D(\mathbf{s}, \tilde{\mathbf{w}})$ does not necessarily equal zero if we choose chart coordinates centred on $\tilde{\mathbf{w}}$. This is the essential characteristic that sets $\underset{\Vdash}{R}$ apart from $\underset{\vdash}{R}$: these two rules cannot be rendered equivalent by simply shifting the coordinate frame.

2.5 Kernel computation from data

To characterize $\mathcal{H}$ in the laboratory, we must derive empirical descriptors of its characteristics. We focus on two established classes of descriptors (Murray, 2011): first-order (Ahumada, 2002) and second-order kernels (Neri, 2010b). These descriptors are not arbitrary: they represent the only logical approach for providing a transparent statistical description of input-output coupling in elementary perceptual judgments (Murray, 2011; Neri, 2018a). In this sense, we are not focusing here on some exotic, abstruse empirical representation of questionable relevance. To the contrary, we are taking as point of reference an indispensable class of descriptors for characterizing noise-based measurements of perceptual processing (Ahumada, 2002; Murray, 2011; Neri, 2018a).

For the first-order class, we compute target-present component h_i and target-absent component $\tilde{h}_i$ separately (Ahumada and Lovell, 1971) (see Section 3.1 for theoretical and empirical motivations behind this distinction). For the second-order class, we only compute the composite descriptor h_{ij} incorporating both target-present and target-absent components (Neri, 2004) (see Section 2.6 for theoretical justification of this latter choice).

We derive first-order kernels via the following calculations (Ahumada, 2002):

$$\begin{aligned} h_i &= \langle n^i \rangle_{\checkmark} - \langle n^i \rangle_{\times} \\ \tilde{h}_i &= \langle \tilde{n}^i \rangle_{\times} - \langle \tilde{n}^i \rangle_{\checkmark} \end{aligned} \tag{6}$$

where subscript $\checkmark$ (or $\times$) refers to the subset of trials on which the observer responded correctly (or incorrectly).

To compute the second-order kernel, we replace averaging $\langle n^i \rangle$ with covariance $\langle n^i n^j \rangle - \langle n^i \rangle \langle n^j \rangle$ in the two expressions above (Neri, 2004), and sum them to obtain:

$$\begin{aligned} h_{ij} = {} & \langle n^i n^j \rangle_{\checkmark} - \langle n^i \rangle_{\checkmark} \langle n^j \rangle_{\checkmark} - \langle n^i n^j \rangle_{\times} + \langle n^i \rangle_{\times} \langle n^j \rangle_{\times} \\ & + \langle \tilde{n}^i \tilde{n}^j \rangle_{\times} - \langle \tilde{n}^i \rangle_{\times} \langle \tilde{n}^j \rangle_{\times} - \langle \tilde{n}^i \tilde{n}^j \rangle_{\checkmark} + \langle \tilde{n}^i \rangle_{\checkmark} \langle \tilde{n}^j \rangle_{\checkmark} \end{aligned} \tag{7}$$

These empirical descriptors share some characteristics with their electrophysiological counterparts, the spike-triggered average (STA) (Rieke, 1999) and spike-triggered covariance (STC) (Schwartz et al., 2002), however they also present important differences. In particular, the second-order kernel above is characterized by a difference operation in which covariance matrices are subtracted for some response classes (h_{ij} is not necessarily expected to be positive definite, while STC always is). This simple fact has fundamental implications for its geometrical interpretation (Section 2.7) that set it apart from STC. In the case of first-order kernels, the difference operation is also present (trailing terms in equations 6), however it bears little conceptual significance for the comparison between these kernels and the STA.

2.6 Kernel derivation under plausible approximations

Because the decisional rules considered in Section 2.4 only contain quadratic elements, their differential output $\Delta R = R(\mathbf{s}) - R(\tilde{\mathbf{s}})$ can be written as:

$$\Delta R = v_i n^i + \tilde{v}_i \tilde{n}^i + v_{ij} n^i n^j + \tilde{v}_{ij} \tilde{n}^i \tilde{n}^j + \text{terms not dependent on } \mathbf{n}, \tilde{\mathbf{n}} \tag{8}$$

which is a simplified second-order Volterra expansion in $\mathbf{n}$ and $\tilde{\mathbf{n}}$. This is the argument submitted to Ψ in expression 2. If we approximate Ψ to first-order (linear), we know from previous work (Neri, 2010b) that:

$$\begin{aligned} h_i &\propto v_i \\ \tilde{h}_i &\propto -\tilde{v}_i \\ h_{ij} &\propto 2(v_{ij} - \tilde{v}_{ij}) \end{aligned} \tag{9}$$

where $\propto$ stands for the same proportionality factor across the three expressions. These results prove reasonably robust for substantial deviations of Ψ from its first-order approximation, with the following caveats in mind. First, if $v_{ij}{=}\tilde{v}_{ij}{=}0$ (i.e. ΔR is linear with no quadratic terms), the expressions above apply for any smooth Ψ, without the need for low-order approximations (Bussgang, 1952; Neri, 2010b). Second, readers may be tempted to infer that one may compute target-present and target-absent components of the second-order kernel separately and that, in similar fashion to the first-order kernels, these components may be simply expressed as $\propto v_{ij}$ and $\propto \tilde{v}_{ij}$. This is only true for a first-order approximation of Ψ, but quickly becomes inapplicable for small deviations from this low-order approximation (Neri, 2010b). For this reason, we only compute the composite second-order kernel h_{ij} derived from both target-present and target-absent noise fields.

We now proceed by rewriting ΔR for each rule in the form of equation 8. For the projection rule (equation 3):

$$\Delta \underset{\perp}{R} = \delta_{ij} w^i s^j - \delta_{ij} w^i \tilde{s}^j = \delta_{ij} w^i n^j + \delta_{ij} w^i \tilde{n}^j + \text{other terms} \tag{10}$$

We therefore have:

$$\underset{\perp}{h}_i = \underset{\perp}{\tilde{h}}_i \propto w_i \tag{11}$$

$$\underset{\perp}{h}_{ij} = 0 \tag{12}$$

This is an established theoretical result (Ahumada, 2002). It has been observed in the laboratory under specific conditions (Beard and Ahumada, 1998; Neri et al., 1999; Neri, 2015) (see Section 3.2 and **Figure 2A**), however it is relatively uncommon that human empirical kernels comply with these specifications: it is more commonly found that $h_i{\neq}\tilde{h}_i$ and $h_{ij}{\neq}0$ (Ahumada and Lovell, 1971; Dai et al., 1996; Barth et al., 1999; Abbey and Eckstein, 2002; Neri and Heeger, 2002; Solomon, 2002; Tjan and Nandy, 2006; Shub and Richards, 2009; Joosten and Neri, 2012; Joosten et al., 2016; Varnet and Lorenzi, 2022).

The absolute distance rule (expression 4) involves measure D (expression 1), which we can expand as:

$$D(\mathbf{s}, \mathbf{w}) = 2g_{ij}(\mathbf{s}) t^i n^j - 2g_{ij}(\mathbf{s}) w^i n^j + g_{ij}(\mathbf{s}) n^i n^j - 2g_{ij}(\mathbf{s}) t^i w^j + g_{ij}(\mathbf{s}) t^i t^j + g_{ij}(\mathbf{s}) w^i w^j$$

This expression may appear to conform to expression 8, however there is an important element of departure: $g_{ij}(\mathbf{s})$ depends on $\mathbf{s}$, which in turn depends on its varying component $\mathbf{n}$. To proceed further, it would be necessary to render this dependence explicit.

At the current stage in our knowledge and understanding of human perception, we do not possess an explicit description for the manner in which the metric varies across

the perceptual manifold. We may assume, for the purpose of attempting a minimal mathematical treatment, that there is an underlying connection with Riemannian characteristics, but we are not in a position to formulate an explicit specification for this connection. To derive expressions that provide any useful level of insight into the perceptual process, we must therefore adopt some approximations. The most important approximation we adopt here is the following:

$$\langle g_{ij}(\mathbf{s})n^{k_1}\cdots n^{k_M}\rangle = g_{ij}\big(\langle\mathbf{s}\rangle\big)\langle n^{k_1}\cdots n^{k_M}\rangle \tag{13}$$

and similarly for the $\tilde{\cdot}$ variants $\tilde{g}_{ij}(\mathbf{s})$, $g_{ij}(\tilde{\mathbf{s}})$, and $\tilde{g}_{ij}(\tilde{\mathbf{s}})$, with $\tilde{n}^k$ substituted for n^k when using $\tilde{\mathbf{s}}$. In other words, it is assumed that the ambient metric is uncorrelated with local noisy modulations along any coordinate around $\langle\mathbf{s}\rangle$ as measured with respect to $\mathbf{w}$. To simplify calculations further, we also adopt the approximation:

$$g_{ij}\big(\langle\mathbf{s}\rangle\big)\approx g_{ij}\big(\langle\tilde{\mathbf{s}}\rangle\big) \tag{14}$$

and similarly for $\tilde{g}_{ij}$. This approximation is reasonable if the intensity of target signal $\mathbf{t}$ is relatively small compared with the noisy modulations applied to $\mathbf{s}$ and $\tilde{\mathbf{s}}$ (in the extreme case of $\mathbf{t}$=0, $\langle\mathbf{s}\rangle$=$\langle\tilde{\mathbf{s}}\rangle$). The above approximation therefore only applies for low-to-moderate stimulus signal-to-noise ratios (SNR), which is the typical SNR regime adopted in kernel estimation experiments (Murray et al., 2002; Murray, 2011) (see Section 3.5 and **Figure 5F**,**I**,**J** for further analysis/discussion of SNR-related issues).

With the above approximations, the entire geometry of the perceptual manifold is summarized by only two metric specifications ($\langle\mathbf{s}\rangle$ and $\langle\tilde{\mathbf{s}}\rangle$ are fixed for a given configuration/characterization): g_{ij}, which captures the ambient geometry around the memory template $\mathbf{w}$ associated with the perceptual representation of the target-to-be-detected, and $\tilde{g}_{ij}$, which captures the ambient geometry around the target-absent template $\tilde{\mathbf{w}}$. This degree of approximation may seem excessive, however we demonstrate that it is adequately matched to the degree of resolution afforded by current empirical measurements of relevant perceptual phenomena (Section 3). In the absence of explicit knowledge about the metric connection and its dependence on stimulus attributes, our only point of reference for empirical validation is provided by first-order and second-order kernel estimates. These estimates are coarse in the following sense: they cannot be obtained for single-shot presentations, but rather must be constructed on the basis of several input-output pairings across thousands of trials (Section 2.5). The resulting degree of characterization is compatible with the degree of approximation afforded by our analytical treatment here (see Sections 3.4, 4.1, and 5.1 for further discussion of relevant issues).

We can now approximate ΔR for the absolute distance rule as:

$$\Delta \underset{\vdash}{R} = 2g_{ij}(w^i - t^i)n^j - 2g_{ij}w^i\tilde{n}^j - g_{ij}n^in^j + g_{ij}\tilde{n}^i\tilde{n}^j + \text{other terms} \tag{15}$$

which conforms to equation 8. Notice that we only allow for this approximation to apply in the context of operating across many trials, which is how we calculate kernels

from data (equations 6–7 in Section 2.5). Using equations 9, we have:

$$\underset{\vdash}{h}_i \propto g_{ij}w^j - g_{ij}t^j \tag{16}$$

$$\underset{\vdash}{\tilde{h}}_i \propto g_{ij}w^j \tag{17}$$

$$\underset{\vdash}{h}_{ij} \propto -2g_{ij} \tag{18}$$

By adopting a similar procedure, we obtain the following approximation for the relative distance rule:

$$\begin{aligned}\Delta\underset{\Vdash}{R} = {} & 2(g_{ij}w^i - \tilde{g}_{ij}\tilde{w}^i - g_{ij}t^i + \tilde{g}_{ij}t^i)n^j - 2(g_{ij}w^i - \tilde{g}_{ij}\tilde{w}^i)\tilde{n}^j \\ & - (g_{ij} - \tilde{g}_{ij})n^i n^j + (g_{ij} - \tilde{g}_{ij})\tilde{n}^i\tilde{n}^j + \text{other terms}\end{aligned}$$

and the following kernel predictions (again from equations 9):

$$\underset{\Vdash}{h}_i \propto g_{ij}w^j - \tilde{g}_{ij}\tilde{w}^j + (\tilde{g}_{ij} - g_{ij})t^j \tag{19}$$

$$\underset{\Vdash}{\tilde{h}}_i \propto g_{ij}w^j - \tilde{g}_{ij}\tilde{w}^j \tag{20}$$

$$\underset{\Vdash}{h}_{ij} \propto 2(\tilde{g}_{ij} - g_{ij}) \tag{21}$$

At this stage, we make the following observation: the relative distance rule subsumes the projection rule as a special case for $g_{ij}{=}\tilde{g}_{ij}{=}\delta_{ij}$ and $\tilde{\mathbf{w}}{=}0$. This can be easily verified by substituting those conditions into the expression for $\Delta\underset{\Vdash}{R}$, which returns $\Delta\underset{\perp}{R}$. We can further relax the above conditions: $\tilde{\mathbf{w}}$ can easily be made to equal the zero vector by simply shifting the coordinate frame, making condition $\tilde{\mathbf{w}}{=}0$ inconsequential; $g_{ij}{=}\tilde{g}_{ij}$ need not be equal to δ_{ij}, as we can absorb g_{ij} into $\mathbf{w}$ by assigning to each element w^i the value returned by $g_{ij}w^j$ (in Euclidean space with Cartesian coordinates, vectors and covectors can be interchanged). We can summarize the above by stating that the relative distance rule collapses onto the projection rule when $g_{ij}{=}\tilde{g}_{ij}$. In the next section (2.7), we provide a geometrical interpretation of this fact.

The above observation does not apply to the absolute distance rule: $\Delta\underset{\vdash}{R}$ can only be made to equal $\Delta\underset{\perp}{R}$ for the degenerate case in which they are both identically 0 (via setting $g_{ij}{=}0$), which is not a viable choice of decision strategy.

2.7 Geometrical interpretation

We can visualize all three decisional rules within a common geometrical framework. We first imagine placing vectors $\mathbf{w}$, $\tilde{\mathbf{w}}$, and $\mathbf{s}$ onto the coordinate frame (**Fig. 1B**). Without loss of generality, we can arbitrarily place the origin of the frame at $\tilde{\mathbf{w}}$, so that $\tilde{\mathbf{w}}$ is the zero-coordinate vector. The projection rule amounts to projecting $\mathbf{s}$ onto $\mathbf{w}$ (black segment in **Fig. 1B**). The absolute distance rule amounts to measuring the distance between $\mathbf{s}$ and $\mathbf{w}$ (green segment in **Fig. 1B**). The relative distance rule amounts to measuring the distance between $\mathbf{s}$ and $\mathbf{w}$ (as in the absolute distance rule), then measuring the distance between $\mathbf{s}$ and $\tilde{\mathbf{w}}$ (red segment in **Fig. 1B**), and finally taking the difference between these two distances.

On the flat plane endowed with a Euclidean metric (bottom part of **Fig. 1B**), the relative distance rule is equivalent to the projection rule (observation at the end of the previous section): when $g_{ij}{=}\tilde{g}_{ij}{=}\delta_{ij}$ (Euclidean space), the two rules return identical

ΔR values. In curved space (or a space endowed with curvilinear coordinates, see Section 5.3 for futher discussion of this point), the two rules differ greatly: in the example shown in **Fig. 1**, distance measurements in flat space (**Fig. 1B**) lead to the conclusion that **s** is more similar to $\tilde{\mathbf{w}}$ because it is closer to this vector (red segment is shorter than green segment); in curved space (**Fig. 1A**), distance measurements lead to the opposite conclusion that **s** is more similar to **w** because it is closer to this vector (green path is shorter than red path).

In our framework, the difference between flat space and curved space is captured by $\tilde{g}_{ij} - g_{ij}$: if this difference is 0, the components of the metric are flat (invariant across space) and the geometry collapses onto **Fig. 1B** (although see Section 3.4 for a demonstration of how this statement is not entirely accurate); if this difference is not zero, the components of the metric vary across space and the geometry lives on the curved surface in **Fig. 1A**. For the relative distance rule, the second-order kernel reflects precisely that difference (equation 21), thus making it a useful empirical descriptor of curvature in perceptual space. This observation is, in a nutshell, the primary motivation for the theoretical framework proposed here.

3 Experimental framework

3.1 Structural failures versus fitting exercises

Our goal is to construct a theory of perceptual processing, not a specific model of this phenomenon. To arbitrate between competing theories, we must avoid engaging in the class of fitting exercises that are often utilized for model selection, because we want mathematical structure, not parameterization, to bear on the arbitration process (Roberts and Pashler, 2000). This means that, if we exclude theory X, we do this because theory X is structurally incapable of accounting for the empirical evidence: it is the architecture of theory X that cannot capture the experimental result, not a specific choice of parameter specification for that architecture (Roberts and Pashler, 2000; Veksler et al., 2015).

For this type of selection programme to be enforceable, the most effective class of relevant experimental findings is such that a specific measurement is either 0 or not 0, and competing theories predict that it should be either 0 or not 0. When a theory predicts 0 (or not 0) and the empirical evidence returns not 0 (or 0), we speak of a "structural failure" of that theory (Neri, 2015). Notice that structural failures of this kind are not all equally compelling: if the theory predicts $\neq 0$ and the evidence provides 0, this is weaker support for excluding the theory than the complementary scenario in which the theory predicts 0 and the evidence provides $\neq 0$, because the absence of a measurable effect does not imply that the effect is absent.

If evidence of the kind outlined above is lacking, the only admissible evidence should be such that the theory makes relevant predictions that do not involve any free parameters (see Section 3.5).

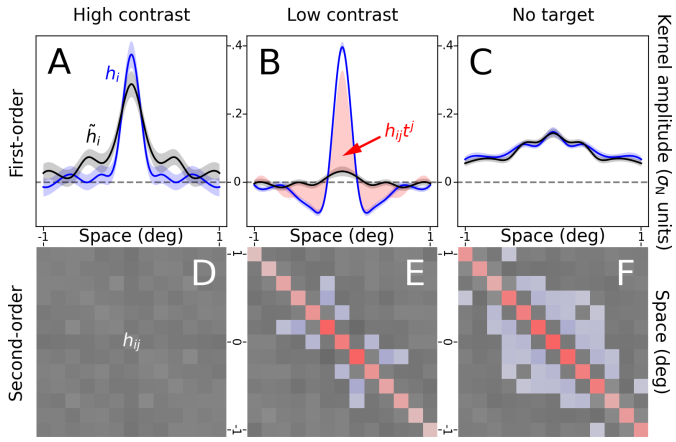

Figure 2: Empirically measured first-order (top row) and second-order (bottom row) kernels may vary independently as a function of specific stimulus manipulations. At high stimulus contrast (left column), target-present (blue) and target-absent (black) first-order kernels are not experimentally distinguishable (**A**), while the second-order kernel is featureless (**D**). At low contrast (middle column), the opposite pattern is observed: target-present and target-absent kernels are markedly different (**B**), while the second-order kernel is highly structured (**E**). When the target signal is removed from the stimulus (right column), target-present and target-absent kernels become indistinguishable by definition (and experimentally, see **C**), however the second-order kernel remains highly structured (**F**). Shaded regions in **A**–**C** show ± 1 SEM. Red/blue-tinted pixels in **D**–**F** indicate significant positive/negative modulations (|Z-score|>2). Red-tinted area in **B** shows theoretically predicted difference between blue and black traces (equation 22, Section 3.5). Kernel amplitude in **A**–**C** (y axis) is plotted in units of external noise SD. Each of the first two columns is computed from ≈35K trials of published data (Neri, 2018b) aggregated across 8 observers. Third column is computed from ≈32K trials of unpublished data aggregated across 10 observers.

3.2 Interdependence (and lack thereof) of first-order and second-order kernel estimates

We focus on two important empirical characteristics of kernel estimates: the presence/absence of potential differences between target-present and target-absent first-order kernels ($h_i \neq \tilde{h}_i / h_i \approx \tilde{h}_i$), and the presence/absence of potential modulations within the second-order kernel ($h_{ij} \neq 0 / h_{ij} \approx 0$). These quantities can be measured robustly with sufficiently large datasets (Ahumada and Lovell, 1971; Dai et al., 1996; Barth et al., 1999; Abbey and Eckstein, 2002; Neri and Heeger, 2002; Solomon, 2002; Tjan and Nandy, 2006; Shub and Richards, 2009; Joosten and Neri, 2012; Joosten et al., 2016; Neri, 2018b; Varnet and Lorenzi, 2022). By "absence" we mean not distinguishable from 0 within the resolution of our empirical measurements. Clearly, it is never possible to state that a given effect is completely absent: there always remains a residual possibility that the effect is actually present, but too small to be measured. Notwithstanding this inevitable difficulty with interpreting experimental results, we regard the resolution of our measurements to be sufficiently high for us to draw meaningful conclusions.

Figure 2 presents three extreme cases of absence/presence for the characteristics just detailed, which encompass the entire range observed in the laboratory. The first two columns in **Figure 2** (panels **A**,**B**,**D**,**E**) come from the same visual experiment (Neri, 2018b), the only difference being stimulus contrast: high in **Figure 2A**,**D**, low in **Figure 2B**,**E** (please refer to the original publication (Neri, 2018b) for further details

on stimulus specification). Observers were asked to discriminate between target-present and target-absent stimuli structured as shown in **Figure 1C**. **Figure 2C** was computed from data collected during a virtually identical experiment (unpublished), with the only difference that no target was added to the target-present stimulus (on one out of three trials, unbeknownst to participants): $\mathbf{t}=0$.

At high contrast, there is little to no measurable difference between target-present (blue) and target-absent (black) kernels (**Figure 2A**), and there are no modulations within the second-order kernel (**Figure 2D**). At low contrast, there is a clear difference between target-present and target-absent kernels (**Figure 2B**), accompanied by substantial structure within the second-order kernel (**Figure 2E**).

The above transition $h_i \sim \tilde{h}_i \rightarrow h_i \neq \tilde{h}_i$, $h_{ij} \sim 0 \rightarrow h_{ij} \neq 0$ raises the following question: does $h_i \sim \tilde{h}_i$ imply $h_{ij}=0$, and/or does $h_i \neq \tilde{h}_i$ imply $h_{ij} \neq 0$?

$h_i \sim \tilde{h}_i$ does not imply $h_{ij} \sim 0$. It is easy to enforce $h_i - \tilde{h}_i = 0$: in the absence of a target signal, target-present and target-absent denominations remain undefined and $h_i = \tilde{h}_i$ by definition, as also confirmed experimentally (**Figure 2C**). When analyzing experimental data under these conditions, trials must be randomly classified as target-present or target-absent, rendering this distinction only useful for gauging the extent of measurement noise (which is small, as indicated by nearly identical estimates in **Figure 2C**). The corresponding h_{ij} is nevertheless clearly structured ($h_{ij} \neq 0$, **Figure 2F**).

$h_i \neq \tilde{h}_i$ does imply $h_{ij} \neq 0$ (at least empirically). It is never observed experimentally that $h_i \neq \tilde{h}_i$ and $h_{ij} \sim 0$, except for pathologically high SNR values (Section 3.5, **Figure 5F**,**H**), when our approximations for deriving kernel estimates (see Section 2.6) become inapplicable. In other words, if we classify all possible scenarios for the two characteristics under examination into binary absent/present categories, only three out of the four possible scenarios have been reported by previous studies: $h_i \neq \tilde{h}_i / h_{ij} \neq 0$ may be absent/absent (**Figure 2A/D**), present/present (**Figure 2B/E**), or absent/present (**Figure 2C/F**).

3.3 (In)compatibility of different rules with experimental evidence

The empirical results in **Figure 2** impose stringent constraints on candidate decisional rules, allowing us to arbitrate between them.

We start with the projection rule ($\perp$). This rule is incompatible with the presence of a measurable difference between target-present and target-absent first-order kernels: it dictates that $\underset{\perp}{h}_i = \underset{\perp}{\tilde{h}}_i$ (equation 11). It is also incompatible with the presence of measurable modulations within the second-order kernel: $\underset{\perp}{h}_{ij}=0$ (equation 12). Of the three experimental scenarios in **Figure 2**, it can only explain the high-contrast condition (left column). Furthermore, in its basic formulation, it is inconsistent with an iconic result in the perceptual literature known as the "dipper" effect (Appendix A). This rule must therefore be discarded because lacking sufficient flexibility to encompass the range exposed by laboratory experiments (Neri, 2015), despite being by far the most common model in sensory neuroscience (Heeger et al., 1996; Brunelli and Poggio, 1997; Dayan and Abbott, 2001; Ostojic and Brunel, 2011; Neri, 2015).

The absolute distance rule ($\vdash$) allows for both absence and presence of a difference between target-present and target-absent first-order kernels: this difference is controlled by the $-g_{ij}t^j$ term in equations 16–17. When g_{ij} is set to 0, $\underset{\vdash}{h}_i = \underset{\vdash}{\tilde{h}}_i$; when g_{ij} is struc-

tured, $\underset{\vdash}{h}_i \neq \underset{\vdash}{\tilde{h}}_i$. However, setting g_{ij} to 0 is not a viable option for three reasons. First, $g_{ij}=0$ enforces $\underset{\vdash}{h}_i = \underset{\vdash}{\tilde{h}}_i = 0$, which is contrary to experimental evidence: when $\underset{\vdash}{h}_i = \underset{\vdash}{\tilde{h}}_i$, they are nevertheless structured (**Figure 2A**,**C**). Second, $g_{ij}=0$ enforces $\underset{\vdash}{h}_{ij}=0$, which is again contrary to experimental evidence: the second-order kernel is often structured (**Figure 2E**,**F**). Third, setting $g_{ij}=0$ enforces $\Delta \underset{\vdash}{R}=0$ (equation 15), making the system unable to produce any meaningful measurement. Like the projection rule, the absolute distance rule must therefore be discarded because lacking sufficient flexibility.

In the relative distance rule ($\Vdash$), the difference between target-present and target-absent first-order kernels is controlled by the $(\tilde{g}_{ij} - g_{ij})t^j$ term in equations 19–20. Furthermore, the first factor in this term $(\tilde{g}_{ij} - g_{ij})$ is proportional to the second-order kernel (equation 21). The relative distance rule therefore allows for both absence and presence of a difference between first-order kernels, and for absence and presence of structured modulations within the second-order kernel: when $\tilde{g}_{ij} = g_{ij}$, $\underset{\Vdash}{h}_i = \underset{\Vdash}{\tilde{h}}_i$ and $\underset{\Vdash}{h}_{ij}=0$; when $\tilde{g}_{ij} \neq g_{ij}$, $\underset{\Vdash}{h}_i \neq \underset{\Vdash}{\tilde{h}}_i$ and $\underset{\Vdash}{h}_{ij} \neq 0$. These two scenarios correspond to left and middle columns in **Figure 2**, respectively. The relative distance rule is also able to capture the scenario in the right column: when $t^j=0$ (no target signal), $\underset{\Vdash}{h}_i = \underset{\Vdash}{\tilde{h}}_i$ (**Figure 2C**) and $\underset{\Vdash}{h}_{ij} \neq 0$ (**Figure 2F**), the latter provided that $\tilde{g}_{ij} \neq g_{ij}$. Finally, the relative distance rule cannot produce a scenario in which $\underset{\Vdash}{h}_i \neq \underset{\Vdash}{\tilde{h}}_i$ and $\underset{\Vdash}{h}_{ij}=0$, because the latter condition requires that $\tilde{g}_{ij} = g_{ij}$, which in turn enforces $\underset{\Vdash}{h}_i = \underset{\Vdash}{\tilde{h}}_i$. In summary, the relative distance rule is sufficiently expressive to capture all known experimental effects, and insufficiently expressive to capture effects that have not been reported experimentally (it also naturally accounts for the dipper effect, see Section A). In this sense, it is the most parsimonious metric rule that is consistent with empirical observations.

3.4 Computer simulations

To confirm and further study the geometrical framework outlined in **Figure 1**, alongside its connection with the $\tilde{g}_{ij} - g_{ij}$ term in the kernel expressions for the relative distance rule (Section 2.7), we implemented this framework in software using Python. Input stimuli consisted of 13-element vectors specified as detailed in Section 2.2 and **Figure 1C–D**. Each stimulus was linearly projected to a 2-element vector via a pair of front-end filters similar to those in **Figure 1E–F**. The 2-element vector mapped to a 2D surface that could take any pre-defined shape in the embedding 3D volume. To compute geodesic distances on the surface, we incorporated the pygeodesic library into our software pipeline. We then applied the relative distance rule to produce binary responses, and analyzed the resulting dataset using identical procedures to those adopted with human data (Section 2.5).

When this procedure is carried out for a flat surface (**Figure 3A**), the resulting first-order kernels overlap ($\underset{\Vdash}{h}_i = \underset{\Vdash}{\tilde{h}}_i$, **Figure 3D**) and the second-order kernel is featureless ($\underset{\Vdash}{h}_{ij}=0$, **Figure 3G**), resembling the experimental measurements in **Figure 2A**,**D**. We therefore refer to kernel measurements characterized by these features with the term "flat-looking". When the same procedure is carried out for a curved surface (**Figure 3B**), the resulting first-order kernels differ ($\underset{\Vdash}{h}_i \neq \underset{\Vdash}{\tilde{h}}_i$, **Figure 3E**) and the second-order kernel is structured ($\underset{\Vdash}{h}_{ij} \neq 0$, **Figure 3H**), resembling the experimental measurements in **Figure 2B**,**E**. We refer to kernel measurements characterized by these features with the

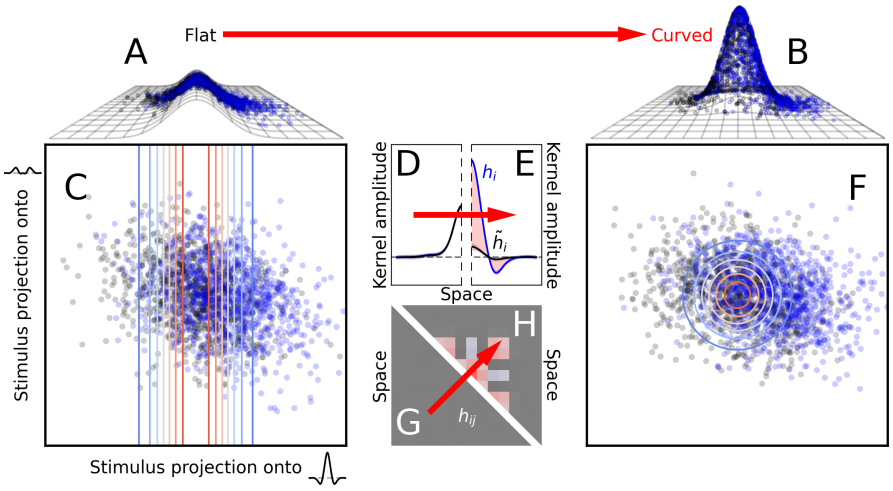

Figure 3: Simulated kernels for the relative distance rule capture the full range of human signatures. Stimuli were structured as in **Figure 1C**. They were projected onto the 2D plane in **C/F** (via linear projection onto the vectors illustrated by the small icons next to the axes in **C**), but distances were measured along geodesics on a potentially curved surface (**A/B**). Blue/black dots indicate target-present/target-absent projections. When the absolute distance rule is applied to distances measured on a flat surface (**A**), the resulting first-order (**D**) and second-order (**G**) kernels are "flat-looking," resembling those obtained at high contrast in human experiments (**Figure 2A**,**D**). When the same procedure is carried out on a curved surface (**B**), the resulting kernels (**E**,**H**) are "curved-looking," resembling those obtained at low contrast in humans (**Figure 2B**,**E**). **D**,**E**,**G**,**H** are plotted to the conventions of **Figure 2**. Red arrows indicate the transition from flat (**A**) to curved (**B**) geometry.

term "curved-looking". When the target signal is removed, first-order kernels become identical (by definition) but the second-order kernel remains structured (not shown), thus capturing the experimental scenario in **Figure 2C**,**F**.

In the above account of the experimental transition between left and middle columns in **Figure 2**, stimulus contrast is not explicitly modelled in any meaningful sense: this parameter is incorporated in the form of curvature, which is specified to be flat (**Figure 3A**) for the purpose of simulating high contrast, and curved (**Figure 3B**) for the purpose of simulating low contrast. In an alternative account of this experimental manipulation, it is possible to capture the resulting effect by specifying a single geometrical landscape with two hills (**Figure 4A**). In this account, contrast is explicitly modelled as a scaling factor applied to the stimulus projection: the low-contrast projection (red dots in **Figure 4B**) spans a smaller range than the high-contrast projection (black dots). Without modifying the geometry, the application of this scaling manipulation produces simulated kernels that adequately capture all salient characteristics of the experimental transition (**Figure 4C–F**).

It may baffle readers that flat-looking kernels (**Figure 4C**,**E**) are produced by a curved geometry (**Figure 4A**). To understand this result, we must consider more closely the meaning of $\tilde{g}_{ij}$ and g_{ij}. In our theorical framework, flat-looking kernels are obtained when $\tilde{g}_{ij}{=}g_{ij}$ (Sections 2.7 and 3.3), where g_{ij} and $\tilde{g}_{ij}$ are ambient metrics. Templates $\mathbf{w}$ and $\tilde{\mathbf{w}}$ point to the summits of the two hills in **Figure 4A**. If we imagine placing the perceptual system at the summit of the hill sampled by low-contrast stimuli (red dots in **Figure 4A**), which corresponds to template $\tilde{\mathbf{w}}$, the system will experience ambient metric $\tilde{g}_{ij}$ when tracing out geodesic paths to the surrounding stimuli. If we then move the system to the summit of the other hill, which corresponds to template $\mathbf{w}$, the system

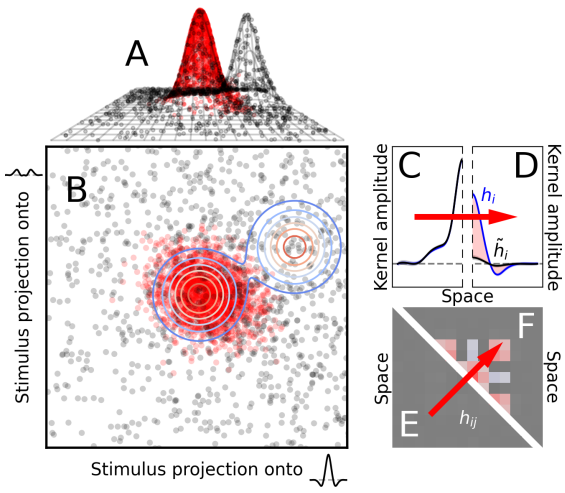

Figure 4: The transition from flat-looking to curved-looking kernels can be obtained without modifying the underlying geometrical landscape. In these simulations, the relative distance rule was applied to a surface with two hills (**A**). Low-contrast stimuli, indicated by red dots (without distinction between target-present and target-absent stimuli), project to the region occupied by only one of the two hills (**B**). The corresponding kernels (**D**,**F**) resemble those obtained in humans at low contrast (**Figure 2B**,**E**). High-contrast stimuli project to a wider region that spans both hills (black dots in **B**), producing kernels (**C**,**E**) that resemble those obtained in humans at high contrast (**Figure 2A**,**D**). While this kernel transition was produced by changing geometry from curved to flat in **Figure 3**, it is produced here without modifying the geometrical landscape, but rather as a consequence of sampling the geometry differently with low- vs high-contrast stimuli. **C**–**F** are plotted to the conventions of **Figure 2**. Red arrows indicate the transition from the high-contrast (black dots in **A**,**B**) to the low-contrast (red dots) projection.

will experience a very different ambient metric g_{ij} when tracing out geodesic paths to the low-contrast stimuli, because these paths will need to cross over to the hill corresponding to the $\tilde{\mathbf{w}}$ template. In other words, the collective projection for low-contrast stimuli looks very different when viewed from the top of the two hills. We can express this configuration as $\tilde{g}_{ij} \neq g_{ij}$, leading to curved-looking kernels (**Figure 4D**,**F**).

When contrast is raised, the collective stimulus projection (black dots in **Figure 4A**) becomes spread across the two hills and samples them to a comparable extent. When travelling from the top of one hill to the stimulus projection along geodesic paths, the system will therefore experience similar ambient metrics regardless of which hill it started from. In other words, the high-contrast projection establishes nearly symmetric sampling of the two hills. We can express this configuration as $\tilde{g}_{ij} = g_{ij}$, leading to flat-looking kernels (**Figure 4C**,**E**). This intuitive interpretation is consistent with our theoretical account (Section 3.3), and is confirmed by the simulations in **Figure 4** (we discuss this and related issues further in Sections 4.1 and 5.1).

The two scenarios simulated in **Figure 3** and **Figure 4** are conceptually different. In the first scenario, the transition from flat-looking to curved-looking kernels is effected by modifying the underlying geometry. In the second scenario, it is effected by modelling contrast explicitly, without modifying the underlying geometry. It is difficult to arbitrate conclusively between these two competing interpretations based on psychophysical evidence, because the manner in which the sensory process is interrogated by standard psychophysical procedures means that the underlying geometry cannot be reconstructed with certainty (see Section 5.1). In essence, when making binary choices of the kind considered here, observers sample only two regions within the perceptual manifold: the region around $\mathbf{w}$, and the region around $\tilde{\mathbf{w}}$. It is possible to attempt extended reconstructions provided some assumptions are incorporated into data analysis (Section 4), however it is not possible to conclusively pinpoint a specific architecture with certainty.

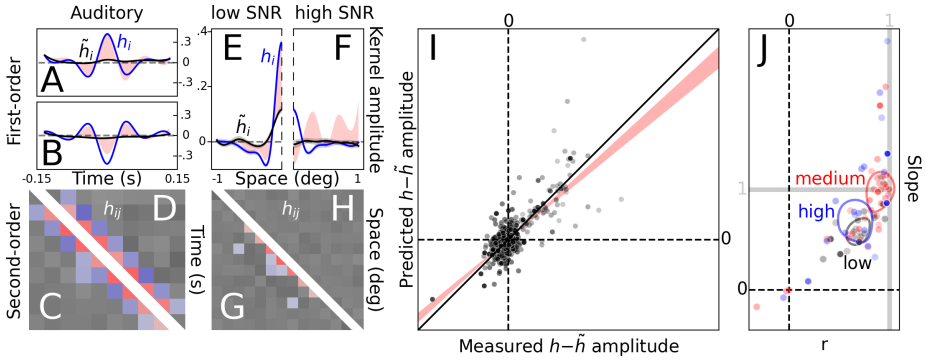

Figure 5: Second-order kernel maps directly to difference between target-present and target-absent first-order kernels with zero free parameters. First-order auditory kernels for detecting loudness increases/decreases (**A/B**) present clear differences between target-present (blue traces) and target-absent kernels (black traces). These differences are fully accounted for (red-tinted regions in **A/B**) by projecting the target signal onto the corresponding second-order kernels (**C/D**). For a stimulus SNR value of 2, which is typical in psychophysical experiments (Murray et al., 2002), the match between measured and predicted difference is equally good in visual experiments (**E**,**G**; see also **Figure 2B**). When SNR is raised to a value of 10, which is well above typical laboratory ranges, the prediction fails (**F**,**H**). **I** plots point-by-point predicted (y axis) versus measured (x axis) differences between target-present and target-absent kernels from a wide range of experiments/conditions, obtained by aggregating ≈613K trials across four published studies (Neri, 2010c, 2013; Joosten et al., 2016; Neri, 2018b). The best linear fit (red-shaded area in **I** shows 95% confidence intervals around fit) is obtained for a slope of 0.86 and intercept∼0, with a correlation coefficient of 0.71. **J** plots slope (y axis) and correlation coefficient (x axis) separately for different kernels (one data point per kernel), sorted into three percentile clusters for low/medium/high SNR (low<2≤medium≤3<high). Oval contours reflect spread of different clusters (for visualization only). Optimal slope values of 1 (indicated by horizontal gray line) and correlation coefficient values approaching 1 (indicated by vertical gray line) are achieved for medium SNR levels (red cluster). Symbol opacity reflects square root of number of trials that contributed to its computation in **I**,**J** (low opacity for fewer trials). Kernels (**A**–**H**) are plotted to the conventions of **Figure 2**. Panels **A**–**D** were computed from 91K trials of published data (Joosten et al., 2016) aggregated across 10 listeners. Panels **E**–**H** were computed from 106K trials of unpublished data aggregated across 9 observers.

3.5 Non-parametric prediction of differential first-order kernel modulations from second-order kernel structure

Equations 19–21 show that our theory predicts a simple relationship between the two empirical characteristics that informed previous sections (difference between target-present and target-absent first-order kernels, structured modulations within second-order kernel). More specifically:

$$2(\underset{\Vdash}{h}_i - \underset{\Vdash}{\tilde{h}}_i) = \underset{\Vdash}{h}_{ij} t^j \tag{22}$$

This relationship is not unique to the relative distance rule (Neri, 2010b) (an identical relationship applies to the absolute distance rule, see equations 16–18).

We know every term in expression 22: $\underset{\Vdash}{h}_i$, $\underset{\Vdash}{\tilde{h}}_i$, and $\underset{\Vdash}{h}_{ij}$ are estimated directly from data, while t^j is specified by the experimenter. There are no free parameters involved.

This relationship holds very tightly in simulated scenarios: the shaded area (red tint) in **Figure 3E** shows the predicted difference between $\underset{\Vdash}{h}_i$ and $\underset{\Vdash}{\tilde{h}}_i$ obtained by applying expression 22 to the second-order kernel in **Figure 3H**. A similar result is demonstrated in **Figure 4D**.

When applied to real data, in general the prediction holds equally well: the red-tinted region in **Figure 2B** shows an example for the low-contrast experiments discussed in Section 3.2, while **Figure 5A**,**B** show two examples from auditory experiments involving detection of loudness increments and decrements, respectively. **Figure 5E**,**F** show examples from unpublished low-contrast visual experiments similar to those in **Figure 2B**. Provided stimulus SNR is kept within a reasonable range (SNR=2 in **Figure 5E**), the prediction is accurate (red-tinted area matches difference between blue and black traces). When SNR is increased to non-physiological levels (SNR=10 in **Figure 5F**), the prediction fails completely (red-tinted area does not match difference between blue and black traces).

The above failure is expected, because our derivations assume low-to-medium-range SNR values (Section 2.6). Notice that SNR could be increased to uncharacteristically high values in the experiments of **Figure 5F** because they involved a very peculiar regime in which contrast was unusually low, remaining below detection threshold. In more typical conditions when stimulus visibility is not a limiting factor, SNR values as high as the one demonstrated in **Figure 5F** (SNR=10) would not permit kernel estimation, because the human observer would be nearly always correct on every trial. On the few trials on which the observer may respond incorrectly, this behavior would likely be attributable to stimulus-decoupled factors, such as blinking or inattention, that are not correlated with input noise. In this specific sense that pertains to viable threshold experiments, SNR values as high as 10 are not physiological (see also (Murray et al., 2002)).

To gauge the applicability of expression 22 more quantitatively and across a wider range of experimental settings, we applied this prediction to $>$600K trials from various conditions/observers/experiments spanning four published studies (Neri, 2010c, 2013; Joosten et al., 2016; Neri, 2018b). **Figure 5I** plots predicted (y axis) versus measured (x axis) amplitude differences between target-present and target-absent first-order kernels. Points scatter around and along the unity line, indicating that the prediction generalizes adequately to a broad range of experimental configurations.

To further study the impact of SNR, we sorted this large dataset into three equally-sized percentile groups corresponding to different SNR levels (low, medium, high), and measured the slope/correlation-coefficient associated with each kernel in each group when plotted separately in the manner of **Figure 5I**. At low and high SNR (black and blue data points in **Figure 5J**), both slope and correlation coefficient fall below the optimal target value of 1 (black/blue ovals fall below the horizontal gray line and to the left of the vertical gray line in **Figure 5J**). At medium SNR (red data points in **Figure 5J**), these two characteristics are very close to 1 (red oval falls near the point of intersection between horizontal and vertical gray lines in **Figure 5J**). This result is consistent with our theory (see below).

As already noted above, our theory does not apply at high SNR. More specifically, equation 22 relies on a linear approximation of Ψ (Section 2.6). This approximation is only reasonable for SNR values below $\approx$3 (Neri, 2010b). At higher SNR values, it becomes necessary to expand the approximation for Ψ to higher orders, however the resulting expressions involve ratios between expansion coefficients (Neri, 2010b) (e.g. between second-order and first-order), and in the best-case scenario these are not easily estimated from empirical measurements (in the worst-case scenario, depending on the

nature of the experimental design, they cannot be estimated at all). Our goal is to remain with zero-free-parameter predictions such as equation 22 (see Section 3.1), so we do not attempt higher-order expansions here.

At low SNR, the amplitude of the target signal is small, weakening the distinction between target-present and target-absent kernel estimates. This configuration inevitably leads to highly unreliable measurements of their difference (if present). In the extreme case of SNR=0 (as in **Figure 2C**), it is meaningless to measure slope/correlation for the prediction from expression 22, because these two characteristics are bound to scatter around 0 (the prediction becomes equal to 0 across the board).

It may be argued that the above results merely reflect differences in data mass allocation to different SNR groups: it is possible that, by chance or by unwitting design, more trials were collected for experiments at medium SNR. Under this scenario, the medium-SNR group would be trivially expected to show higher correlation coefficients as a consequence of more reliable measurements. We can exclude this interpretation on the basis of two observations. First, the number of trials allocated to each kernel estimate is slightly larger for the low/high SNR groups (Q1–median–Q4: 5.1K–5.8K–7K/5.2K–5.8K–6.9K) compared with the medium SNR group (5.1K–5.6K–6.5K). Second, there is no obvious relationship between symbol opacity (which scales with square root of number of trials allocated to the corresponding data point) and vicinity to target slope/correlation-coefficient values across data points in **Figure 5J**. We therefore conclude that the trend exposed by the scatter plots in **Figure 5I**,**J** does not reflect artefactual differences in data mass allocation across conditions.

4 Reconstruction of underlying geometry directly from data

If we adopt some simplifying assumptions, we can reconstruct and visualize the manner in which the metric varies across the perceptual manifold directly from data, without introducing any free parameters. For this purpose, we therefore allow $g_{ij}(\mathbf{s})$ and $\tilde{g}_{ij}(\mathbf{s})$ to vary with $\mathbf{s}$.

4.1 Reconstruction in theory

Our first assumption is that the metric is rank 2, and that the eigenvectors are fixed for a given set of experimental conditions. For example, for the experiments involving contrast manipulations (left/middle columns in **Figure 2**), we assume that the same two eigenvectors apply to all stimuli across all contrast levels. To avoid excessive subscripting/superscripting in writing out explicit expressions for the projection of $\mathbf{s}$, $\tilde{\mathbf{s}}$, $\mathbf{w}$, and $\tilde{\mathbf{w}}$ onto the two eigenvectors, we adopt the following conventions: the two scalar values obtained by projecting stimulus $\mathbf{s}$ onto the two eigenvectors are s_1 and s_2; the two scalar values obtained by projecting template $\tilde{\mathbf{w}}$ onto the two eigenvectors are w_1 and w_2; the two eigenvalues associated with g_{ij} are λ_1 and λ_2; similar notation applies to $\tilde{\mathbf{s}}$, $\tilde{\mathbf{w}}$, and $\tilde{g}_{ij}$ with the addition of the $\tilde{}$ specifier.

Second, we assume that the perceptual process operates in accordance with the relative distance rule. Adopting the above assumption/notation, we can rewrite the output

in response to stimulus $\mathbf{s}$ using the following simple expression:

$$\underset{\Vdash}{R}(\mathbf{s})=\tilde{\lambda}_1(\mathbf{s})(s_1-\tilde{w}_1)^2+\tilde{\lambda}_2(\mathbf{s})(s_2-\tilde{w}_2)^2-\lambda(\mathbf{s})_1(s_1-w_1)^2-\lambda(\mathbf{s})_2(s_2-w_2)^2 \quad (23)$$

where the dependence on $\mathbf{s}$ has been displaced from $g_{ij}/\tilde{g}_{ij}$ onto eigenvalues $\lambda_1/\tilde{\lambda}_1$ and $\lambda_2/\tilde{\lambda}_2$, which essentially amounts to dimensionality reduction.

We express the Z-scored probability of producing a correct response as $\Phi^{-1}[\Psi(\Delta\underset{\Vdash}{R})]$ (Section 2.4) where Φ is the cumulative normal distribution function. We assume (as is customary) that Ψ is the cumulative Gaussian distribution function with standard deviation (SD) η, the SD of the internal noise source expressed as a multiple of the SD of the fluctuations induced onto the decision variable (R) by the external noise source (Green and Swets, 1966). We now approximate this function to first-order. For moderate SNR values, this approximation will be equivalent to a first-order approximation around 0 (which corresponds to SNR=0). The Z-scored probability will therefore be:

$$Z \propto \frac{\Delta\underset{\Vdash}{R}}{\eta} \quad (24)$$

The proportionality constant above can be safely assumed to remain fixed across experimental manipulations that do not involve large changes in SNR (Neri, 2010b).

We now choose a specific target-present stimulus $\mathbf{s}_\star$, and compute the Z-scored probability that it is classified as target-present when paired with all possible $\tilde{\mathbf{s}}$:

$$Z(\mathbf{s}_\star)=\frac{1}{\eta}\langle\Delta\underset{\Vdash}{R}(\mathbf{s}_\star,\tilde{\mathbf{s}})\rangle_{\tilde{\mathbf{s}}}=\frac{1}{\eta}[\underset{\Vdash}{R}(\mathbf{s}_\star)-\langle\underset{\Vdash}{R}(\tilde{\mathbf{s}})\rangle]$$

where the term $\langle\underset{\Vdash}{R}(\tilde{\mathbf{s}})\rangle$ is clearly independent of $\mathbf{s}_\star$.

Next, we choose a specific target-absent stimulus $\tilde{\mathbf{s}}_\star$ that projects to the same scalar values s_1/s_2 returned by the the projection of $\mathbf{s}_\star$. Similarly to the procedure used with $\mathbf{s}_\star$, we consider the probability that $\tilde{\mathbf{s}}_\star$ is classified as target-present when paired with all possible $\mathbf{s}$: $1-\Psi[\langle\Delta\underset{\Vdash}{R}(\mathbf{s},\tilde{\mathbf{s}}_\star)\rangle_{\mathbf{s}}]$. When Z-scored, this simplifies to:

$$Z(\tilde{\mathbf{s}}_\star)\propto\frac{1}{\eta}[\underset{\Vdash}{R}(\tilde{\mathbf{s}}_\star)-\langle\underset{\Vdash}{R}(\mathbf{s})\rangle]$$

where the term $\langle\underset{\Vdash}{R}(\mathbf{s})\rangle$ is clearly independent of $\tilde{\mathbf{s}}_\star$.

We now average the two Z-scored probabilities above:

$$\eta Z(\mathbf{s}_\star,\tilde{\mathbf{s}}_\star)\propto\underset{\Vdash}{R}(\mathbf{s}_\star)+\underset{\Vdash}{R}(\tilde{\mathbf{s}}_\star)+\text{terms not dependent on }\mathbf{s}_\star,\tilde{\mathbf{s}}_\star$$

When written out explicitly (via equation 23), we obtain the following polynomial in s_1/s_2:

$$\begin{aligned}\eta Z(\mathbf{s}_\star,\tilde{\mathbf{s}}_\star)\propto\sum_{i=1}^{2}&[\tilde{\lambda}_i(\mathbf{s}_\star)+\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star)-\lambda_i(\mathbf{s}_\star)-\lambda_i(\tilde{\mathbf{s}}_\star)]s_i^2\\&+2[\lambda_i(\mathbf{s}_\star)w_i+\lambda_i(\tilde{\mathbf{s}}_\star)w_i-\tilde{\lambda}_i(\mathbf{s}_\star)\tilde{w}_i-\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star)\tilde{w}_i]s_i\\&+[\tilde{\lambda}_i(\mathbf{s}_\star)+\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star)]\tilde{w}_i^2-[\lambda_i(\mathbf{s}_\star)+\lambda_i(\tilde{\mathbf{s}}_\star)]w_i^2\end{aligned} \quad (25)$$

We want to study how the quadratic surface specified by Z varies as a function of small perturbations around s_1/s_2, to gain some insight into the manner in which $\lambda_1/\tilde{\lambda}_1$ and

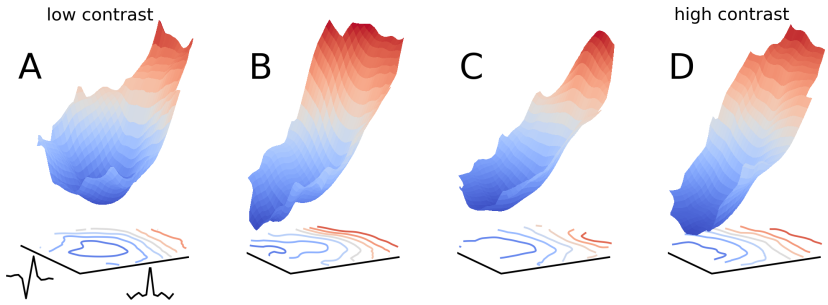

Figure 6: Reconstructed geometry for different contrast levels. Surface in **A** plots the probability of classifying a given stimulus as target-present (z axis, corresponding to quantity Z in equation 25) as a function of the two values (mapped to x and y axes) associated with that stimulus when projected onto the two eigenvectors shown by black traces near x/y axes (these eigenvectors were computed directly from the second-order kernel derived from the entire dataset across all contrast levels, see Section 4.2). In practice, projected values were binned for an arbitrarily chosen bin partition, and all stimuli/trials within a given bin contributed to the estimation of the corresponding probability (see main text for details). Consistent with corresponding kernel-based observations (left/middle columns in **Figure 2**), the reconstructed geometry at low contrast is curved (**A**), and becomes progressively planar/flat as contrast is increased (**B** to **D**). Surfaces were computed from 28K trials aggregated across 8 observers at SNR=1 (Neri, 2018b).

$\lambda_2/\tilde{\lambda}_2$ vary as a function of $\mathbf{s}_\star$ and $\tilde{\mathbf{s}}_\star$. We make the reasonable assumption that the various λ factors vary more slowly with $\mathbf{s}_\star/\tilde{\mathbf{s}}_\star$ than the projected values associated with those stimuli (s_1/s_2). Failing this assumption, it becomes impossible to resolve/visualize λ variations, because they would exceed the resolution of our visualization strategy.

First, we examine the case in which the metric is constant across perceptual space (flat geometry): the λ factors do not depende on $\mathbf{s}_\star$ and $\tilde{\mathbf{s}}_\star$, and $\lambda_i{=}\tilde{\lambda}_i$. In this case, Z can be visualized as planar (equation 25 becomes linear in s_1 and s_2). For the degenerate case $w_i{=}\tilde{w}_i$, in which the system carries no discriminatory power (the two memory templates map to the same vector), $Z{=}0$ (chance performance).

Second, we examine the case in which the metric changes primarily with respect to the memory template from which distance is computed ($\lambda_i{\neq}\tilde{\lambda}_i$), and less with respect to the specific choice of $\mathbf{s}_\star/\tilde{\mathbf{s}}_\star$ to which distance is computed ($\lambda_i(\mathbf{s}_\star){\approx}\lambda_i(\tilde{\mathbf{s}}_\star)$ and $\tilde{\lambda}_i(\mathbf{s}_\star){\approx}\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star)$). In this case, Z can be visualized as curved (equation 25 becomes quadratic in s_1 and s_2).

There is a third scenario in which the metric changes with respect to both memory templates and input stimuli. In general, Z will be visualized as curved under these conditions. However, it is possible for it to appear flat if $\mathbf{s}_\star$ and $\tilde{\mathbf{s}}_\star$ are chosen so that $\tilde{\lambda}_i(\mathbf{s}_\star){+}\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star){=}\lambda_i(\mathbf{s}_\star){+}\lambda_i(\tilde{\mathbf{s}}_\star)$. This is unlikely, but not impossible. Consider for example two zero-mean eigenvectors u_i (even) and v_i (odd). Choose $\mathbf{s}_\star$ to be an indicator function for positive values of u_i (equal to 1 whenever u_i is positive, 0 otherwise), and $\tilde{\mathbf{s}}_\star$ to be the negative of an indicator function for negative values of u_i (equal to -1 whenever u_i is negative, 0 otherwise). $\mathbf{s}_\star$ and $\tilde{\mathbf{s}}_\star$ return the same values when projected onto u_i and v_i, however they are sufficiently different from one another to allow for the reasonable possibility that $\lambda_i(\mathbf{s}_\star){\neq}\lambda_i(\tilde{\mathbf{s}}_\star)$. As long as $\tilde{\lambda}_i(\mathbf{s}_\star)$ and $\tilde{\lambda}_i(\tilde{\mathbf{s}}_\star)$ vary in mirror fashion to $\lambda_i(\mathbf{s}_\star)$ and $\lambda_i(\tilde{\mathbf{s}}_\star)$ to keep their respective sums balanced, Z will look flat in the vicinity of the projected values s_1 and s_2. This scenario is not dissimilar from the scenario simulated in **Figure 4**, where a specific choice of stimulus distribution (high-contrast, black dots in **Figure 4B**) returns flat-looking kernels (**Figure 4C**,**E**) for

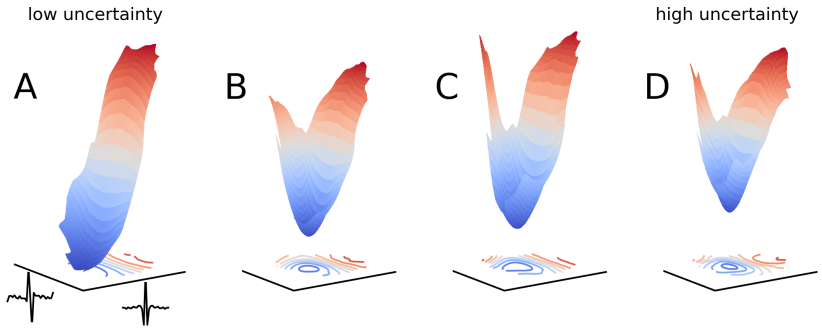

Figure 7: Reconstructed geometry for different uncertainty levels. Plotted to the same conventions of **Figure 6** using a different published dataset. When observers are allowed to focus on a restricted region of space where the target may appear (low uncertainty), the reconstructed geometry is flat (**A**). When the target is allowed to appear within increasingly wider regions of space (higher uncertainty), the underlying geometry becomes progressively more curved (**B** to **D**). Surfaces were computed from 94K trials aggregated across 10 observers (Neri, 2010c).

a curved geometry (**Figure 4A**).

4.2 Reconstruction in practice

In practice, the reconstruction method must be subjected to an additional form of aggregation. Although in theory we can "choose $\tilde{\mathbf{s}}_\star$ such that its projected values match those for $\mathbf{s}_\star$," this is not possible in practice: we cannot estimate the eigenvectors before running the experiment. In similar fashion, although in theory we can state "when paired with all possible $\tilde{\mathbf{s}}$," we cannot repeat infinite measurements for a specific stimulus $\mathbf{s}_\star$ when paired with all possible $\tilde{\mathbf{s}}$, let alone do this for many choices of $\mathbf{s}_\star$. When applying the reconstruction method to real data, we must bin the output range of s_1/s_2 and compute Z for all trials that fall within each bin. This practical implementation of the methodology further reduces its resolving power.

To estimate the eigenvectors associated with the metric, we rely on the assumption that they remain fixed throughout (for a given dataset). If we accept that the second-order kernel h_{ij} can be expressed as $\tilde{g}_{ij}-g_{ij}$, its eigenvectors will be those same eigenvectors (except for the comprehensively flat case in which $\tilde{g}_{ij}=g_{ij}$ across the entire dataset). We can therefore estimate the two eigenvectors from the experimentally derived h_{ij} computed across the entire dataset (examples are shown by the two black traces near the axes at the bottom of **Figure 6A**). We can then project all stimuli used for a given experimental condition (e.g. low contrast) onto those eigenvectors, and compute Z.

In equation 25, the lefthand side contains ηZ, not just Z: to estimate the internal geometry of the response landscape (righthand side) prior to the application of the choice transducer Ψ, we must take into account internal decisional noise η. This quantity can be estimated experimentally using the double-pass technique (Burgess and Colborne, 1988; Neri, 2010a), which is the approach we adopt here (please refer to the original publications detailing internal noise characteristics for the datasets analyzed in this section, namely (Neri, 2018b) and (Neri, 2010c)). In the following, we therefore plot ηZ.

Figure 6 shows the outcome of applying the above procedure to the dataset already examined in Section 3.2, involving the transition from curved-looking to flat-looking

kernels with increasing contrast (left/middle columns in **Figure 2**). In line with those observations, the reconstruction method returns a curved surface at low contrast (**Figure 6A**) that becomes progressively planar as contrast is raised (**Figure 6B–D**). Notice that, even when the geometry is predominantly curved (see blue-tinted region in **Figure 6A**), some portions of perceptual space may nevertheless be flat (red-tinted region).

In **Figure 7**, the reconstruction method is applied to a different dataset involving different levels of spatial uncertainty (Neri, 2010c). In this setting, kernel derivation is complicated by the fact that the target signal was not uniquely specified, in that it was allowed to vary its spatial location within a predetermined uncertainty window. In this case, kernel analysis can still be carried out using a technique termed signal-clamping (Tjan and Nandy, 2006), whereby noise samples are spatially registered to align with the target signal (Neri, 2010c). This ralignment procedure can only be applied to target-present noise samples, however, and not to target-absent samples, which may impact the applicability of the reconstruction process. Notwithstanding this limitation, it is interesting to observe that the reconstructed geometry varies from flat to curved with increasing uncertainty (**Figure 7A** to **Figure 7D**).

5 Discussion

5.1 Mathematics and perception

Perception in its most general sense (such as qualia) is far too complex a phenomenon to be available for mathematical treatment. It would already be an extraordinary achievement to establish a mathematical theory of perceptual processing in the sense of experimental psychology, that is in the sense of a phenomenon that can be quantified through behavior. If we restrict ourselves to the latter sense, we still find that not much progress has been made since the seminal work in the 1960s that led to the development of signal detection theory (Green and Swets, 1966) (SDT). A few other mathematical operators have been established for neuronal processing, such as energy models (Adelson and Bergen, 1991) and gain control (Carandini and Heeger, 2011), but their applicability to perceptual processing has been difficult to ascertain conclusively.

It seems relevant to ask: why has perception proven so difficult to capture using mathematical tools? There are various reasons potentially underlying this difficulty, ranging from the high variability associated with psychological phenomena to the inherent complexity and richness of perceptual experiences. At the same time, it should be recognized that these challenges have not hampered the mathematical characterization of neural activity to the same extent, despite those challenges being arguably equally relevant to neural representations: starting in the 1950s with the Hodgkin-Huxley model and subsequent developments in single-neuron biophysics, and progressing onto manifold accounts of multiunit activity in recent years (Kriegeskorte and Wei, 2021; Chung and Abbott, 2021; Langdon et al., 2023), it seems evident that neuromathematics (Ermentrout and Terman, 2010; Petitot, 2017) has provided compelling accounts of neurophysiological signals to a degree of detail and depth that has not been achieved for psychological and perceptual phenomena. The question is why.

We believe a large part of the answer lies in the extreme degree of information

compression that is effected by animals when taking behavioral decisions (Zheng and Meister, 2024; Wu et al., 2016) ($\mathcal{H}$ and Ψ in our formalism): animals receive a staggering amount of sensory data during natural behavior; this information is digested by their neural machinery to generate relatively rich perceptual representations; these representations are then used by animals to produce behavioral decisions in the form of motor outputs that ensure their survival, often involving little more than binary choices between discrete behaviors. The reduction in information rate associated with each of these transitions is so extreme that it can hardly be put into perspective. To provide a rough idea of the compression incurred by the translation from neural representation to behavioral output (second transition above), this can be estimated to involve a contraction from >1 gigabits/s to 10 bits/s (Zheng and Meister, 2024), which is at least a 10^8-fold reduction in information rate.

It follows from the above considerations that behavior provides an extremely impoverished readout of perceptual processes, thus making it difficult to impose serious constraints on candidate mathematical models (even low-dimensional ones) aimed at capturing behavior. For example, a sensitivity-based characteristic like the dipper function (Appendix A, **Figure 9B**) can be explained using a variety of different modeling schemes (Solomon, 2009), without clear evidence in favour of one or the other: the behavioral signature is just too poor a constraint to support effective model selection. In this study, we rely primarily on higher-dimensional behavioral descriptors (perceptual kernels, Section 2.5). These descriptors arguably provide a richer characterization of sensory behavior than sensitivity-based measurements (or at least complementary to the latter (Neri, 2018a)), however they too present important limitations: because of the compression associated with binary behavioral outputs, it becomes necessary to acquire many instances of such outputs to achieve a sufficient depth of empirical characterization (the kernels presented in this study come from dozens of thousands of trials).

The above considerations allow us to appreciate the non-trivial nature of the result in **Figure 5I-J**. This is a zero-free-parameter prediction of rich behavioral descriptors based on an articulated mathematical framework. It is remarkable that human behavior can be modelled to such an extent of precision using mathematics, and is in no way to be taken for granted: we have presented examples for which the theoretical prediction does fail under specific conditions (**Figures 5F**). Although this failure is to some extent expected from (and therefore indirectly consistent with) theory, it also demonstrates that the match between empirical descriptors and theoretical prediction is not a given: it is rooted in genuine compliance of human behavior with our theoretical treatment, which is ultimately based on mathematical expressions. If it is reasonable to find it unreasonable that mathematics is so profoundly effective at capturing natural phenomena in the physical sciences (Wigner, 1960), it seems reasonable to find it astounding that mathematics can capture so accurately a phenomenon as complex as human sensory behavior, notwithstanding the limited domain for which this result is demonstrated in the present study.

5.2 Relations to prior literature

It is legitimate to wonder whether/how the proposed framework may differ (or not) from existing accounts of sensory processing. More specifically, it may be argued that the

geometrical account proposed in this study amounts to nothing more than rephrasing nonlinear frameworks for system characterization/identification (Golden et al., 2016; Marmarelis, 2004), such as the Volterra expansion. These frameworks have a long history in systems neuroscience (Marmarelis and Marmarelis, 1978; Westwick and Kearney, 2003), indeed the derivation of first-order/second-order kernels via noise-based reverse correlation techniques refers back to Volterra-like formulations (Section 2.6), both in physiology (Schetzen, 1980) and in psychophysics (Heath, 2000; Neri, 2004, 2010b).

Our approach in this study is fundamentally different for the following reasons. It was motivated by the specific goal of recasting perceptual processing using the simplest imaginable notion of what it means to take sensory measurements of the outside world: the notion of metric distance in perceptual space (Section 2.3). This is arguably the most natural choice for approaching the problem, to the extent that it may strike some readers as overly naive.

Once we commit to a unified account based on metric distance, the appropriate framework is intrinsic geometry. The Volterra expansion does not, *per se*, say anything about a specific notion of perceptual distance: it is a non-parametric general-purpose approximator for a wide class of functionals without any explicit connection with any specific construct, whether geometrical or otherwise. For example, the response of a system (R in our notation) may be successfully characterized by a Volterra expansion of the input stimulus ($\mathbf{s}$) via the introduction of specific kernels (Marmarelis, 2004), but this does not say anything about the potential connection between the structure of those kernels and the perceptual operation that led to said structure.

To understand the above statement more concretely in direct connection with the material presented here, consider the transition from flat-looking to curved-looking kernels (left column to middle column in **Figure 2**). Both can be modelled using Volterra expansions, but this formulation does not say anything about whether their different characteristics originate from the same perceptual operation or from different operations, nor does it provide clearly interpretable information about what those operations may be. In our framework, the two characteristics originate from one and the same operation: measuring distances in perceptual space. The transition reflects a change in the intrinsic shape of perceptual space, but said space is subjected to the same operation for the purpose of generating a perceptual judgment.

For an additional example of the difference between our framework and the Volterra formalism, consider the manner in which we proceed when approximating the metric to low-rank in Section 4.1. Using this approximation/assumption, we can derive the eigenvectors of g directly from h_{ij}. This approach is predicated on the notion that h_{ij} can be expressed as the sum of different g's, however. Without an explicit connection between h_{ij} and g_{ij} (equation 22), there is nothing to tell us that we can apply the above procedure for inferring the eigenvectors of g_{ij} from h_{ij}. This explicit connection is instantiated by our geometrical framework. In the Volterra formalism, the conceptual significance of the eigenvectors of h_{ij} remains unclear/unspecified.

More generally, our approach defines a simple geometrical framework that can potentially incorporate various cognitive phenomena (such as attention) in the form of intuitively graspable constructs, such as positional shifts of template vectors or disturbances in the distance-measuring process. Under a general-purpose approximator such

as the Volterra expansion, those phenomena would merely involve compiling a list of induced kernel alterations with no unifying conceptual structure. At this stage, we do not know whether the potential of our framework for encompassing and understanding a wider range of cognitive phenomena may be realized or not, but we hope to address this issue in future work.

In a line of research that shares some conceptual characteristics with ours, some authors have proposed hyperbolic accounts of various perceptual phemonena, ranging from color vision (Yilmaz, 1962; Farup, 2014) to stereopsis (Luneburg, 1950; Indow, 1967). In these proposals, the perceptual metric is hyperbolic and specific cognitive phenomena, such as adaptation in color vision (Yilmaz, 1962) and movement speed in motion perception (Caelli et al., 1978), are associated with Lorentzian transformations. These accounts differ from ours in that, similar to special relativity, they live in a flat geometry: the associated second-order kernel is featureless, because the components of the metric are fixed. In our account, the observation of curved-looking kernels necessitates that said components vary across perceptual space. Furthermore, experimental evidence in support of those hyperbolic accounts is either not available or rather limited in its constraining power, at least in comparison with the depth of empirical scrutiny afforded by the methodologies employed in this study.

We conclude this section by drawing a clear distinction between the approach advocated here and existing mathematical treatments of various problems in vision science that rely on tools from differential geometry (Koenderink, 1984; Hoffman, 1989; Petitot, 2017). Those tools have played a central role in seminal work on the instrinsic geometrical structure of visual space (Koenderink, 1990). The frameworks developed by these prior studies "live" in the physical world, so to speak, in the sense of its ecological architecture (Gibson, 1979) and/or the optical image it projects (Koenderink, 1984; Turski, 2023), and as such they do not speak directly to the geometry of perceptual operations. Conversely, our framework lives in perceptual space: all relevant architectural constraints are imposed by the perceptual operation of computing geodesic distance, with little (at most indirect) input from the intrinsic geometrical structure of the physical world. Our account is also entirely local: the geometry is expanded around a single point in the optical image, with nothing to say about extended physical space and related phenomena (such as Gestalt rules of good continuation (Sarti et al., 2009)). In the future, it may be possible to specify the perceptual equivalent of field equations to describe how metric structure varies across space as a function of certain image characteristics (such as object boundaries or other salient features in the environment), however this seems a distant goal at the present moment. In conclusion, our framework is neither superior nor preferable to those discussed in this paragraph: it simply speaks to a different problem.

5.3 Curvature: intrinsic property, or curvilinear chart?

An important issue that was not explicitly addressed in this article relates to the coordinates adopted for specifying perceptual charts. Throughout this study, we have assumed Gaussian coordinates: those defined by orthogonal geodesics at a point. This assumption is implicitly incorporated into our simulations (Section 3.4) by measuring geodesic distance. Using this coordinate system, changes in the metric tensor transparently re-

flect genuine changes in curvature of the underlying geometry. What is the impact on our conclusions of allowing charts to be defined by curvilinear coordinates?

In a sense, the above issue is a non-issue, because it is not possible to distinguish between changes in curvature and changes in chart coordinates: we have no access to an underlying "true" geometry. In physics applications such as general relativity, for example, the notion of genuine curvature refers back to the physical world (matter distribution in particular). In psychological/perceptual space, there is no clear way of defining what is "real" and what is not: what is the difference between saying that the metric tensor defined in Gaussian coordinates varies across perceptual space because perceptual space is curved, and that the metric tensor defined in curvilinear coordinates varies across perceptual space despite said space being flat? In both cases, the perceptual system defines chart coordinates as well as underlying space, so from the viewpoint of metric measurements it is not possible to arbitrate between those two interpretations.

Notwithstanding the above considerations, it is possible to hypothesize a specific type of neural architecture for which the distinction between coordinate-induced and curvature-induced changes of the metric tensor takes on provisional meaning. If we posit the existence of an independent neural geometry at the level of brain activity, for example as represented by the spiking output of a neuronal subpopulation within a relevant sensory area (Sohn et al., 2019; Vyas et al., 2020; Langdon et al., 2023), and that this geometry exists as a distinct phenomenon from the perceptual process that instantiates charts on it for the purpose of measuring distance and generating perceptual judgments, then we can ascribe curvature-induced changes to the former representation and coordinate-induced changes to the latter process. In other words, we are hypothesizing the existence of a "neural reality" that is independent of the perceptual process, and we ascribe genuine curvature (or lack thereof) to this "real" space. The perceptual process then becomes an observer of said reality: this observer may produce measurements that appear curved, even though the underlying neural reality is flat, merely as a consequence of laying out charts with curvilinear coordinates. We discuss this hypothesis further below.

5.4 The distinction between sensing, perceiving, and performing perceptual judgments

Our theory involves several intermediate constructs and computations preceding the final motor output produced by the system (Section 2). A relevant question is to what extent these "hidden" processes occur independently of whether they result in an overt motor output or not. For example, we may ask whether the geometrical landscape is instantiated by the brain of an anesthetized/asleep/ unconscious individual (Chaudhuri et al., 2019; Gardner et al., 2022): if visual stimuli are flashed in front of his/her eyes, is the associated brain activity organized around a geometrical structure that retains the same characteristics elicited in the conscious state? Similarly, we may ask whether a specific chart is explicitly selected and represented by the sensory process of an individual who is not required to make perceptual judgments: is the mere act of perceiving the visual stimulus accompanied by the instantiation of metric components dictated by a spontaneously selected chart? Further, we may ask whether memory templates for

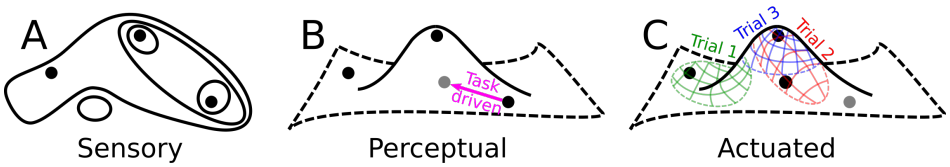

Figure 8: Speculative neurogeometrical scheme for the unfolding of perceptual processing from sensory encoding to motor action. We hypothesize that, in the absence of conscious perception, sensory representations are endowed with topological structure dictated by neural substrates (**A**). Perception enriches the topology with manifold architecture and a specific geometrical structure (**B**), which is encoded by stable descriptors in neural population activity. During task training, memory templates are recalibrated to represent task-relevant signals (magenta arrow). The geometry is not rendered explicit until the perceptual system is queried for the purpose of expressing a perceptual judgment, when it then selects a specific chart and performs actionable measurements on the underlying geometry (**C**). Every time the system is queried, a different chart may be selected (red/blue/green elements in **C**), which may appear in the form of trial-to-trial variations in neural activity. This scheme would account (among other things) for the co-existence of stable and heterogeneous characteristics at the level of neural activity (Low et al., 2018; Langdon and Engel, 2022).

specific stimuli are created spontaneously upon presentation of noiseless exemplars, whether they require explicit instructions/directions (e.g. "this is what your target looks like"), and/or whether they are only represented at the moment of making a perceptual judgment as to whether a given stimulus conforms to the template or not.

Needless to say, we currently do not have answers to these questions, so we can only speculate. We put forward a bold hypothesis in which we distinguish three components/stages: brain activity devoid of manifold structure (no atlas, **Figure 8A**); impromptu representation of manifold structure with Riemannian tools and addition/refinement of task-specific templates (**Figure 8B**); temporary choice and deployment of a specific chart (**Figure 8C**). We associate these three stages with three distinct states in which the human observer may be operating.

We hypothesize that, in the absence of conscious (i.e. potentially actionable) perception, there is no manifold structure. This does not mean that there is no structure at all: neural activity may be organized according to an identifiable topology (**Figure 8A**) with relatable notions, such as conceptually "meaningful" subsets (e.g. the subset of all subparts belonging to a given object, the subset of all objects belonging to a given category, and so on). When we speak of manifold structure, we specifically refer to the representation of geometrical constructs that can be subjected to actionable computations. In the above, we use the term actionable with reference to a state/representation that can support action in the behavioral sense of motor output.

In the presence of conscious perception, geometrical structures are fully instantiated (**Figure 8B**), but they are not represented in computable coordinates with reference to a specific chart. This may be the state of an observer who is "looking at the world" without being assigned a specific behavioral task. Of course agents are constantly performing behavioral tasks, often self-initiated, but we hypothesize that the observer may be just "perceiving without acting" at least with reference to certain aspects of the sensory world. For example, the observer may be listening to a sound stream and performing actions in response to auditory information, however he/she may also be concomitantly sensing visual stimuli in the environment, without necessarily performing actions in response to visual information. We hypothesize that, under these conditions, neural ac-

tivity is encoding the various geometrical constructs we have used in this study (such as the metric tensor) using representational schemes that are chart-independent.

Perception is inextricably informed by templates (sometimes termed priors (Mamassian et al., 2002)), some genetically encoded, others learnt throughout development. We hypothesize that these templates are spontaneously represented during conscious perception, however most of them will not be relevant to the specific behavioral task faced by the animal at a given time, so no distance measurements are computed for these templates. When observers are introduced to a new task, they will create new templates corresponding to the signals that are relevant to the task, possibly by modifying pre-existing templates (Hafting et al., 2005; Gardner et al., 2022) (magenta arrow in **Figure 8B**). For example, if observers are asked to detect a bright bar against a gray background, such as in the experiments considered in Section 3.2, observers will recall pre-existing templates for objects that are similar to a bright bar, clone one of them, and fine-tune its characteristics to align with the newly introduced target.

Finally, when observers are either prompted or forced to produce an explicit motor output based on a perceptual judgment, their perceptual system deploys a suitable chart (**Figure 8C**): all geometrical constructs within the charted region of perceptual space are assigned components specified by the selected coordinates, explicitly representing the machinery that is necessary to perform distance measurements. It then becomes possible to implement the operations described in Section 2, and generate corresponding motor behavior. The extent to which the selected chart aligns with geodesic paths may be controlled by cognitive factors, such as attention. For example, the curved-to-flat transition in **Figure 6** may reflect actual changes in the geometry of the neural representation, as it is driven by stimulus changes and is not under cognitive control (Neri, 2018b), while the flat-to-curved transition in **Figure 7** may reflect chart-induced misalignments associated with increasing uncertainty, as this factor is controlled by the state of the observer (Neri, 2010c). Furthermore, different charts may be deployed on different instances when a behavioral output is produced in response to the same stimulus (different colors in **Figure 8C**), i.e. with reference to a stable geometry for the neural landscape. This chart-based interpretation would naturally explain the co-existence of stable and heterogeneous/transient characteristics at the level of neural population responses (Low et al., 2018; Langdon and Engel, 2022): the stable component would represent the underlying instrinsic geometry, while the transient and variable component would represent the deployment of ever-changing (but consistent when overlapping) charts.

At present, it is not possible to assess the applicability of the hypothetical scheme proposed above. However, geometrical techniques are gaining popularity in the analysis of neural population activity (Sohn et al., 2019; Vyas et al., 2020; Chung and Abbott, 2021), potentially supporting characterizations that dovetail the behavioral framework developed here. In the future, we may have access to concomitant neural and behavioral measurements that are sufficiently rich to enable geometrical reconstructions such as those in **Figures 6–7**. Comprehensive measurements of this kind would support detailed assessment of the relationship between neural and behavioral components of sensory processing with reference to the hypotheses put forward in the preceding paragraphs.

Acknowledgments

This study was supported by the Agence nationale de la recherche (grants ANR-16-CE28-0016, ANR-19-CE28-0010-01751, ANR- 10-LABX-0087, ANR-10-IDEX-0001-02, ANR-22-CE93-0013), CNRS, and IIT.

A The iconic "dipper" effect

A.1 Dipper effect: theory

Although our primary experimental point of reference is represented by kernel estimates (**Section 3.2**), we also consider one of the most iconic results in visual psychophysics: the dipper effect (Solomon, 2009). This is a counterintuitive phenomenon whereby the just noticeable difference δ_{thr} between a given stimulus of intensity p and another stimulus of intensity $p{+}\delta$ initially becomes gradually smaller as p increases, but then becomes gradually larger as p further increases beyond a certain value. In other words, δ_{thr} does not vary monotonically with p, as often assumed by coarse accounts of sensory discrimination such as Weber's law, which dictates that $\delta_{\text{thr}}{\propto}p$.

We assume one fixed form of internal noise. More specifically we assume that, every time the perceptual system performs a distance measurement, this measurement is subject to an estimation error (Section A.2), analogous to taking measurements of physical quantities in the real world. In further analogy with many such physical measurements, we also assume that the magnitude of such errors scales linearly with distance to be measured: the larger the distance, the proportionally larger the standard deviation of the distribution associated with multiple measurements of said distance. We refer to this type of noise with the term "Weber's noise."

We start with the projection rule. In the simplified case of a typical dipper experiment, all variables involved are simply scalars: $\underset{\perp}{R}(\mathbf{s}) = w(p{+}\delta)$ and $\underset{\perp}{R}(\tilde{\mathbf{s}}) = wp$, where w is the weight assigned by the system to the stimulus. To obtain approximate analytical expressions for δ_{thr}, we recall that the resolving power of the system (its sensitivity) can be defined as follows (Green and Swets, 1966):

$$d' = \frac{R(\mathbf{s}) - R(\tilde{\mathbf{s}})}{\sigma[R(\tilde{\mathbf{s}})]} \tag{26}$$

where $\sigma[x]$ is the standard deviation of variable x. It should be noted that, technically, the above definition for d' only makes sense when the distributions of variables $R(\mathbf{s})$ and $R(\tilde{\mathbf{s}})$ have comparable standard deviations (Green and Swets, 1966). This is not the case for Weber's noise, but we adopt this definition here as a reasonable approximation, to be validated by computer simulations later (**Figure 9**).

Under the projection rule, $\underset{\perp}{R}(\tilde{\mathbf{s}})$ involves only one distance measurement (equation 3), so $\sigma[R(\tilde{\mathbf{s}})]{\propto}R(\tilde{\mathbf{s}})$ and $d'{\propto}\frac{\delta}{p}$. At threshold performance (constant d'), this implies $\delta_{\text{thr}}{\propto}p$ (Weber's law). In its most basic form, the projection rule cannot therefore account for the dipper effect.

We now consider the relative distance rule. Under this rule, $\underset{\Vdash}{R}(\tilde{\mathbf{s}})$ involves two distance measurements (equation 5): $D(\tilde{\mathbf{s}}, \tilde{\mathbf{w}}){=}(p - \tilde{w})^2$ and $D(\tilde{\mathbf{s}}, \mathbf{w}){=}(p - w)^2$. We therefore have that $\underset{\Vdash}{R}(\tilde{\mathbf{s}}) = (p{-}\tilde{w})^2{-}(p{-}w)^2$ and $\sigma[R(\tilde{\mathbf{s}})] = \eta\sqrt{(p - \tilde{w})^4 + (p - w)^4}$, where we are taking into account both measurements in expressing σ, and η is a proportionality factor that dictates the overall intensity of internal noise. Without loss of generality, we can set $\tilde{w}{=}0$. When substituting the above expressions alongside $\underset{\Vdash}{R}(\mathbf{s}){=}(p + \delta)^2 - (p + \delta - w)^2$ into equation 26 and setting $d'{=}1$ to enforce threshold, we obtain:

$$\delta_{\text{thr}} = \frac{\eta\sqrt{p^4 + (p - w)^4}}{2w}$$

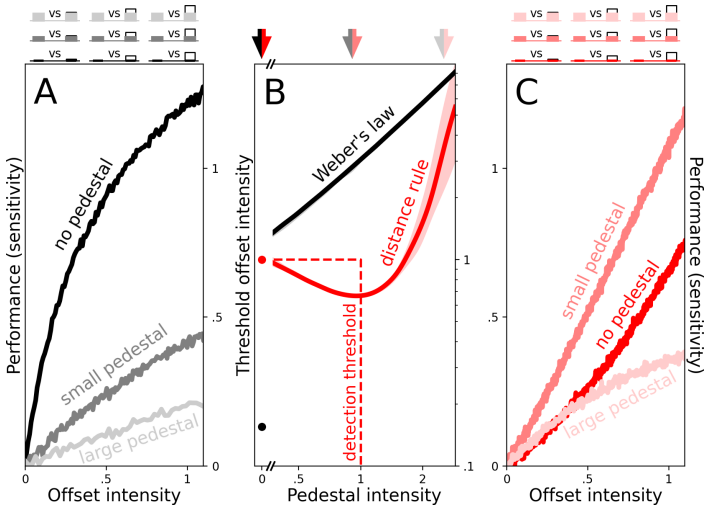

Figure 9: The relative distance rule naturally predicts the experimentally established dipper effect (Raab et al., 1963; Campbell and Kulikowski, 1966; Solomon, 2009). In this phenomenon, observers are asked to select the largest of two stimuli: one specified by a given pedestal intensity, the other one by an offset intensity added onto the same pedestal intensity (see icons above **A**, where offset is indicated by white rectangular element). As offset intensity is increased (x axis in **A**), discrimination performance (y axis) increases (traces in **A**). For a measuring system adopting the projection rule (black elements), and whose measurements are corrupted by intrinsic noise that scales with stimulus intensity (Weber's noise), the rate of performance increase as a function of offset intensity scales with pedestal intensity: it is fastest for zero pedestal intensity (black trace in **A**), slower for small pedestal intensity (dark-gray trace), and much slower for large pedestal intensity (light-gray trace). The threshold offset intensity associated with a fixed performance level (y axis in **B**) scales linearly with pedestal intensity (x axis) in accordance with Weber's law (black line in **B**). For a measuring system adopting the relative distance rule (red elements) and subjected to the same Weber's noise (added to each distance measurement), the fastest rate of performance increase is not observed for zero pedestal intensity (red trace in **C**), but for an intermediate small value of pedestal intensity (dark-salmon trace in **C**). The corresponding threshold-offset-vs-pedestal-intensity function is therefore non-monotonic (red trace in **B**), with a trough that aligns with the pedestal intensity (dashed vertical line) that corresponds to the threshold offset intensity associated with zero pedestal intensity (dashed horizontal line, detection threshold). All numerical values are in units of this detection threshold. Shaded regions in **B** show 95% boundaries around median (line) across 100 iterations of simulated experiments.

This expression is not monotonic: its derivative with respect to p vanishes for $p = w/2$. This value is similar to the detection threshold: $\delta_{\text{thr}}{=}w\eta/2$ for $p{=}0$. When $\eta{\approx}1$, which is typically the case in human sensory processing (Neri, 2010a), the dipper is therefore expected for a pedestal value near detection threshold, as reported experimentally (Raab et al., 1963; Campbell and Kulikowski, 1966; Solomon, 2009).

A.2 Dipper effect: simulations

We confirmed the above conclusions with computer simulations. This is necessary because the expression for sensitivity (equation 26) does not necessarily reflect the process of measuring a psychometric curve in a 2-alternative-forced-choice pedestal experiment, and estimating δ_{thr} directly from said curve. Furthermore, the notion of internal noise that scales linearly with stimulus intensity creates difficulties around $p{=}0$ (detection): for the projection rule, for example, this leads to the anomalous result that

the response to the pedestal becomes noiseless ($\sigma[R(\tilde{\mathbf{s}})]=0$). In computer simulations, we circumvented this issue by adding a small amount of "dark noise" (Barlow et al., 1957; Fuortes and Yeandle, 1964): a fixed factor added to the standard deviation of the internal noise source that remains present at zero stimulus intensity, and for which there is empirical support in human vision (Barlow et al., 1957; Srebro and Jaffe, 1968).

We simulated internal noise as exponentially distributed, with mean (and consequently standard deviation) set to stimulus intensity ($\eta=1$) plus the dark-noise factor mentioned above. Our choice of the exponential distribution was motivated by various considerations, for example the desirability of selecting a 1-parameter distribution with standard deviation that naturally scales with the mean. Furthermore, there is empirical evidence that internal noise in human sensory processing is Laplace-distributed (Neri, 2013), which is consistent with perceptual judgments based on differences of eponentially distributed variables. Finally, although very speculative and unspecified at this stage, it is not inconceivable that the process of geodesic exploration may share some distributional similarities with heat diffusion (Crane et al., 2017) and maximum-entropy energy allocation, making the Boltzmann distribution a potential candidate for modelling the spread of distance estimates over perceptual space. The current state of affairs is, unfortunately, that we do not know what the distribution of internal noise in human sensory processing conforms to. It has been assumed Gaussian without direct empirical evidence (Neri, 2013), however this assumption is almost certainly incorrect (Neri, 2013; Goris et al., 2014). For the purpose of our conclusions here, the choice of distribution is not critical, provided noise intensity scales with stimulus response.

We simulated end-to-end psychophysical experiments in which the proportion of correct responses was measured over 200 trials of scalar stimuli defined by p and $p+\delta$. As offset intensity δ is increased (along x axis in **Figure 9A**,**C**), sensitivity (Z-scored proportion of correct responses, plotted on y axis) also increases (traces in **Figure 9A**,**C**). For a system that implements the projection rule (**Figure 9A**), the rate of increase becomes progressively slower as pedestal intensity p is increased (see Section A.1). For example, when $p=0$ (detection, see black icons above **Figure 9A**) sensitivity increases rapidly with offset intensity (black trace in **Figure 9A**). When p is increased to a slightly larger value (dark-gray icons), sensitivity increases more slowly with offset intensity (dark-gray trace), and this effect becomes more pronounced when p is increased to an even larger value (light-gray icons/trace).

From the above sensitivity-vs-offset traces, we measured threshold offset δ_{thr}: the δ value for which sensitivity reaches a fixed arbitrary value ($1/\sqrt{2}$ in these simulations). When plotted (y axis in **Figure 9B**) against pedestal intensity p (x axis), this quantity increases linearly for the projection rule (black trace), in conformity with Weber's law.

The relative distance rule presents a different behavior (**Figure 9C**). When pedestal intensity is increased from $p=0$ (red elements in **Figure 9C**) to a slightly larger value (dark-salmon elements), sensitivity increases more rapidly with offset intensity (compare red with dark-salmon traces). As pedestal intensity is increased further (light-salmon elements), sensitivity increases more slowly (compare red with light-salmon traces). This non-monotonic behavior is reflected by the red trace in **Figure 9B**. As expected from theory (see Section A.1), this trace reaches a minimum for $p=1$, which corresponds to the value of δ_{thr} when $p=0$ (all values have been normalized by this specific δ_{thr} value).

References

Abbey, C. K. and Eckstein, M. P. (2002). Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J Vis*, 2:66–78.

Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision. *Computational models of visual processing (Landy M and Movshon AJ eds), Cambridge: MIT Press*, pages 3–20.

Ahumada, A. and Lovell, J. (1971). Stimulus features in signal detection. *J. Acoust. Soc. Am.*, 49:1751–1756.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *J Vis*, 2:121–131.

Barack, D. L. and Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, 22(6):359–371.

Barlow, H. B., Fitzhugh, R., and Kuffler, S. W. (1957). Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol.*, 137(3):338–354.

Barth, E., Beard, B. L., and Ahumada, A. J. J. (1999). Nonlinear features in vernier acuity. *Human Vision and Electronic Imaging III. SPIE Proceedings (Rogowitz BE and Pappas TN eds), San Jose, CA: SPIE*, pages 88–96.

Beard, B. L. and Ahumada, A. J. (1998). A technique to extract relevant image features for visual tasks. *Proceedings of SPIE*, 3299:79–85.

Brunelli, R. and Poggio, T. (1997). Template matching: matched spatial filters and beyond. *Pattern Recognition*, 30:751–768.

Burgess, A. E. and Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *J Opt Soc Am A*, 5:617–627.

Bussgang, J. J. (1952). Cross-correlation functions of amplitude-distorted Gaussian signals. *MIT Research Laboratory Electricity Technical Report*, 216:1–14.

Caelli, T., Hoffman, W., and Lindman, H. (1978). Subjective lorentz transformations and the perception of motion. *J. Opt. Soc. Am.*, 68(3):402–411.

Campbell, F. W. and Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *J Physiol*, 187(2):437–445.

Carandini, M. (2024). Sensory choices as logistic classification. *Neuron*, S0896-6273(24)00451-3.

Carandini, M. and Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.*, 13:51–62.

Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22(9):1512–1520.

Chung, S. and Abbott, L. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70:137–144. Computational Neuroscience.

Crane, K., Weischedel, C., and Wardetzky, M. (2017). The heat method for distance computation. *Commun. ACM*, 60(11):90–99.

Dai, H., Nguyen, Q., and Green, D. M. (1996). Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty. *J. Acoust. Soc. Am.*, 99:2298–2306.

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press.

Ermentrout, G. and Terman, D. H. (2010). *Mathematical Foundations of Neuroscience*. Springer: Cambridge University Press.

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.*, 9(4):292–303.

Farup, I. (2014). Hyperbolic geometry for colour metrics. *Optics Express*, 22(10):12369.

Fuortes, M. G. and Yeandle, S. (1964). Probability of occurrence of discrete potential waves in the eye of limulus. *J. Gen. Physiol.*, 47(3):443–463.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., and Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

Golden, J. R., Vilankar, K. P., Wu, M. C., and Field, D. J. (2016). Conjectures regarding the nonlinear geometry of visual neurons. *Vision Research*, 120:74–92.

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nat. Neurosci.*, 17(6):858–865.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806.

Heath, R. A. (2000). *Nonlinear dynamics: techniques and applications in psychology*. New Jersey: Lawrence Erlbaum Associates.

Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. (1996). Computational models of cortical visual processing. *Proc. Natl. Acad. Sci. U.S.A.*, 93:623–627.

Hoffman, W. C. (1989). The visual cortex is a contact bundle. *Applied Mathematics and Computation*, 32(2–3):137–167.

Indow, T. (1967). Two interpretations of binocular visual space: Hyperbolic and euclidean. *Annals of the Japan Association for Philosophy of Science*, 3(2):51–64.

Joosten, E. R., Shamma, S. A., Lorenzi, C., and Neri, P. (2016). Dynamic reweighting of auditory modulation filters. *PLoS Comput. Biol.*, 12(7):e1005019.

Joosten, E. R. M. and Neri, P. (2012). Human pitch detectors are tuned on a fine scale, but accessed on a coarse scale". *Biological Cybernetics*, 106:465–482.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, 50(5):363–370.

Koenderink, J. J. (1990). *Solid shape*. MIT Press, Cambridge, MA, United States.

Kriegeskorte, N. and Wei, X.-X. (2021). Neural tuning and representational geometry. *Nature Reviews Neuroscience*, 22(11):703–718.

Langdon, C. and Engel, T. A. (2022). Latent circuit inference from heterogeneous neural responses during cognitive tasks. *bioRxiv*.

Langdon, C., Genkin, M., and Engel, T. A. (2023). A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience*, 24(6):363–377.

Low, R. J., Lewallen, S., Aronov, D., Nevers, R., and Tank, D. W. (2018). Probing variability in a cognitive map using manifold inference from neural dynamics. *bioRxiv*.

Luneburg, R. K. (1950). The metric of binocular visual space*. *Journal of the Optical Society of America*, 40(10):627.

Lyu, C., Abbott, L. F., and Maimon, G. (2021). Building an allocentric travelling direction signal via vector computation. *Nature*, 601(7891):92–97.

Mamassian, P., Landy, M., Maloney, L., and Lewicki, S. (2002). 1 bayesian modelling of visual perception. *Probabilistic Models of the Brain: Perception and Neural Function*.

Marmarelis, P. Z. and Marmarelis, V. Z. (1978). *Analysis of Physiological Systems: the White-Noise Approach*. New York: Plenum Press.

Marmarelis, V. Z. (2004). *Nonlinear Dynamic Modeling of Physiological Systems*. Piscataway: New Jersey: Wiley IEEE Press.

Murray, R. F. (2011). Classification images: A review. *J Vis*, 11 (5):1–25 (doi:10.1167/11.5.2).

Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2002). Optimal methods for calculating classification images: weighted sums. *J Vis*, 2:79–104.

Neri, P. (2004). Estimation of nonlinear psychophysical kernels. *J Vis*, 4:82–91.

Neri, P. (2010a). How inherently noisy is human sensory processing? *Psychon Bull Rev*, 17:802–808.

Neri, P. (2010b). Stochastic characterization of small-scale algorithms for human sensory processing. *Chaos*, 20:045118.

Neri, P. (2010c). Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. *Front Comput Neurosci*, 4:151.

Neri, P. (2013). The statistical distribution of noisy transmission in human sensors. *J Neural Eng*, 10(1):016014.

Neri, P. (2015). The elementary operations of human vision are not reducible to template matching. *PLoS Comput. Biol.*, 11(11):e1004499.

Neri, P. (2018a). Classification images as descriptive statistics. *J. Mathematical Psychology*, 82:26–37.

Neri, P. (2018b). The empirical characteristics of human pattern vision defy theoretically-driven expectations. *PLoS Comput. Biol.*, 14(12):e1006585.

Neri, P. (2022). Deep networks may capture biological behavior for shallow, but not deep, empirical characterizations. *Neural Netw*, 152:244–266.

Neri, P. (2024). Local geometry of elementary visual computations. *MODVIS 2024 Abstract (Computational Models of Edges & Contours).*

Neri, P. and Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nat. Neurosci.*, 5:812–816.

Neri, P., Parker, A. J., and Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, 401:695–698.

Ostojic, S. and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. *PLoS Comput. Biol.*, 7(1):e1001056.

Petitot, J. (2017). *Elements of Neurogeometry*. Springer.

Pritchett, L. M. and Murray, R. F. (2015). Classification images reveal decision variables and strategies in forced choice tasks. *Proc. Natl. Acad. Sci. U.S.A.*, 112(23):7321–7326.

Raab, D. H., Osman, E., and Rich, E. (1963). Intensity Discrimination, the "Pedestal" Effect, and "Negative Masking" with White-Noise Stimuli. *The Journal of the Acoustical Society of America*, 35(7):1053–1053.

Rieke, F. (1999). *Spikes: exploring the neural code*. Cambridge, MA: MIT Press.

Roberts, S. and Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing. *Psychol. Rev.*, 107(2):358–367.

Sarti, A., Citti, G., and Petitot, J. (2009). Functional geometry of the horizontal connectivity in the primary visual cortex. *Journal of Physiology-Paris*, 103(1–2):37–45.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley.

Schütt, H. H. and Wichmann, F. A. (2017). An image-computable psychophysical spatial vision model. *J Vis*, 17(12):12.

Schwartz, O., Chichilnisky, E., and Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 269–276. MIT Press.

Shub, D. E. and Richards, V. M. (2009). Psychophysical spectro-temporal receptive fields in an auditory task. *Hear. Res.*, 251:1–9.

Sohn, H., Narain, D., Meirhaeghe, N., and Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, 103(5):934–947.e5.

Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *J Vis*, 2:105–120.

Solomon, J. A. (2009). The history of dipper functions. *Atten Percept Psychophys*, 71:435–443.

Srebro, R. and Jaffe, A. (1968). Origin of dark noise in human foveal vision. *J. Opt. Soc. Am.*, 58(9):1313.

Tjan, B. S. and Nandy, A. S. (2006). Classification images with uncertainty. *J Vis*, 6:387–413.

Turski, J. (2023). Riemannian geometries of visual space: Variable curvature and horizon. *Mathematical Methods in the Applied Sciences*, 46(8):9298–9324.

Varnet, L. and Lorenzi, C. (2022). Probing temporal modulation detection in white noise using intrinsic envelope fluctuations: A reverse-correlation study. *J. Acoust. Soc. Am.*, 151(2):1353.

Veksler, V. D., Myers, C. W., and Gluck, K. A. (2015). Model flexibility analysis. *Psychol. Rev.*, 122(4):755–769.

Vyas, S., Golub, M. D., Sussillo, D., and Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, 43(1):249–275.

Westwick, D. T. and Kearney, R. E. (2003). *Identification of Nonlinear Physiological Systems*. Piscataway: New Jersey: Wiley IEEE Press.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. richard courant lecture in mathematical sciences delivered at new york university, may 11, 1959. *Communications on Pure and Applied Mathematics*, 13(1):1–14.

Wu, T., Dufford, A. J., Mackie, M.-A., Egan, L. J., and Fan, J. (2016). The capacity of cognitive control estimated from a perceptual decision making task. *Scientific Reports*, 6(1).

Yamins, D. L. and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.*, 19(3):356–365.

Yilmaz, H. (1962). On color perception. *The Bulletin of Mathematical Biophysics*, 24(1):5–29.

Zheng, J. and Meister, M. (2024). The unbearable slowness of being: Why do we live at 10 bits/s? *Neuron*, 10.1016/j.neuron.2024.11.008.

Zhou, J., Duong, L. R., and Simoncelli, E. P. (2024). A unified framework for perceived magnitude and discriminability of sensory stimuli. *Proc Natl Acad Sci U S A*, 121(25):e2312293121.