Measuring Stimulus Information Transfer Between Neural Populations through the Communication Subspace

Oren Weiss[1-2] and Ruben Coen-Cagli[1-3]

[1]*Department of Systems and Computational Biology*
[2]*Dominick P. Purpura Department of Neuroscience*
[3]*Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Bronx, NY 10461, USA*

November 6, 2024

Abstract

Sensory processing arises from the communication between neural populations across multiple brain areas. While the widespread presence of neural response variability shared throughout a neural population limits the amount of stimulus-related information those populations can accurately represent, how this variability affects the interareal communication of sensory information is unknown. We propose a mathematical framework to understand the impact of neural population response variability on sensory information transmission. We combine linear Fisher information, a metric connecting stimulus representation and variability, with the framework of communication subspaces, which suggests that functional mappings between cortical populations are low-dimensional relative to the space of population activity patterns. From this, we partition Fisher information depending on the alignment between the population covariance and the mean tuning direction projected onto the communication subspace or its orthogonal complement. We provide mathematical and numerical analyses of our proposed decomposition of Fisher information and examine theoretical scenarios that demonstrate how to leverage communication subspaces for flexible routing and gating of stimulus information. This work will provide researchers investigating interareal communication with a theoretical lens through which to understand sensory information transmission and guide experimental design.

1   Introduction

Complex brain functions, such as behavior and perception, depend on interactions among anatomically and functionally distinct brain regions (Seguin et al. 2023). In sensory areas, neural populations represent the external world by transforming sensory input through local computations; the resulting signal is then transmitted to other cortical areas for further processing. For example, in the mammalian visual system, as visual input moves through the anatomical hierarchy, neurons tend to become increasingly selective ("tuned") to more complex visual features, from orientation tuning in primary visual cortex (V1) to selectivity for conjuctions of oriented features and for texture in mid-level visual areas V2, V4 and selectivity for object features in higher visual areas (Riesenhuber and Poggio 1999; DiCarlo et al. 2012). This increased selectivity is often characterized by relating changes in mean firing rates throughout neural populations to changes in external input.

Other aspects of neural population activity beyond the firing rate are largely ignored in studies of hierarchical sensory processing. One such feature is the variability of neural responses to repeated presentation of the same stimulus. Neural variability is present throughout the brain (Stein et al. 2005; Goris et al. 2014) and can limit a neuron's representational capacity (Tolhurst et al. 1983; Geisler and Albrecht 1997). In particular, this neural response variability is often shared throughout a neural population (Cohen and Kohn 2011; Panzeri et al. 2022), and the structure of this shared variability determines the amount of sensory information a population can encode (Shadlen and Newsome 1998; Rumyantsev et al. 2020; Averbeck et al. 2006; Moreno-Bote et al. 2014; Kohn, Coen-Cagli, et al. 2016; Kafashan et al. 2021; Kriegeskorte and Wei 2021). Furthermore, this variability might also constrain the transmission of information between populations in different cortical areas (Zylberberg et al. 2017; Huang, Pouget, et al. 2022; Rowland et al. 2023). With the advance of experimental techniques to record from multiple cortical areas simultaneously (Urai et al. 2022; Abdeladim et al. 2023), new theoretical tools are needed to understand the interplay between neural population variability and sensory transformations (Kang and Druckmann 2020; Kass et al. 2023).

To make progress towards this goal, here we consider a scheme for interareal communication proposed in recent work examining functional interactions in multi-area recordings. Researchers have found that the ability of one cortical area[1] (the "source" or "sender" area) to influence the activity of a downstream cortical area (the "target" or "receiver" area) depends on how well activity patterns in the source area align with specific directions in the space of neural population activity (Kaufman et al. 2014; Semedo, Zandvakili, et al. 2019; Javadzadeh and Hofer 2022). These axes are referred to as the *communication subspace*, as only fluctuations in activity that lie in this subspace get "communicated" to (i.e. are predictive of the activity of) a downstream region. Empirically, the dimensionality of the communication subspace is significantly less than the dimensionality of both the source and target population activity spaces (Semedo, Zandvakili, et al. 2019; Iyer et al. 2021; Srinath, Ruff, et al. 2021). Communication subspaces have thus been suggested as a mechanism for

---

[1]This paper will use the words "area" and "population" interchangeably. When one considers *intra-areal* communication between subpopulations in the same region of cortex using the same techniques used to characterize interareal communication, it is more correct to use "population" (e.g., the "source population").

signal multiplexing, in which multiple different signals can be simultaneously represented by a single neural population and sent to different downstream target areas (MacDowell et al. 2023; Barreiro et al. 2024), and as a means of flexible information routing, in which information is selectively communicated to downstream areas in a context-dependent manner (Kohn, Jasper, et al. 2020; Srinath, Ruff, et al. 2021; Srinath, Ni, et al. 2024; Ritz and Shenhav 2024).

Leveraging this scheme for interareal communication, in this paper we introduce a mathematical framework for studying in detail the transmission of sensory information across neural populations. We focus on the transmission of stimulus-related information between two brain areas. This is different from most studies of functional interactions, including existing work on the subspace, which do not distinguish explicitly between transmission of signals that convey stimulus information and other unrelated shared factors (e.g., endogenous global state fluctuations, Seguin et al. 2023). By examining stimulus-related information transfer, we can concentrate on how the content of a signal (i.e., the stimulus) is transmitted via communication subspaces instead of examining how much of the signal is shared between two areas, regardless of what information it contains (Ince et al. 2015; Celotto et al. 2023). Our method will combine communication subspaces with Fisher information, a widely used measure of stimulus-related information in a neural population that depends on the alignment between the trial-averaged population activity across stimulus values (the signal axis) and the population covariance (Beck et al. 2011; Moreno-Bote et al. 2014).

We aim to decompose the Fisher information in the source and target areas depending on how the neural population activity aligns with the communication subspace. This will allow us to investigate various aspects of communication separately, such as how information is shared between these areas, how information is transmitted from the source area to the target area, and how the received information affects the target area. We provide extensive mathematical derivations of this decomposition and how components of the Fisher information in each area relate to Fisher information lying in different subspaces. To illustrate the explanatory power of our framework, we will then demonstrate how flexible communication of sensory information can be achieved by modifying activity in the source or target areas to change the alignment of population activity to different subspaces of the population activity space.

Our results will guide researchers interested in understanding how stimulus-related information is propagated through the brain via communication subspaces. Because Fisher information (Kanitscheider et al. 2015) and the low-rank linear map (Semedo, Zandvakili, et al. 2019) can be efficiently estimated from neural population recordings with limited data, this method of decomposing Fisher information can easily be used as a data-analytic technique for explicitly measuring stimulus information transmitted through the communication subspace.

2   Methods

2.1   Defining Communication Subspaces via Reduced-Rank Regression

Consider the combined activity (e.g., spike counts) of all neurons in the source area and arrange them into a population activity matrix $\mathbf{x} \in \mathbb{R}^{q_x \times N}$, such that the i-th column contains the activity of neuron i across N identical experimental trials. We similarly define the target population activity matrix $\mathbf{y} \in \mathbb{R}^{q_y \times N}$. We assume that the source area activity is normally distributed across trials:

$$\mathbf{x}(\theta) = \mathbf{f}_x(\theta) + \boldsymbol{\eta}_x \qquad \boldsymbol{\eta}_x \sim \mathcal{N}(0, \boldsymbol{\Sigma}_x) \tag{1}$$

With $\boldsymbol{\Sigma}_x$ representing the cross-trial noise covariance in the source area and $\mathbf{f}_x$ the tuning curve with respect to the stimulus parameter θ, which describes how the mean neural activity across experimental trials changes with stimulus. We will assume that all noise covariance matrices are stimulus-independent.

Reduced-rank regression searches for a low-rank matrix (for a prespecified rank k) that best approximates the full-rank linear regression matrix of source activity $\mathbf{x}$ on target activity $\mathbf{y}$ (Izenman 1975; Tso 1981; Davies and Tso 1982). This can be written as:

$$\mathbf{y}(\theta) = \mathbf{B}_{\mathrm{cs}}\mathbf{x} + \mathbf{r}_y(\theta) + \boldsymbol{\eta}_y \qquad \boldsymbol{\eta}_y \sim \mathcal{N}(0, \sigma_2^2 \mathbf{I}_{q_y}) \tag{2}$$

$$\mathbf{B}_{\mathrm{cs}} = \mathbf{U}_k \mathbf{U}_k^\top \mathbf{B}_{\mathrm{OLS}} \tag{3}$$

Where $\mathbf{r}_y(\theta)$ is the residual tuning in the target area not explained by tuning from the source area (i.e., the tuning curve in the target area is defined as $\mathbf{f}_y = \mathbf{B}_{\mathrm{cs}}\mathbf{f}_x + \mathbf{r}_y$), $\mathbf{B}_{\mathrm{OLS}}$ is the full-rank linear regression matrix of $\mathbf{x}$ on $\mathbf{y}$, $\mathbf{U}_k$ are the top-k principal components of $\mathbf{B}_{\mathrm{OLS}}\mathbf{x}$ (equivalently, these are the top k left singular vectors of $\boldsymbol{\Sigma}_{yx}(\boldsymbol{\Sigma}_x^{-1/2})^\top$, with $\boldsymbol{\Sigma}_{yx}$ the cross-covariance between $\mathbf{y}$ and $\mathbf{x}$, and $\boldsymbol{\Sigma}_x^{-1/2}$ is the inverse Cholesky factor of $\boldsymbol{\Sigma}_x$, see Izenman 2008, Section 6.3.2), and σ_2^2 is the residual variance that is minimized in the reduced-rank regression. With this, the noise covariance in the target population, $\boldsymbol{\Sigma}_y$, is

$$\boldsymbol{\Sigma}_y = \mathbf{B}_{\mathrm{cs}} \boldsymbol{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top + \sigma_2^2 \mathbf{I}_{q_y} \tag{4}$$

Activity patterns in the source population (i.e., linear combinations of the columns of $\mathbf{x}$) that are not communicated to the downstream target population are said to lie in the source population's private subspace. In other words, an activity pattern v lies in the private subspace if $\mathbf{B}_{\mathrm{cs}}v = 0$. Therefore, the private subspace is equivalent to the kernel or null space of $\mathbf{B}_{\mathrm{cs}}$ (written $\ker(\mathbf{B}_{\mathrm{cs}})$, which is how we will define the private subspace). Since the vector spaces we are considering in this paper are finite-dimensional Euclidean spaces endowed with the standard inner product, we can then define the communication subspace as the orthogonal complement of the private subspace, $\ker(\mathbf{B}_{\mathrm{cs}})^\perp$, which is equivalent to the row space of $\mathbf{B}_{\mathrm{cs}}$, $\mathrm{Im}(\mathbf{B}_{\mathrm{cs}}^\top)$. Intuitively, the elements of the communication subspace are those activity patterns that are communicated via the reduced-rank regression mapping $\mathbf{B}_{\mathrm{cs}}$, as vectors in this subspace correspond one-to-one with vectors in $\mathrm{Im}(\mathbf{B}_{\mathrm{cs}})$ (Axler 2024,

Proposition 6.67).

For Section 2.2, we will want to consider projections onto the communication or private subspaces in the source area. Note that once we define one projection, say the projection onto the communication subspace, which we denote $\mathbf{P}_{cs}$, then the projection onto the private subspace, $\mathbf{Q}_{cs}$, is $\mathbf{I} - \mathbf{P}_{cs}$. To define these projections, we will use the (Moore-Penrose) pseudoinverse (written $\mathbf{A}^+$), which is a generalization of the inverse of a matrix or linear map to non-invertible transformations (Penrose 1955). As detailed in Appendix A, we can define the projection onto the communication subspace as

$$\mathbf{P}_{cs} := \mathbf{B}_{cs}^+ \mathbf{B}_{cs} \tag{5}$$

In the preceding discussion, we have implicitly assumed that the reduced-rank regression matrix and all its derived quantities (i.e., the communication and private subspaces and their associated projections) are independent of the stimulus parameter θ. We justified this assumption because the closed-form for the reduced-rank regression matrix is a function of the noise covariance matrices (Izenman 1975), which we assumed were stimulus-independent. When applying our method to experimental data, even in the presence of stimulus-dependent covariance, we expect that this will be a reasonable approximation when considering small differences in the stimulus parameter. More generally, researchers should consider if invariant subspaces are a good approximation for their dataset.

2.2 Fisher Information Decomposition

The main quantities derived in this section, and their meaning, are summarized in Table 1.

The linear Fisher information $\mathbf{J}(\mathbf{z}; \theta)$ computed from neural activity $\mathbf{z}$ with tuning curve $\mathbf{f}$ and noise covariance matrix $\mathbf{\Sigma}$ is written as (Kanitscheider et al. 2015):

$$\mathbf{J}(\mathbf{z}; \theta) = d\mathbf{f}^\top \mathbf{\Sigma}^{-1} d\mathbf{f} \tag{6}$$

Where $d\mathbf{f}$ is the derivative of the tuning curve, which is computed between two conditions using the finite difference quotient (i.e., $d\mathbf{f} = \dfrac{\mathbf{f}(\theta_2) - \mathbf{f}(\theta_1)}{\theta_2 - \theta_1}$, for θ_1 and θ_2 the two different stimulus conditions). This also defines the signal axis in population activity space (Panzeri et al. 2022). The linear Fisher information represents the smallest change in the stimulus that can be discriminated from a neural population using an optimal linear decoder (Kohn, Coen-Cagli, et al. 2016). This is a component of the full Fisher information; in this paper, we will exclusively focus on the linear Fisher information and thus refer to Eq 6 as "Fisher information".

2.2.1 Fisher Information in the communication and private subspaces

An important quantity for our goal of studying information transmission between areas, is the Fisher information in the communication subspace. There are two possible definitions for this quantity. The first is the Fisher information present in the source population projected

into the communication subspace, namely

$$\mathbf{J}(\mathbf{P}_{\mathrm{cs}}\mathbf{x};\theta) = d\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \left(\mathbf{P}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{P}_{\mathrm{cs}}\right)^+ \mathbf{P}_{\mathrm{cs}}d\mathbf{f}_x \tag{7}$$

This quantity is closely related to a similar quantity that measures a binary discrimination threshold $(d')^2$ after using a targeted dimensionality reduction technique (Ebrahimi et al. 2022). The other definition is the Fisher information present in the activity patterns mapped by the reduced-rank regression matrix $\mathbf{B}_{\mathrm{cs}}$:

$$\mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x};\theta) = d\mathbf{f}_x^\top \mathbf{B}_{\mathrm{cs}}^\top \left(\mathbf{B}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{B}_{\mathrm{cs}}^\top\right)^+ \mathbf{B}_{\mathrm{cs}}d\mathbf{f}_x \tag{8}$$

In both expressions, we have to use the pseudoinverse rather than the full matrix inverse, as both matrices are rank-deficient ($\mathbf{P}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{P}_{\mathrm{cs}}$ is a $q_x \times q_x$ matrix of rank k while $\mathbf{B}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{B}_{\mathrm{cs}}^\top$ is a $q_y \times q_y$ matrix of the same rank).

These two definitions should be equivalent, as $\mathbf{B}_{\mathrm{cs}}$, restricted to the communication subspace, is an isomorphism onto its range, and a simple calculation shows that invertible linear transformations of activity will not change the amount of Fisher information (Huang, Pouget, et al. 2022). However, that calculation does not necessarily translate once we move from matrix inverses into pseudoinverses because, in general, the pseudoinverse of a product of matrices is not the product of their pseudoinverses (Hartwig 1986). Nonetheless, we can prove that these two expressions are equivalent (Appendix B). Thus, we have a well-defined expression for the amount of Fisher information in the communication subspace, which we will denote $\mathbf{J}_{\mathrm{cs}}$:

$$\mathbf{J}_{\mathrm{cs}} := \mathbf{J}(\mathbf{P}_{\mathrm{cs}}\mathbf{x};\theta) = d\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \left(\mathbf{P}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{P}_{\mathrm{cs}}\right)^+ \mathbf{P}_{\mathrm{cs}}d\mathbf{f}_x \tag{9}$$

We can similarly define the information in the private subspace as:

$$\mathbf{J}_{\mathrm{priv}} := \mathbf{J}(\mathbf{Q}_{\mathrm{cs}}\mathbf{x};\theta) = d\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}} \left(\mathbf{Q}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{Q}_{\mathrm{cs}}\right)^+ \mathbf{Q}_{\mathrm{cs}}d\mathbf{f}_x \tag{10}$$

By the data processing inequality for (full) Fisher information (Zamir 1998), we have that the (linear) Fisher information in the source population is an upper bound on the information contained in the communication and private subspaces (see also Appendix B, Lemma 2 which provides a more direct proof of these inequalities as a consequence):

$$\mathbf{J}(\mathbf{x};\theta) \geq \max\left\{\mathbf{J}_{\mathrm{cs}}, \mathbf{J}_{\mathrm{priv}}\right\} \tag{11}$$

2.2.2 Decomposing Fisher Information in the source and target areas

Using the identity $\mathbf{f}_x = \mathbf{P}_{\mathrm{cs}}\mathbf{f}_x + \mathbf{Q}_{\mathrm{cs}}\mathbf{f}_x$ (by definition, $\mathbf{P}_{\mathrm{cs}} + \mathbf{Q}_{\mathrm{cs}} = \mathbf{I}$) and Eq 6, it is easy to derive the following decomposition of the source Fisher information:

$$\mathbf{J}(\mathbf{x};\theta) = d\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}}\mathbf{\Sigma}_x^{-1}\mathbf{P}_{\mathrm{cs}}d\mathbf{f}_x + d\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}}\mathbf{\Sigma}_x^{-1}\mathbf{Q}_{\mathrm{cs}}d\mathbf{f}_x + 2d\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}}\mathbf{\Sigma}_x^{-1}\mathbf{Q}_{\mathrm{cs}}d\mathbf{f}_x \tag{12}$$

We can similarly decompose the target Fisher information using the identity $\mathbf{f}_y = \mathbf{B}_{\mathrm{cs}}\mathbf{f}_x + \mathbf{r}_y$:

$$\mathbf{J}(\mathbf{y};\theta) = d\mathbf{f}_x^\top \mathbf{B}_{\mathrm{cs}}^\top \mathbf{\Sigma}_y^{-1}\mathbf{B}_{\mathrm{cs}}d\mathbf{f}_x + d\mathbf{r}_y^\top \mathbf{\Sigma}_y^{-1}d\mathbf{r}_y + 2d\mathbf{f}_x^\top \mathbf{B}_{\mathrm{cs}}^\top \mathbf{\Sigma}_y^{-1}d\mathbf{r}_y \tag{13}$$

6

We now provide an interpretation of the different terms in the decompositions above. We will begin with the source Fisher information decomposition and consider the first two terms, which we will define as:

$$\mathbf{CI}_{\mathrm{cs}} = \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \boldsymbol{\Sigma}_x^{-1} \mathbf{P}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x \tag{14}$$

$$\mathbf{CI}_{\mathrm{priv}} = \mathrm{d}\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}} \boldsymbol{\Sigma}_x^{-1} \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x \tag{15}$$

In Appendix C we show (using a technique similar to Zhang 2011, Eq 7.11), that:

$$\mathbf{CI}_{\mathrm{cs}} \geq \mathbf{J}(\mathbf{P}_{\mathrm{cs}}\mathbf{x}; \theta) \tag{16}$$

$$\mathbf{CI}_{\mathrm{priv}} \geq \mathbf{J}(\mathbf{Q}_{\mathrm{cs}}\mathbf{x}; \theta) \tag{17}$$

Our understanding of Eqs 14 and 15 is that they are the Fisher information *contributed* to either the communication (Eq 14) or private (Eq 15) subspaces, as they are the Fisher information for a neural population whose tuning curve lies entirely in a subspace (i.e., if $\mathbf{P}_{\mathrm{cs}}\mathbf{f}_x = \mathbf{f}_x$, then the other terms in Eq 12 are 0 and $\mathbf{J}(\mathbf{x}; \theta) = \mathbf{CI}_{\mathrm{cs}}$, and similarly for the private subspace). It is tempting to think that these contributed information terms are themselves bounded by the total source population Fisher information (for example, $\mathbf{J}(\mathbf{x}; \theta) \geq \mathbf{CI}_{\mathrm{cs}}$). However, this is not true in general as we show in Appendix D.

The final term of the source Fisher information decomposition (Eq 12) is a cross-term that measures the angle between $\mathbf{P}_{\mathrm{cs}}\mathrm{d}\mathbf{f}_x$ and $\mathbf{Q}_{\mathrm{cs}}\mathrm{d}\mathbf{f}_x$ relative to the overall covariance in the source area $\boldsymbol{\Sigma}_x$. Mathematically, this accounts for the fact that contributed information to the communication and private subspaces can exceed the overall source Fisher information; this cross-term can be thought of as correcting for overcounting when adding these contributed information quantities. Although the communication and private subspaces are orthogonal, that does not mean they are independent (Semedo, Gokcen, et al. 2020). In particular, the information in these subspaces can overlap, as activity in the communication subspace is predictive of the activity in the private subspace and vice versa (Semedo, Zandvakili, et al. 2019). For this reason, we will call this term the *shared information* between the communication and private subspaces:

$$\mathbf{SI}_1 = \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \boldsymbol{\Sigma}_x^{-1} \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x. \tag{18}$$

Next, we will consider the components of the target Fisher information decomposition (Eq 13). The first term, which we will call the *impactful information*, is the amount of Fisher information in the target population when $\mathbf{r}_y = 0$, which is when the mean tuning of the source activity completely determines the tuning of the target population. We will denote this term as $\mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$, which is similar in structure to the Fisher information that lies in the communication subspace (Eq 7), except that $\mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$ includes covariance arising from the additional noise $\boldsymbol{\eta}_y$ (Eq 4). These terms are related by:

$$\mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x}; \theta) \geq \mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x} + \boldsymbol{\eta}_y; \theta) \tag{19}$$

While reduced-rank regression requires that the target noise be a diagonal matrix with isotropic noise (i.e., $\boldsymbol{\eta}_y \sim \mathcal{N}(0, \sigma_2^2\mathbf{I})$, see Eq 2), the above relation also holds if $\boldsymbol{\eta}_y$ has an

arbitrary covariance matrix Σ_r. This is important for possible extensions to the methods proposed here that would apply to related dimensionality reduction techniques such as canonical correlation analysis that have been used in the literature to characterize cortical communication (Semedo, Jasper, et al. 2022; Semedo, Gokcen, et al. 2020; Gokcen, Jasper, Xu, et al. 2024). There are two ways to demonstrate that this relationship holds. Indirectly, this can be seen as another consequence of the data processing inequality for Fisher information (Zamir 1998), as the chain relationship $\theta \to \mathbf{B}_{cs}\mathbf{x} \to \mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y$ is trivially a Markov chain. The more direct argument is provided in Appendix E.

The second term in the target Fisher information (Eq 13) will be called the *residual information* (**RI**). This term would reflect the stimulus-related information present in the target area when no tuning information is linearly communicated between the source and target (i.e., when $\mathbf{B}_{cs}\mathbf{f}_x = 0$, $\mathbf{J}(\mathbf{y};\theta) = \mathrm{d}\mathbf{r}_y^\top \Sigma_y^{-1} \mathrm{d}\mathbf{r}_y$) or equivalently, after subtracting the influence of the source area on the target.

Lastly, the final term in the target Fisher information decomposition (Eq 13) provides a measure of the interaction between the impactful information and the residual information, as it measures the angle between $\mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x$ and $\mathrm{d}\mathbf{r}_y$ relative to principal axes of the target covariance matrix Σ_y. For this reason, we will call it the *synergistic/redundant information* between the residual and transmitted tuning:

$$\mathbf{SI}_2 = \mathrm{d}\mathbf{f}_x^\top \mathbf{B}_{cs}^\top \Sigma_y^{-1} \mathrm{d}\mathbf{r}_y \tag{20}$$

This paper will not examine these terms as we assume no residual tuning for simplicity (see Section 2.3).

2.2.3 Considerations on estimation from finite data

With our choice to focus on Fisher information, applications of the framework proposed here to real data can leverage existing methods to estimate linear Fisher information efficiently and accurately for large neural populations. These methods include linear decoding (Kohn, Coen-Cagli, et al. 2016) and direct bias-corrected estimators (Kanitscheider et al. 2015). The direct estimators are straightforward; they require measuring the empirical tuning derivatives (by finite difference) and noise covariances, projecting them through the appropriate mapping matrices, and applying analytical bias corrections. Using linear decoding requires some additional considerations. Specifically, some terms of our decompositions do not immediately correspond to decoding measured activity. For instance, $\mathbf{CI}_{cs}$ uses the full noise covariance of the source area, but only the tuning derivatives projected onto the subspace (not the full tuning of the source area, see Eq 14), so it cannot be estimated by decoding the source area activity. However, in this case, it is easy to construct a transformation of the target area activity $\mathbf{x} - \mathbf{Q}_{cs}\mathbf{f}_x$, which has exactly those tuning derivatives and covariance. Therefore, $\mathbf{CI}_{cs}$ can be estimated by decoding after applying the transformation above. Similarly, the impactful information term can be estimated with a linear decoder computed from $\mathbf{y} - \mathbf{f}_y + \mathbf{B}_{cs}\mathbf{f}_x$. Lastly, the term denoted $\mathbf{SI}_1$ can be computed from data by using the linear decoders for $\mathbf{x}$, $\mathbf{CI}_{cs}$ and $\mathbf{CI}_{priv}$ (discussed above) with Eq 12 and, similarly, we can also estimate $\mathbf{SI}_2$ from data by combining estimates for $\mathbf{J}(\mathbf{y};\theta)$, $\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y;\theta)$ and **RI**.

2.3   Simulations

In Section 3, we will analyze Fisher information and derived quantities (the quantities in Eqs 9, and 10, 12, and 13) computed directly from simulated tuning curves, reduced-rank regression matrices, and covariance matrices. For the baseline simulations performed, we need to specify the tuning curves $\mathbf{f}_x, \mathbf{f}_y$, the source covariance $\mathbf{\Sigma}_x$, the reduced-rank regression matrix $\mathbf{B}_{cs}$, and the residual variance σ_2^2. The number of neurons in the source (q_x) and target (q_y) area will be 50. We will assume that there is no residual tuning in the target area, such that all the stimulus-related information in the target area arises from the feedforward contribution of the source area ($\mathbf{f}_y = \mathbf{B}_{cs}\mathbf{f}_x$). The tuning in the source area will be von-Mises tuning curves, which resemble the orientation selectivity of neurons in V1. The stimulus parameter, θ, will take two values: $\theta = 0, \frac{\pi}{2}$. The i-th component of $\mathbf{f}_x$, $f_{x,i}$ is parameterized as

$$f_{x,i}(\theta) = b_i + g_i \exp(\kappa_i \cos(\theta - \phi_i)) \tag{21}$$

Where b_i is the bias, g_i is the amplitude, ϕ_i is the preferred orientation for neuron i, and κ_i controls the width of the tuning curve. Across all neurons in the source area, $b_i = 2, g_i = 30/e$. The value of ϕ_i is drawn from $\mathcal{U}([0, \pi])$, while the value for κ_i is drawn from $\mathcal{U}([0.5, 2])$. The correlation structure is generated as follows: first, we generate a limited-range correlation matrix $\mathbf{R}_\phi$ from the preferred orientations ϕ_i (with spatial scale $L = 1$ and correlation scale $c_0 = 0.3$, see Ecker et al. 2011). Then we draw a random correlation matrix $\mathbf{R}_{LKJ}$ from LKJ distribution with β parameter fixed at 30 (Lewandowski et al. 2009) and set a mixing proportion $s = 0.8$ to create the source correlation matrix $\mathbf{R}_1 = s\mathbf{R}_\phi + (1-s)\mathbf{R}_{LKJ}$. This allows for some negative noise correlations between neurons, as all limited-range correlations are positive. To form the noise covariance matrix $\mathbf{\Sigma}_x$, we set the variances to be proportional to the mean ("Poisson-like variability"): for all neurons i, we took the mean of the tuning curve across the two stimulus values $\overline{f_{x,i}} = \frac{1}{2}(f_{x,i}(\frac{\pi}{2}) + f_{x,i}(0))$, created a diagonal standard deviation matrix $\mathbf{S}$ with diagonal entries $\overline{f_{x,i}}^{1/2}$, and set $\mathbf{\Sigma}_x = \mathbf{S}\mathbf{R}_1\mathbf{S}$. The reduced-rank regression matrix is generated by taking a truncated SVD of a random Gaussian matrix: we generate a $q_y \times q_x$ matrix $\mathbf{B}$ with entries sampled from $\mathcal{N}(0, 1/\sqrt{q_y})$, compute the SVD $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, and retain only the top k singular values and vectors. We fixed the dimensionality of the communication subspace (the number of singular values to retain during truncation) at $k = 5$, on par with empirical data (Semedo, Zandvakili, et al. 2019). Finally, we fix the residual variance $\sigma_2^2 = 15$.

Additionally, we will modify the structure of the tuning curves and covariance matrices to exemplify certain properties of the information decomposition we are proposing and how this would influence using communication subspaces as a flexible stimulus-related information propagation scheme. These manipulations will be done to assess how manipulations of the model structure affect Fisher information quantities. This allows direct comparisons (per simulation) of the information quantities computed from the initially generated parameters and their perturbations.

3   Results

The communication subspace framework explains fluctuations in activity in a target neural population from a low-dimensional subset of activity in a source population (Fig 1). Intuitively, only information in this predictive subspace (the communication subspace) should be inherited by the activity in the target area. In contrast, information in the orthogonal complement (the private subspace) is not communicated to this area and can be used for local processing or communicated to a different target area (Semedo, Zandvakili, et al. 2019). The linkage between the two areas, the communication subspace, is analogous to a bottleneck through which information present in the source area must flow to get to the target area (although this is distinct from the information bottleneck technique proposed by Tishby et al. 2000). This paper examines these concepts in a more rigorous setting by investigating how stimulus-related information is transmitted through this system using Fisher information (Eq 6), which measures how well the stimulus values can be linearly decoded from neural population activity.
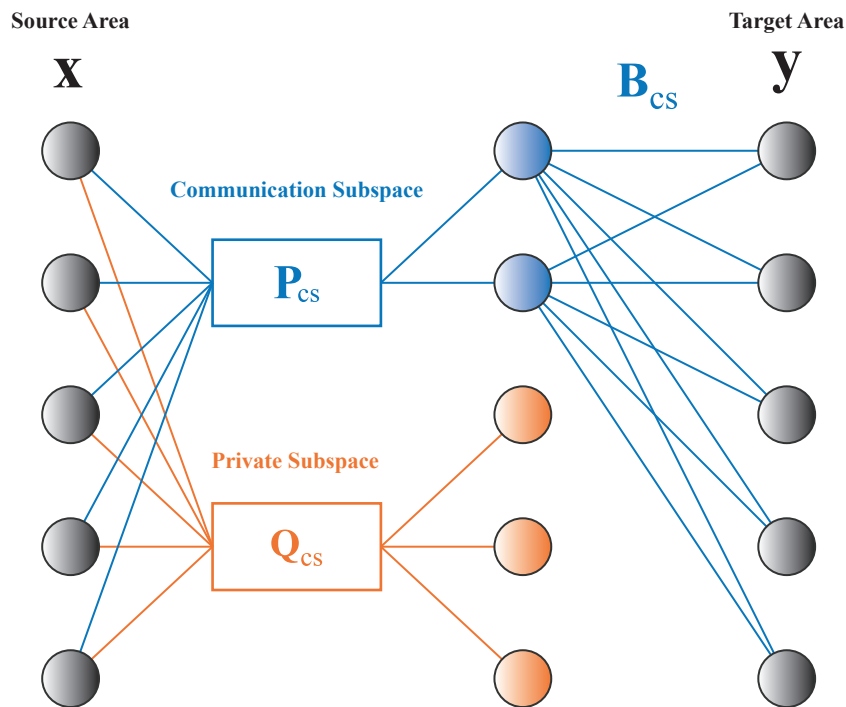

Figure 1. Communication Subspace

Representation of communication between two cortical areas (source and target) as a linear autoencoder. Linear combinations of neural activity in the source area can be divided into two orthogonal subspaces: the communication subspace (blue) and the private subspace (orange). Only activity in the communication subspace is used to linearly predict activity in the target area through the map $\mathbf{B}_{cs}$. The activity in the private subspace lies in the kernel of $\mathbf{B}_{cs}$, so is mapped to 0.

Symbol	Meaning	Definition (Reference)
$\mathbf{J}(\mathbf{x};\theta)$	Source Fisher Information	$d\mathbf{f}_x^\top \Sigma_x^{-1} d\mathbf{f}_x$ (Eq 12)
$\mathbf{J}(\mathbf{P}_{cs}\mathbf{x};\theta)$	Communicated Information	$d\mathbf{f}_x^\top \mathbf{P}_{cs} \left(\mathbf{P}_{cs}\Sigma_x \mathbf{P}_{cs}\right)^+ \mathbf{P}_{cs} d\mathbf{f}_x$ (Eq 7, Fig 2A)
$\mathbf{J}(\mathbf{Q}_{cs}\mathbf{x};\theta)$	Private Information	$d\mathbf{f}_x^\top \mathbf{Q}_{cs} \left(\mathbf{Q}_{cs}\Sigma_x \mathbf{Q}_{cs}\right)^+ \mathbf{Q}_{cs} d\mathbf{f}_x$ (Eq 10)
$\mathbf{CI}_{cs}$	Contributed Information to Communication Subspace	$d\mathbf{f}_x^\top \mathbf{P}_{cs}\Sigma_x^{-1}\mathbf{P}_{cs} d\mathbf{f}_x$ (Eq 14, Fig 2B)
$\mathbf{CI}_{priv}$	Contributed Information to Private Subspace	$d\mathbf{f}_x^\top \mathbf{Q}_{cs}\Sigma_x^{-1}\mathbf{Q}_{cs} d\mathbf{f}_x$ (Eq 15, Fig 2C)
$\mathbf{SI}_1$	Shared Information between Communication and Private Subspaces	$2 d\mathbf{f}_x^\top \mathbf{P}_{cs}\Sigma_x^{-1}\mathbf{Q}_{cs} d\mathbf{f}_x$ (Fig 3)
$\mathbf{J}(\mathbf{y};\theta)$	Target Fisher Information	$d\mathbf{f}_y^\top \Sigma_y^{-1} d\mathbf{f}_y$ (Eq 13)
$\mathbf{J}(\mathbf{B}_{cs}\mathbf{x};\theta)$	Mapped Information	$d\mathbf{f}_x^\top \mathbf{B}_{cs}^\top \left(\mathbf{B}_{cs}\Sigma_x \mathbf{B}_{cs}^\top\right)^+ \mathbf{B}_{cs} d\mathbf{f}_x$ (Eq 8, Fig 2A)
$\mathbf{J}(\mathbf{B}_{cs}\mathbf{x}+\eta_y;\theta)$	Impactful Information	$d\mathbf{f}_x^\top \mathbf{B}_{cs}^\top \Sigma_y^{-1}\mathbf{B}_{cs} d\mathbf{f}_x$
$\mathbf{RI}$	Residual Information	$d\mathbf{r}_y^\top \Sigma_y^{-1} d\mathbf{r}_y$
$\mathbf{SI}_2$	Synergistic/Redundant Information	$2 d\mathbf{f}_x^\top \mathbf{B}_{cs}^\top \Sigma_y^{-1} d\mathbf{r}_y$

Table 1. Table containing definitions of the relevant Fisher information-theoretic terms.

3.1 Measuring stimulus-releated information in communication subspaces

We first illustrate our proposed framework to characterize stimulus-related information transmission through the communication subspace, using linear Fisher Information $\mathbf{J}$ about a stimulus variable θ (details and proofs in Section 2.2; see also Fig 1 and Table 1 for definitions of terms). There are two possible definitions for the information "in" the communication subspace: the information computed from the source activity projected onto the communication subspace ($\mathbf{J}(\mathbf{P}_{cs}\mathbf{x};\theta)$, Eq 7) and the information computed from the source activity mapped ono the target via the reduced-rank regression matrix ($\mathbf{J}(\mathbf{B}_{cs}\mathbf{x};\theta)$, Eq 8). We showed analytically and with simulations (Fig 2A) that these quantities are equal (the common term written $\mathbf{J}_{cs}$, Eq 7). Similarly, we also defined the term $\mathbf{J}_{priv}$ (Eq 10) denoting Fisher information in the private subspace.

Next, we derived equations (Eqs 12 and 13) that decompose information in the source ($\mathbf{x}$) and target ($\mathbf{y}$) areas with respect to the low-rank linear mapping between the two areas (Eq 2). In the decomposition for the source area, there are two terms ($\mathbf{CI}_{cs}$ and $\mathbf{CI}_{priv}$, Eqs 14 and 15) that are very similar to the definitions for the Fisher information in the communication and private subspaces ($\mathbf{J}_{cs}$ and $\mathbf{J}_{priv}$), except that they use the full source area covariance Σ_x as opposed to the covariance of the source activity after being projected into the corresponding subspace. We proved a lemma (Appendix C) which demonstrated that $\mathbf{CI}_{cs}$ and $\mathbf{CI}_{priv}$ provide upper bounds for $\mathbf{J}_{cs}$ and $\mathbf{J}_{priv}$ (Eqs 16 and 17; simulations in Fig 2B and 2C). For this reason, we named $\mathbf{CI}_{cs}$ and $\mathbf{CI}_{priv}$ the *contributed information* to either the communication or private subspace. Lastly, we showed that $\mathbf{J}_{cs}$ is an upper bound on the information present in the target area when all stimulus-related activity in the target population is entirely driven by the source population ($\mathbf{J}(\mathbf{B}_{cs}\mathbf{x}+\eta_y;\theta)$, Fig 2D). Therefore, we termed this latter quantity *impactful information* because it measures how much the

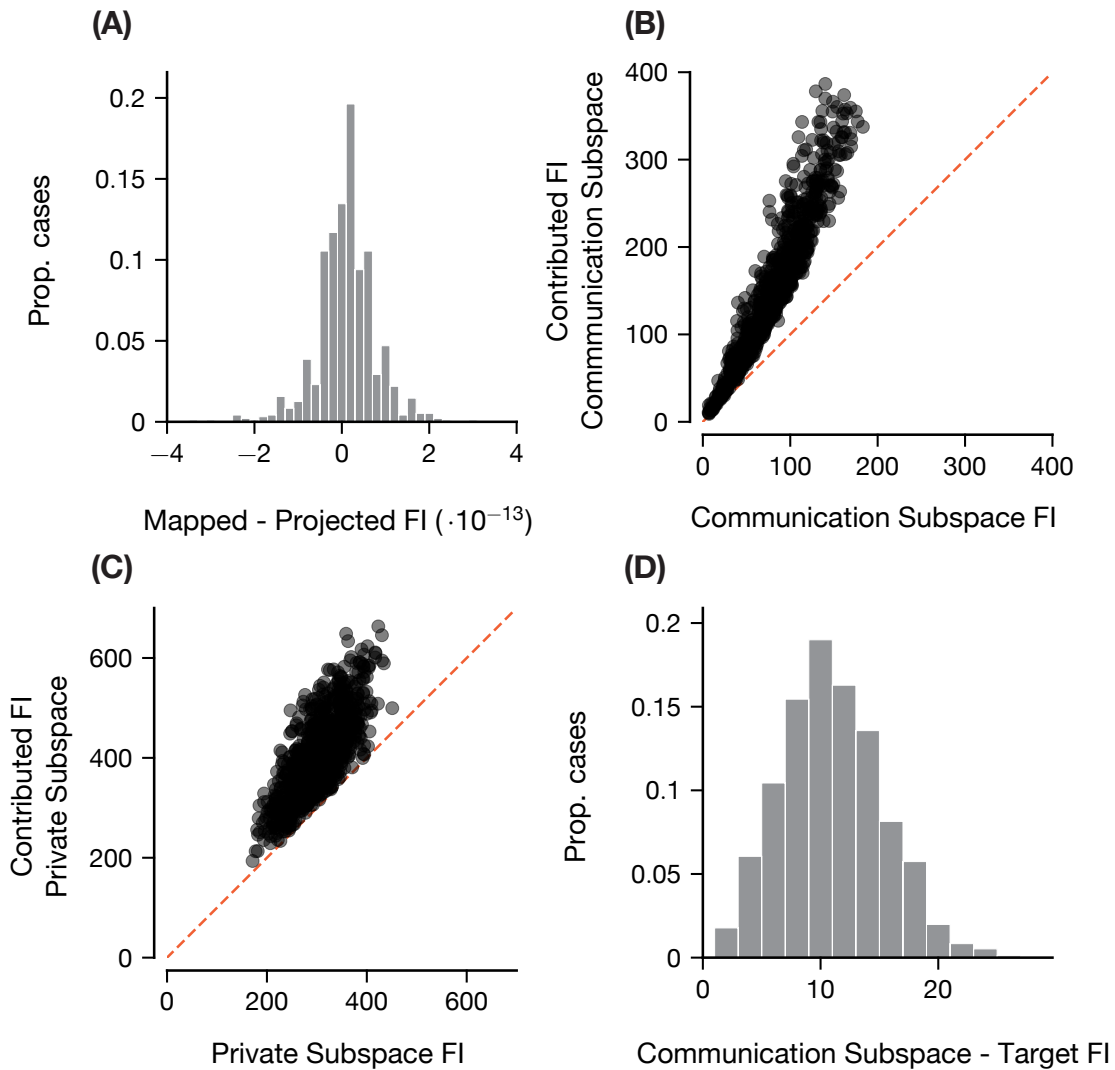communicated information impacts the target area.

Figure 2. Validation of Relationships between Fisher Information Theoretic Terms
Simulated population means and covariance matrices were generated as described
in Section 2.3. Fisher information theoretic terms were computed for each simulation. 1000
simulated mean vectors and noise covariance matrices were generated.
(A) Information in the communication subspace $\mathbf{J}_{cs}$ has a consistent definition (Eq 30):
$\mathbf{J}(\mathbf{B}_{cs}\mathbf{x}; \theta)$ (Eq 8) $= \mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta)$ (Eq 7)
(B) Contributed information to the communication subspace is an upper bound on
information in the communication subspace (Eq 16): $\mathbf{CI}_{cs}$ (Eq 14)$\geq \mathbf{J}_{cs}$
(C) Contributed information to the private subspace is an upper bound on information in the
private subspace (Eq 17): $\mathbf{CI}_{priv}$ (Eq 15) $\geq \mathbf{J}_{priv}$ (Eq 10)
(D) Information in the communication subspace is an upper bound on the impactful
information (information received from the source population, Eq 19):
$\mathbf{J}_{cs} \geq \mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$. Note that in these simulations the impactful information is
equivalent to the Target information because $\mathbf{r}_y = 0$.

12

The Fisher information decomposition in the source area (Eq 12) contains an additional term, $\mathbf{SI}_1$. As described in Section 2.2, this term measures the angle between the projections of the signal axis ($\mathrm{d}\mathbf{f}_x$) onto the communication and private subspace relative to principal axes of the source population activity space (i.e., the eigenvectors of Σ_x). Because of this, we claimed that this term accounts for how much information is shared between the communication and private subspaces, and we called it *shared information*. In particular, as the other terms in the source Fisher information decomposition (the contributed information terms $\mathbf{CI}_{cs}$ and $\mathbf{CI}_{priv}$) can be larger than the overall source Fisher information, the shared information term corrects for overcounting when combining these two terms. We tested this idea by measuring how $\mathbf{SI}_1$ changes with the size of the communication subspace (Fig 3A). For low- and high-dimensional communication subspaces (relative to the dimensionality of the source area), the magnitude of the cross-term is relatively small because there is less overlapping information between the communication and private subspaces due to their complementary sizes (e.g., when the communication subspace is low-dimensional the private subspace is high-dimensional, and vice versa). When the communication and private subspaces are of comparable dimensionality, the shared information is highly negative, which is due to the communication and private subspaces containing redundant information. We quantified this by comparing the shared information and the sum of Fisher information in the communication and private subspaces ($\mathbf{J}_{cs} + \mathbf{J}_{priv}$, Fig 3B). There is a strong negative correlation (-0.75), indicating that the more overall information in both the communication and private subspace, the more likely this information is not unique to these subspaces. This provides some evidence that the shared information term is responsible for measuring the extent to which the communication and private subspace have overlapping or redundant stimulus-related information.

In summary, we have derived mathematical equations for decomposing Fisher information in the source and target areas with respect to a low-rank linear map that describes the functional connectivity between these areas. We verified the relationships among the key quantities in this decomposition analytically and numerically (Fig 2). We then showed how 1) the impact on the target area (impactful information) does not simply depend on the communicated information, but also on the variability in the target area, with important consequences for the optimality of transmission and, as we show later, for selective gating at the target; 2) one term in the source Fisher information decomposition, the shared information $\mathbf{SI}_1$, describes the extent to which the communication and private subspaces contain the same information about the stimulus (Fig 3). These simulations validate our intuitions (detailed in Section 2.2) regarding the interpretation of the different components of the Fisher information decompositions. Next, we demonstrate a few scenarios in which our theory that can help us better understand information transmission across brain areas.
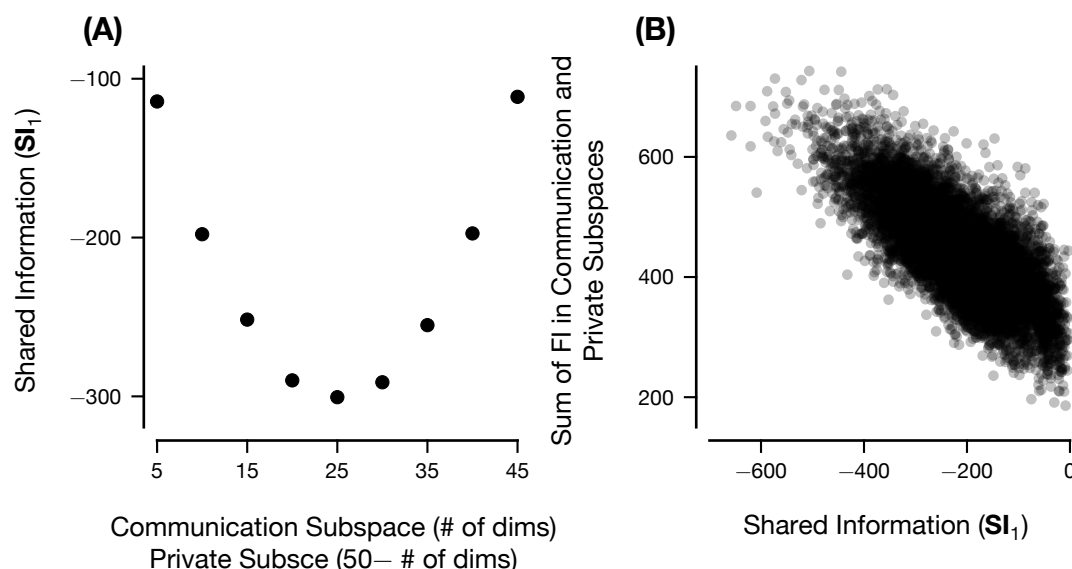
(A)

(B)

Figure 3. Shared Information between Communication and Private Subspace

(A) The relationship between the mean shared information term in the source area $\mathbf{SI}_1$ and the dimensionality of the communication subspace (or equivalently, 50 (the dimensionality of the source population) $-$ the dimensionality of the private subspace).

(B) How the magnitude of the shared information in the source area varies with the sum of Fisher information in the communication subspace and private subspace $\mathbf{J}_{cs} + \mathbf{J}_{priv}$.

Each point in (A) represents the average across 1000 simulated experiments with different tuning curves and covariance matrices. Dimension of the communication subspace was varied between 5 and 45. Each point in (B) represents a single simulation with fixed number of communication subspace dimensions between 5 and 45.

3.2 Selectively Altering Communication of Information by Modifying Tuning

3.2.1 Selectively Increasing Information in the Private Subspace

One of the scenarios presented in the original communication subspaces paper concerns the routing of information to two downstream cortical areas (Semedo, Zandvakili, et al. 2019). In this scenario, two downstream target populations (T_1 and T_2) receive input from a single source area (S) through two different reduced-rank regressions, corresponding to two different communication subspaces. How would the source area S be able to selectively communicate with downstream target area T_1 without affecting communication between S and T_2 and vice versa? Semedo, Zandvakili, et al. propose one solution: if the private subspace for communication between S and T_2 is contained in the communication subspace that exists between S and T_1 (symbolically, $\ker(S \to T_2) \subset \ker(S \to T_1)^\perp$), then fluctuations in activity in S that lie entirely in $\ker(S \to T_2)$ are communicated to T_1 but do not affect information transfer from $S \to T_2$. This alignment between the communication and private subspaces would allow the source area to represent multiple signals simultaneously and transmit different signals to distinct downstream areas (Jun et al. 2022; MacDowell et al. 2023).

We will formalize this observation using our Fisher information decomposition. Consider the case in which one source area communicates with one target area. First, we will demonstrate that one can increase the amount of information in the private subspace without affecting the communication between source and target. Then, conversely, we will consider how to increase information in the communication subspace without affecting information in the private subspace. This is a simplification of the scenario we previously presented, but maintains the core idea: if we can increase information in the private subspace without altering the information in the communication subspace or target area, then a different target area, whose communication subspace contains this private subspace, can receive this information.

To demonstrate this possibility, we first need to show that we can increase information in the private subspace without changing communication to the target area. Suppose a source area with tuning curve $\mathbf{f}_x$ is communicating with a target area via the low-rank linear mapping $\mathbf{B}_{\mathrm{cs}}$. Then we can decompose $\mathbf{f}_x = \mathbf{P}_{\mathrm{cs}}\mathbf{f}_x + \mathbf{Q}_{\mathrm{cs}}\mathbf{f}_x$, where $\mathbf{Q}_{\mathrm{cs}}$ is the projection onto $\ker(\mathbf{B}_{\mathrm{cs}})$ (see Section 2.2). Now, we will modify the tuning curve by increasing the projection of the tuning curve onto the private subspace:

$$\mathbf{f}_x \to \mathbf{f}_x^{\uparrow\mathrm{priv}} = \mathbf{P}_{\mathrm{cs}}\mathbf{f}_x + C\mathbf{Q}_{\mathrm{cs}}\mathbf{f}_x, \quad C > 1 \tag{22}$$

We will call the source activity with this new tuning curve $\mathbf{x}^{\uparrow\mathrm{priv}}$. To begin, we assume that the covariance matrix for the modified source activity is unchanged ($\mathbf{\Sigma}_x^{\uparrow\mathrm{priv}} = \mathbf{\Sigma}_x$).
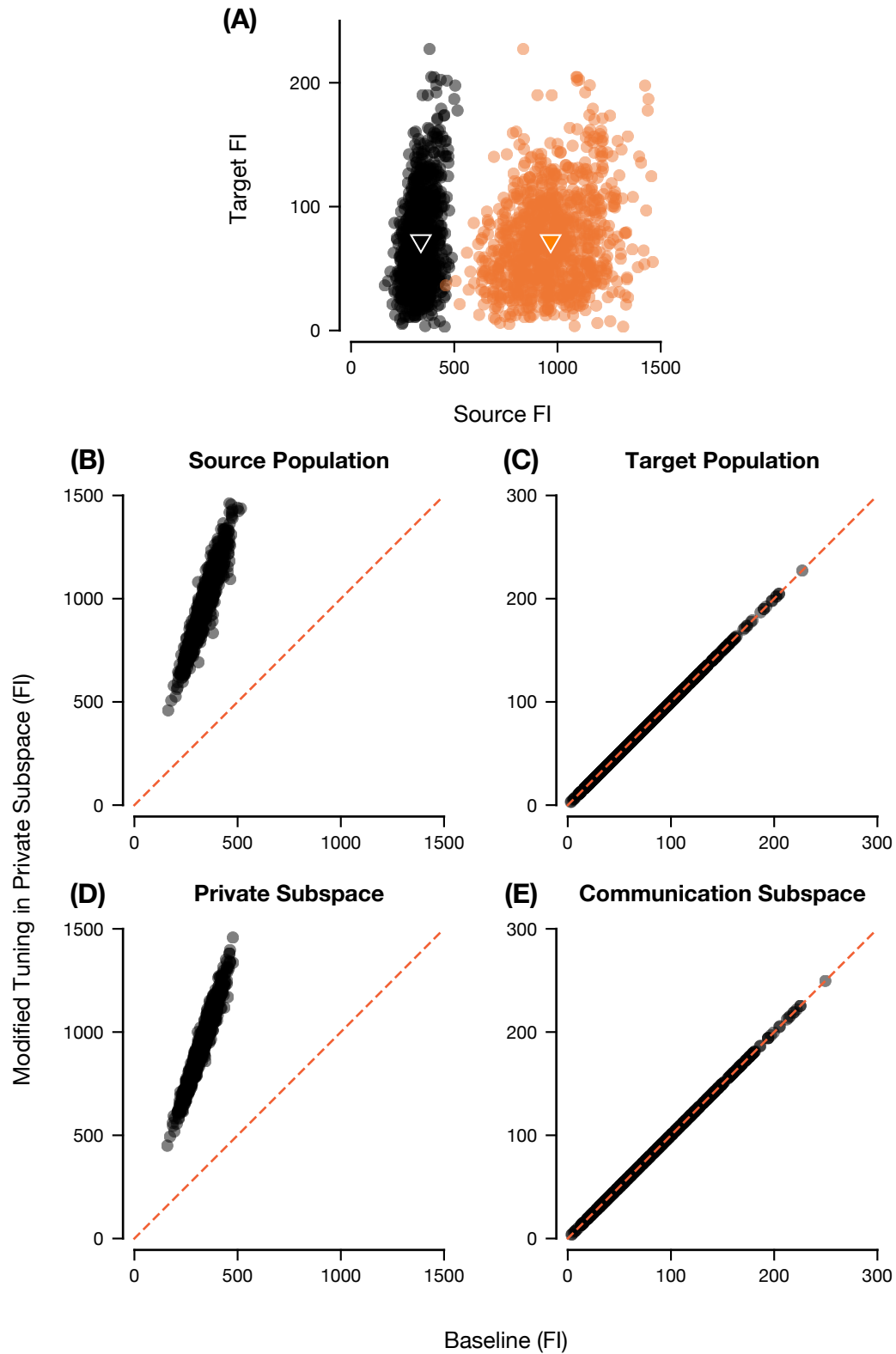
Figure 4 *(Caption on next page.)*

16

Figure 4. Changes in Fisher Information after Modifying Tuning in the Private Subspace

(A) Scatter plot of Fisher information (FI) in the source area vs. FI in the target area before (black) and after (orange) modifying the private tuning. The triangular marks indicate the means of the respective point clouds.

(B) Fisher information in the source area before and after modifying the private tuning. The x-axis corresponds to the abscissa of the black scatter points in (A), while the y-axis corresponds to the abscissa of the orange scatter points in (A).

(C) Fisher information in the target area before and after modifying the private tuning. The x-axis corresponds to the ordinate of the black scatter points in (A), while the y-axis corresponds to the ordinate of the orange scatter points in (A).

(D-E) Fisher information in the private subspace (D) and communication subspace (E) before and after modifying tuning in the private subspace.

1000 simulated tuning curves and covariance matrices were created. For each simulation, the tuning curve was modified using Eq 22 with $C = 1.5$, with each neuron (i.e., element of the tuning curve vector $\mathbf{f}_x$) having some additional random jitter for visualization purposes.

Fig 4 shows the results of 1000 simulations in which we increased the private tuning (Eq 22 with $C = 1.5$). We can see in Fig 4A that this modification causes the relationship between the source Fisher information and target Fisher information to shift primarily to the right (black is the baseline, orange is after modifying the tuning curve), indicating that we are increasing the source Fisher information without affecting the target Fisher information. This can be seen directly in the following panels in which we compare the source Fisher information (Fig 4B) and target Fisher information (Fig 4C) before and after increasing the private tuning. Clearly, after modifying the private tuning, the source Fisher information increases while the target Fisher information stays the same. In essence, the increase in the source Fisher information is because we are increasing the amount of information in the private subspace (Fig 4D), which we can think of as a component of the source Fisher information. Thus, increasing the information in the private subspace should increase the source Fisher information[2]. Additionally, as we are not changing the amount of information in the communication subspace (Fig 4E), there should be no change in the amount of impactful information $\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$, which for our simulations is exactly the amount of information in the target area. We also derived these effects mathematically by directly comparing the Fisher information terms (Appendix F).

As discussed in Section 2.3, we created the baseline source covariance matrix such that the variance would be equal to the mean. However, by modifying the source area tuning curve, we have broken this power-law relationship that is often observed in neural recordings. To address this, we can also modify the covariance matrix using the same procedure used to generate the source covariance matrix with the new tuning curve (Eq 22) and compare the

---

[2]Strictly speaking, the information in the private subspace is not a component of the source Fisher information; instead, the *contributed information* to the private subspace (Eq 15) is a component of the source Fisher information and the information in the private subspace is merely a lower bound for both the source Fisher information (Eq 11) and the contributed information (Eq 17). Nonetheless, we provide a mathematical justification for the fact that increasing the information in the private subspace tends to increase the information in the source area (Eq 39).
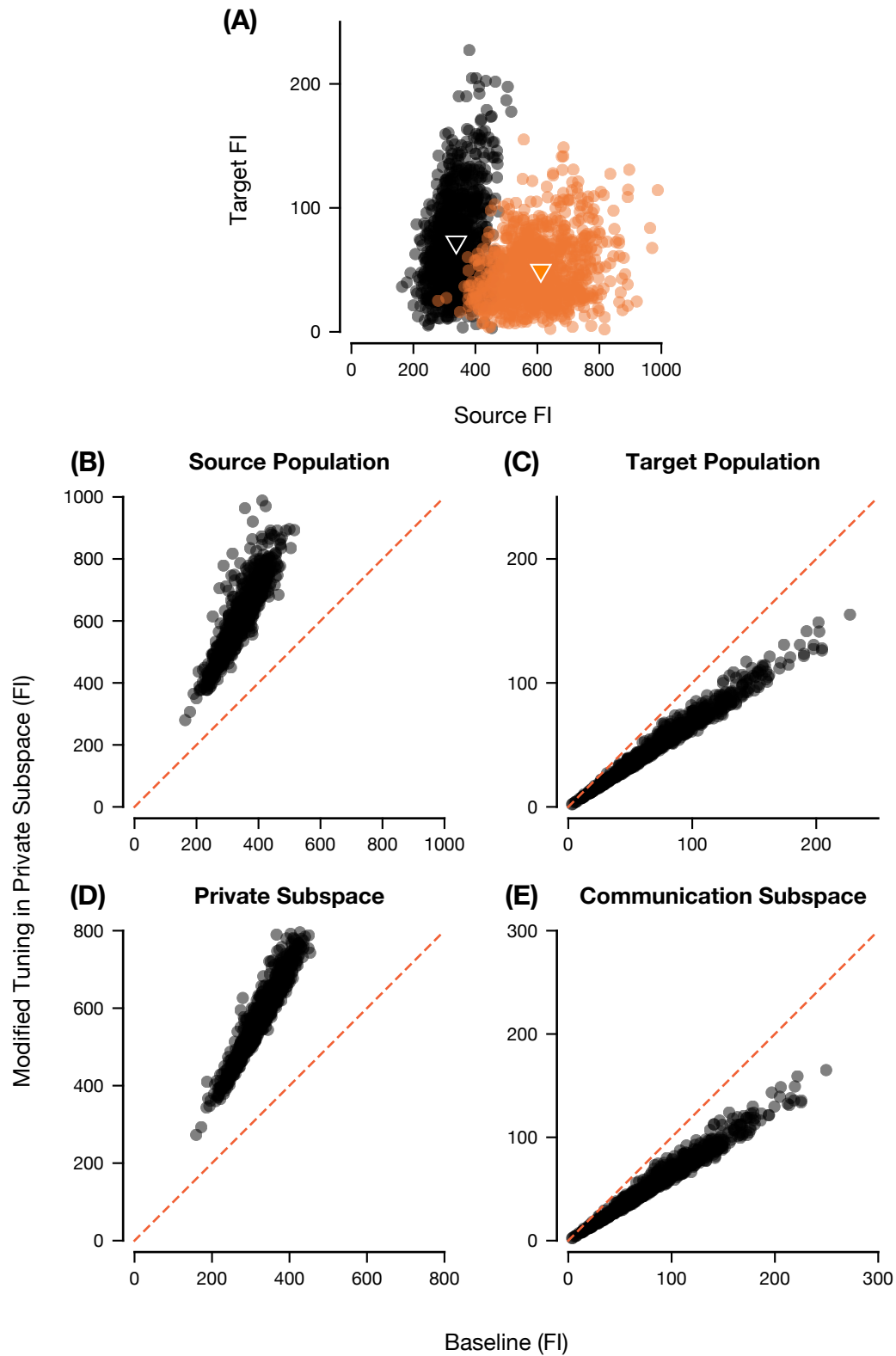
17

Figure 5 *(Caption on next page.)*

Figure 5. Changes in Fisher Information after Modifying Tuning and Covariance in the Private Subspace

Same as Fig 4 except we modify the covariance matrix so that the variance is equal to the mean.

Fisher information quantities for $\mathbf{x}$ and $\mathbf{x}^{\uparrow\mathrm{priv}}$ (Fig 5). In this case, many of the same conclusions apply numerically; however, the Fisher information in the communication subspace and the target area decrease (Fig 5C, E). This is likely occurring because, by increasing tuning in the private subspace, we tend to increase the average across-stimulus response of neurons in the source area, which increases the response variance. If this also causes the variability in the communication subspace to increase (i.e., if the eigenvalues of $\mathbf{P}_{\mathrm{cs}}\boldsymbol{\Sigma}_x^{\uparrow\mathrm{priv}}\mathbf{P}_{\mathrm{cs}}$ are greater than those for $\mathbf{P}_{\mathrm{cs}}\boldsymbol{\Sigma}_x\mathbf{P}_{\mathrm{cs}}$), then the Fisher information quantities computed from the covariance in the communication subspace will tend to decrease. Intuitively, since the Fisher information depends on the (pseudo)inverse of the covariance matrix (Eq 6), by increasing variability in this subspace we are "dividing" by a larger quantity, thus decreasing the information. This is not entirely true, however, as the effect will also depend on how the principal axes of activity in the communication subspace (i.e., the eigenvectors of $\mathbf{P}_{\mathrm{cs}}\boldsymbol{\Sigma}_x\mathbf{P}_{\mathrm{cs}}$) align with the projected signal axis $\mathbf{P}_{\mathrm{cs}}\mathbf{df}_x$ (Zylberberg et al. 2017, Eq 71) and the size of the subspace being modified; it is possible for the Fisher information to increase after this modification (see Fig 8D and the explanation related to Fig 7).

In summary, the analysis above demonstrates that stimulus information can be increased selectively in one area without affecting all its targets, by changing tuning in the private subspace (Fig 4). However, it also cautions that if the tuning modification is accompanied by changes in noise covariance, then transmission to the target may be impaired (Fig 5), which is an important consideration for realistic circuit models of selective communication (see Section 4).

3.2.2 Selectively Increasing Information in the Communication Subspace

We can use a similar method as in Section 3.2.1 to increase the strength of communication between a source and target area while not altering the amount of information in the private subspace. To do so, we replace the tuning curve with one in which we increase the contribution of the projection onto the communication subspace:

$$\mathbf{f}_x \rightarrow \mathbf{f}_x^{\uparrow\mathrm{comm}} = C\mathbf{P}_{\mathrm{cs}}\mathbf{f}_x + \mathbf{Q}_{\mathrm{cs}}\mathbf{f}_x, \quad C > 1 \tag{23}$$

We will refer to the source population activity with this modified tuning curve as $\mathbf{x}^{\uparrow\mathrm{comm}}$ (again note the covariance matrix will initially be unchanged).
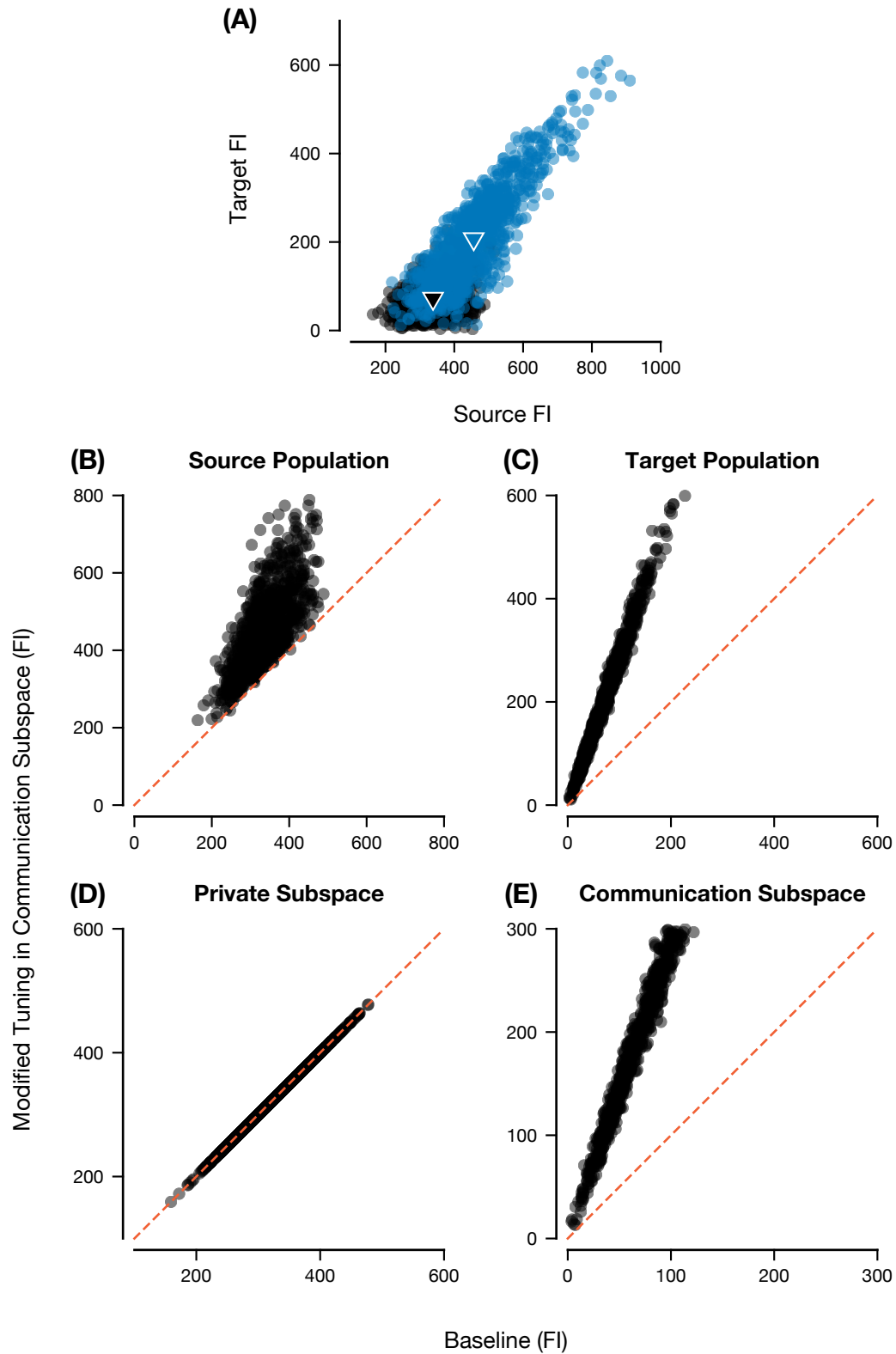
19

Figure 6 *(Caption on next page.)*

Figure 6. Changes in Fisher Information after Modifying Tuning in Communication Subspace

(A) Scatter plot of Fisher information (FI) in the source area vs. FI in the target area before (black) and after (blue) modifying the tuning in the communication subspace (blue). The triangular marks indicate the means of the respective point clouds.

(B) Fisher information in the source area before and after modifying the tuning in the communication subspace. The x-axis corresponds to the abscissa of the black scatter points in (A), while the y-axis corresponds to the abscissa of the blue scatter points in (A).

(C) Fisher information in the target area before and after modifying the tuning in the communication subspace. The x-axis corresponds to the ordinate of the black scatter points in (A), while the y-axis corresponds to the ordinate of the blue scatter points in (A).

(D-E) Fisher information in the private subspace (D) and communication subspace (E) before and after modifying tuning in the communication subspace.

1000 simulated tuning curves and covariance matrices were created. For each simulation, the tuning curve was modified using Eq 23 with $C = 1.5$, with each neuron (i.e., element of the tuning curve vector $\mathbf{f}_x$) having some additional random jitter for visualization purposes.

Fig 6 is analogous to Fig 4, except we modified tuning in the communication subspace instead of modifying the private tuning (Eq 23 with $C = 1.5$). Before modification, the baseline tuning curves and covariance matrices are the same in Figs 4 and 6. Looking at Fig 6A and comparing the positions of the means (triangular marks) of the point clouds for both the baseline (black) and modified tuning in the communication subspace (blue), it is clear that this modification to the communication subspace causes the relationship between source and target Fisher information to shift up and to the right, indicating an increase in both the source and target areas. The increase in Fisher information in the source and target populations is due to the increase of Fisher information in the communication subspace (Fig 6E, but again note the same caveat applies here as when we considered increasing information in the private subspace). In contrast, the Fisher information in the private subspace does not change (Fig 6D). Mathematically, these effects can all be explained similarly to the change of tuning in the private subspace (Appendix F).

We also noted that the increase in source area (Fig 6B) is much more modest, especially in comparison to the increase in source area Fisher information when increasing private tuning (Fig 4B). The reason for this is due to the difference in dimensionality between the communication subspace and the private subspace. In our simulations (and in neural data (Semedo, Zandvakili, et al. 2019)), the dimensionality of the communication subspace is significantly less than the dimensionality of the neural population space (the dimension of the communication subspace in our simulations is 5, while the number of neurons in the source population is 50). For this reason, when we increase tuning in the private subspace, it is approximately equal to scaling the whole population mean tuning curve, which causes a noticeable effect on the amount of Fisher information in the source area (and Fisher information contained in the private subspace, Fig 4B and 4D). Conversely, increasing tuning in the communication subspace modifies a small portion of activity in the source area, so the effect of Fisher information in the source area is not as large (Fig 6B and notice the smaller magnitude of information in the communication subspace, Fig 6C). We can see this in Fig 7, in which we increase the dimensionality of the communication subspace and
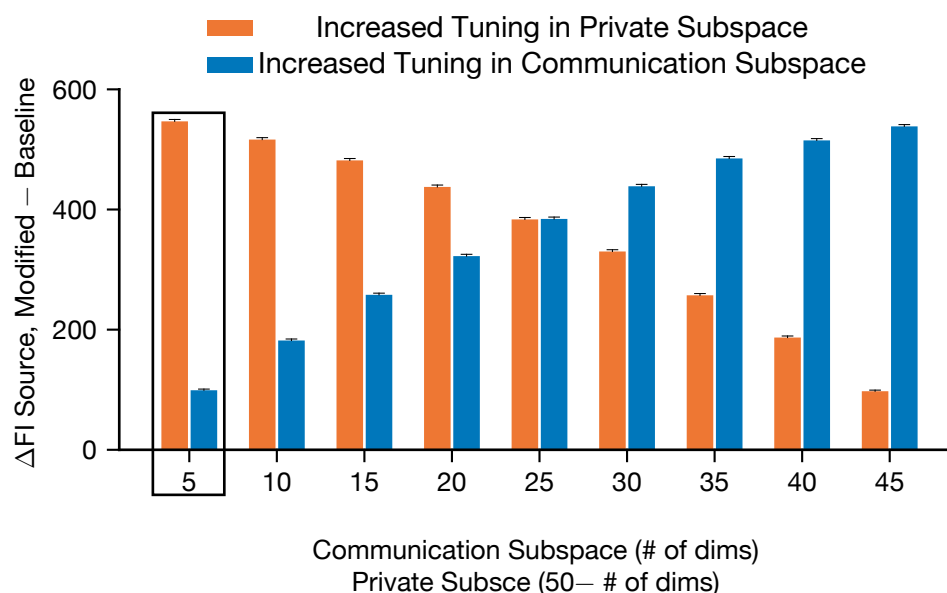
21

Figure 7. Increase in Fisher Information in Source Area after Modifying Tuning Scales with Size of Subspace

Difference between Fisher information in the source area with increased tuning in either the private subspace (orange) or communication subspace (blue) to Fisher information in the source area at baseline ($\mathbf{J}(\mathbf{P}_{cs}\mathbf{x}^{\uparrow \mathrm{comm\ or\ priv}}; \theta) - \mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta)$). Each bar represents the average difference across 1000 simulations with the dimensionality of the communication subspace fixed for a value between 5 (which corresponds to the simulations in Figs 4 and 5 (orange bar), and Figs 6 and 8 (blue bar)) and 45. By construction, the dimensionality of the private subspace for each simulation is 50 (the number of neurons in the source area) $-$ the dimensionality of the communication subspace. As with Fig 6, the tuning curves were modified using Eq 23 with $C = 1.5$ with some per-neuron jitter. Error bars are the standard error of the mean difference.

measure how much the Fisher information in the source area increases when we modify tuning in the communication (blue bars) or private (orange bars) subspace. Clearly, as the dimensionality of the communication subspace increases, the Fisher information in the source area after modifying tuning in the communication subspace increases (relative to the unmodified baseline), while the same monotonic effect is observed for modifying tuning in the private subspace. This also explains the difference in scale between the source and target Fisher information quantities in general, as only the stimulus-related changes lying in the low-dimensional communication subspace are transmitted to the target area (Figs 4A and 6A).

If we modify both tuning curves and covariance covariance matrix to maintain the power-law relationship between the variance and the mean, we find qualitatively similar effects (Fig 8). The only noticeable difference is for the Fisher information in the private subspace; for some simulations, the Fisher information increases after modifying the tuning curve in the communication subspace and the covariance matrix, while for others it decreases (Fig 8D). However, for the vast majority of cases, these differences were not significant (934/1000 simulations had absolute value of the z-scored difference ≤ 1.96). The reason there is a stronger effect on the Fisher information in the communication subspace after increasing private tuning (Fig 5E) than for the Fisher information in the private subspace here (Fig 8D) is likely also due to the dimensionality of these subspaces in our simulations. As explained previously, for our simulation, increasing tuning in the private subspace should nearly increase tuning for the whole space, so the variance of all neurons in the source area will increase. Increasing variability in all directions will tend to also increase variability in the communication subspace, so the Fisher information should decrease. On the other hand, because the effect of increasing tuning in the communication subspace does not greatly effect the tuning curve magnitude, the covariance matrix will not change substantially.
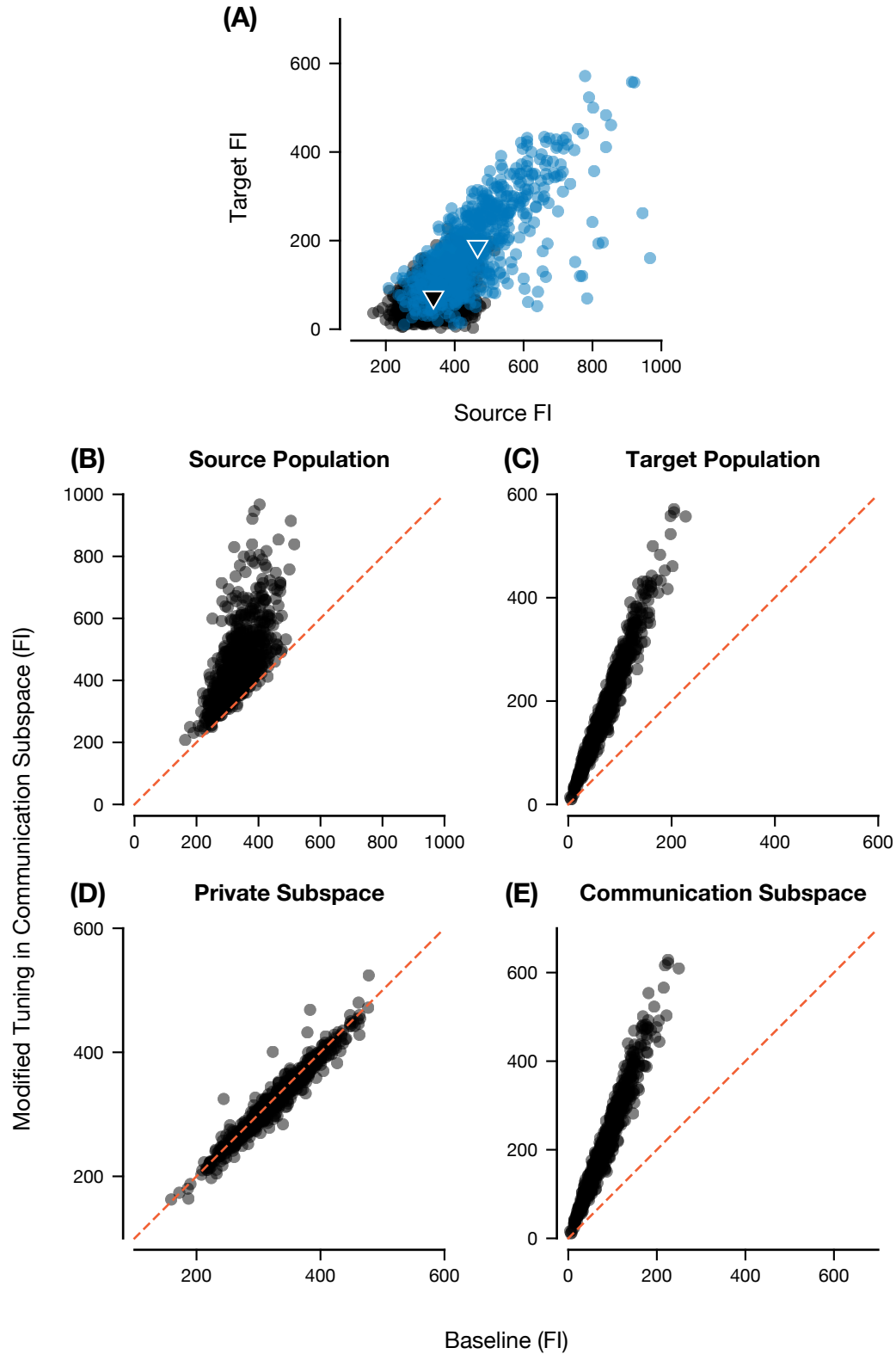
Figure 8 *(Caption on next page.)*

Figure 8. Changes in Fisher Information after Modifying Tuning and Covariance in the Communication Subspace

Same as Fig 6 except we modify the covariance matrix so that the variance is equal to the mean.

In summary, we have shown how we can selectively alter the amount of stimulus-related information communicated between the source and target areas by modifying the amount of mean stimulus tuning contributed by the private subspace (Figs 4 and 5) or communication subspace (Figs 6 and 8). We also found that the effect of these modifications on Fisher information scales with the dimensionality of the subspace where we increase tuning (Fig 7). These results demonstrate how functional networks in the brain might exploit the flexibility afforded by communication subspaces (Semedo, Zandvakili, et al. 2019) to flexibly route stimulus information between different downstream target areas.

3.3   Modifying Communication by Changing Target Covariance Structure

In this previous section, we investigated how to modulate communication of stimulus information by altering activity in the source area. Now, we will consider how changing activity in the target area can affect the strength of communication, which is referred to as input signal gating or selective reception (Gisiger and Boukadoum 2011; Pesaran et al. 2021; Barbosa et al. 2023). This is also related to multisensory integration, where different upstream source areas represent separate sensory modalities (e.g., vision and hearing), and a single downstream target must selectively combine information from each area (Angelaki et al. 2009). This integration of evidence requires the target area to weigh information from the upstream source areas in different ways depending on task demands (e.g., there will be scenarios when the target area will "ignore" information coming from a particular source area). Since we assumed that the target area has no residual tuning, we can only modify activity in the target area by changing the residual noise covariance or the structure of the reduced-rank regression matrix. Because changing the reduced-rank regression matrix will also change the communication and private subspaces, we will leave this fixed and focus on how changing the noise covariance in the target area can affect communication. To measure this, we will examine the relationship between the Fisher information in the communication subspace ($\mathbf{J}_{cs}$) and the impactful information in the target area ($\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$), which in our simulations is equivalent to the total target Fisher information (see Section 2.3). As we have previously shown, the Fisher information in the communication subspace is an upper bound for the impactful information in the target area (Section 2.2.2 and AppendixE). We can interpret the tightness of this upper bound (i.e., how close $\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$ is to $\mathbf{J}_{cs}$) as a measure of how information is being read out from the target area.

One way to weaken communication between the source and target area is to increase the noise variance in the target area (Fig 9B). In Section 2.1, we discussed how the residual variance in the target area is the quantity that is being minimized (i.e., the regression matrix is chosen such that it minimizes prediction error, which is the unexplained variance (Izenman

2008, Eq 6.78)). By increasing the variance in the target area, we are, in effect, decreasing the predictive power of the regression, such that activity patterns in the target are no longer well described as linear combinations of activity patterns in the communication subspace as they become more "random" from trial-to-trial (compare Fig 9A and Fig 9B). This decreases the amount of information decodable from the target population, explaining why the Fisher information in the target area decreases, thus "weakening" the communication between the source and target.
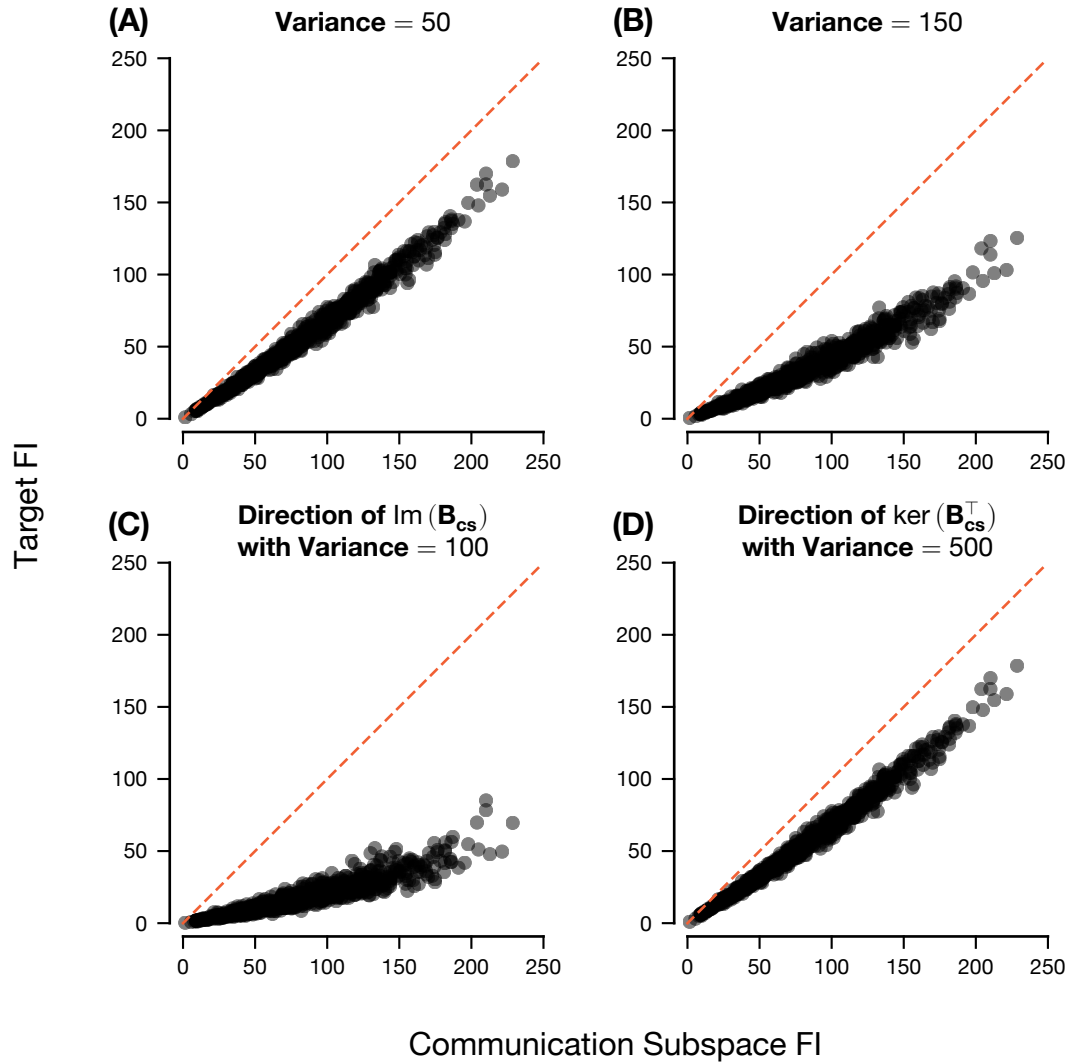
Figure 9. Changing Noise Covariance in Target Area can Selectively Alter Communication

Plotting the Fisher information (FI) in the communication subspace $\mathbf{J}_{cs}$ vs. the Fisher information in the target population, which for our simulations is equal to $\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$. Note that $\mathbf{J}_{cs}$ is the same across all panels (see Eq 7). Each panel shows this relationship with a different target covariance structure $\boldsymbol{\Sigma}_y = \mathbf{B}_{cs}\boldsymbol{\Sigma}_x\mathbf{B}_{cs}^\top + \boldsymbol{\Sigma}_r$ (Eq 4, with $\boldsymbol{\Sigma}_r = \sigma_2^2\mathbf{I}$). In each case, the "inherited covariance matrix" $\mathbf{B}_{cs}\boldsymbol{\Sigma}_x\mathbf{B}_{cs}^\top$ is left unchanged. Thus, each panel will have a different form of the matrix $\boldsymbol{\Sigma}_r$.

(A) Baseline simulation with $\boldsymbol{\Sigma}_r = 50\mathbf{I}$.
(B) Simulation with $\boldsymbol{\Sigma}_r = 150\mathbf{I}$.
(C) Simulation with $\boldsymbol{\Sigma}_r = 50\mathbf{I} + 100\,\mathrm{Corr}\left(\mathrm{Im}(\mathbf{B}_{cs})\right)$. To do so, we found an orthonormal basis for the column space of $\mathbf{B}_{cs}$, formed a projection matrix from this basis, converted this projection matrix to a correlation matrix and multiplied by 100.
(D) Simulation with $\boldsymbol{\Sigma}_r = 50\mathbf{I} + 500\,\mathrm{Corr}\left(\mathrm{ker}(\mathbf{B}_{cs}^\top)\right)$. We used the same technique as in (C) but for the null space of $\mathbf{B}_{cs}^\top$.

However, this modification increases variability uniformly in every direction in the

target area. In the multisensory integration scenario, in which there are two source areas, say S_1 and S_2, communicating via low-rank linear maps $\mathbf{B}_1, \mathbf{B}_2$ to one target area T, this would decrease information transfer from both areas. How would we modify the target area covariance to decrease communication between S_1 and T without affecting the information transmitted from S_2 to T? Let's assume that $\mathbf{B}_1, \mathbf{B}_2$ are sufficiently low-rank such that their column spaces are linearly independent (i.e., $\text{Im}(\mathbf{B}_1) \cap \text{Im}(\mathbf{B}_2) = \{0\}$). In this case, we can introduce noise correlations along the axes of $\text{Im}(\mathbf{B}_1)$. Intuitively, much like increasing the variance in all directions equally (Fig 9B), increasing variability in the direction that contains the mapped activity from one source ($\text{Im}(\mathbf{B}_1)$) will decrease the extent to which the low-rank map determines activity in the target area (i.e., reduces the predictive power of $\mathbf{B}_1$). However, by selectively increasing the variability in this direction, any direction orthogonal to this subspace (which includes $\text{Im}(\mathbf{B}_2)$), should have no change in variability. In other words, the predictive power of the reduced-rank regression from the other source area ($\mathbf{B}_2$) should be left unchanged.

We can reduce this to a scenario with one source area and one target area (with corresponding map $\mathbf{B}_{cs}$) while maintaining the same core idea: if we modify the noise covariance of the target area by introducing correlations in the direction of $\text{Im}(\mathbf{B}_{cs})$ (which we call the impactful direction), we can decrease the information transfer from the source to the target, but if we modify the covariance in the orthogonal direction ($\text{Im}(\mathbf{B}_{cs})^\perp = \ker(\mathbf{B}_{cs}^\top)$, Axler 2024, Proposition 7.6), then this should not affect the information transfer. As predicted, adding noise covariance in the target area that lie along the subspace $\text{Im}(\mathbf{B}_{cs})$ decreases the impactful information in the target area relative to the Fisher information in the communication subspace (Fig 9C). Notably, this effect is more pronounced here than for the case of increasing the variability equally in all directions (Fig 9B) even though the residual covariance matrices have the same overall variance. This occurs because we are "allocating" proportionally more of the variability to the impactful directions; if we increase variability in all directions, the overall effect will depend on how this variance is projected onto the impactful directions. On the other hand, if we increase variability in the non-impactful directions (i.e., the orthogonal complement of $\text{Im}(\mathbf{B}_{cs})$), then we see that information transmission is left unchanged, even if there is substantially stronger noise in this direction (Fig 9D).

In summary, we have shown through our Fisher information decomposition that we can modify the target activity to "silence" communication between the source and target area by introducing noise correlations in the subspace of the target area which is well predicted by the source area ($\text{Im}(\mathbf{B}_{cs})$, Fig 9C). We also showed that this effect is selective: increasing variability in any orthogonal direction does not influence how much information the target area inherits from the source area (Fig 9D).

4   Discussion

We have introduced a method for decomposing the stimulus-related information in cortical areas that communicate through a low-rank linear mapping (Fig 1A). We provided theoretical proofs (see Section 2.2) and numerical analyses (Figs 2 and 3) that suggest how we can interpret the individual terms (see Table 1 for descriptions of these terms) in the

Fisher information decompositions. Through numerical simulations, we illustrated how to levrage communication subspaces for flexible stimulus-information routing and gating. We demonstrated how one can selectively increase the amount of information in the source area without affecting information in the target area by increasing the information in the private subspace (Figs 4 and 5). Additionally, we showed how to strengthen communication between areas by increasing information in the communication subspace (Figs 6 and 8), and that the magnitude of this effect depends on the dimensionality of the communication subspace (Fig 7). Lastly, we showed how the structure of the covariance in the target area can also influence communication by changing how much impact the information in the source area has on the target area (Fig 9).

Although we demonstrated our framework with simplified simulations using synthetic tuning and covariance, the insights gained here can guide future work on information transmission in realistic networks. For instance, we assumed no residual tuning in the target area, therefore we did not study the residual information **RI** and the synergistic/redundant information $\mathbf{SI}_2$. We chose not to investigate these terms to focus on several examples of how the concept of communication subspaces can be used to explain flexible information routing and signal gating. However, in experimental data, there will likely be residual tuning, which can arise from, for example, unrecorded source neurons feeding information to the target population. Future work could study these quantities by using more realistic models of activity in a source and target area (e.g., V1 and V2), for instance using hierarchical network models that reproduces the firing rate statistics of communicating neural populations (e.g., Perich and Rajan 2020; Perich, Arlt, et al. 2021; Barbosa et al. 2023; Mastrogiuseppe and Ostojic 2018; Hahn et al. 2019; Huang, Pouget, et al. 2022). Our results of Figs 4 and 6 offer concrete demonstrations of how joint changes in tuning and covariance–which are unavoidable in realistic dynamic recurrent circuit models (e.g., Seriès et al. 2004))–can impact stimulus information transmission. The framework proposed here can thus be leveraged to study quantitatively how realistic network models implement flexible or selective communication.

Despite much work on functional connectiviy, only few studies have investigated how stimulus-related information propagates between cortical areas. Panzeri and colleagues have developed a new information-theoretic measure called Feature-specific Information Transfer (FIT) to quantify stimulus-related information transfer (Celotto et al. 2023; Pica et al. 2019). Their method is based on mutual information rather than Fisher information. Although these quantities are related (Brunel and Nadal 1998; Wei and Stocker 2016), computing mutual information from neural population data is difficult due to the number of trials needed to measure the empirical multivariate probability distributions involved (Timme and Lapish 2018; Huang and Zhang 2018). As such, the FIT measure has only been applied to pairs of 1-dimensional signals (e.g. two channels in EEG recordings). In contrast, our framework has potential applicability to large population data. because, to estimate the terms in our Fisher information decompositions, one can leverage 1) techniques to efficiently compute the reduced-rank regression coefficient matrix and the associated projections (Anderson 1999; Giraud 2011; Negahban and Wainwright 2011; Hua et al. 2001); and 2) existing efficient estimators of Fisher information such as analytical bias-corrected estimation (Kanitscheider et al. 2015) and linear decoding (Kohn, Coen-Cagli, et al. 2016). As explained in Section 2.2.3, although decoding can only be applied straightforwardly to some terms in our decomposition,

all the other terms can still be seen as the Fisher information for transformed measures of neural population activity (e.g., $\mathbf{CI}_{cs}$ is the linear Fisher information that corresponds to decoding the stimulus identity from $\mathbf{x} - \mathbf{Q}_{cs}\mathbf{f}_x$).

The study most similar to ours examined the Fisher information in the output layer of a two-layer neural network (Zylberberg et al. 2017; see also Renart and van Rossum 2012). There are a few subtle differences between their work and our contribution. On the one hand, they examined how a static linear or nonlinear transformation of the output layer (i.e., the target population) changes how stimulus-related information can be decoded, while our work focuses exclusively on linear transformations. On the other hand, their analysis requires a full-rank feedforward connectivity matrix, whereas we consider low-rank connectivity. The most crucial difference, however, is the difference in perspectives that drove our analyses. Their primary concern was to understand the relationship between the capacity of a population to represent information and the ability of that population to send information to another population. Different from that work, we were not focused on what source population covariances maximize information transfer; instead, we investigated the functional implications of communication subspaces for the transmission of information (but see below for a possible extension to our framework that would consider how to optimize information transfer through a communication subspace). Nonetheless, both studies are ultimately interested in the geometry of noise correlations in neural populations and how this affects information propagation. A better understanding of this interaction through the lens of communication subspaces, possibly by combining our information decompositions with the insights of Zylberberg et al., is vital for future work.

One limitation of our proposed information decomposition is the assumption of linearity of predictability, which, in turn, leads to our analyses being based on linear geometry in Euclidean spaces. Recent work has suggested that neural population activity may lie on a "neural manifold" (Chung and Abbott 2021; Gallego et al. 2017) that can be highly nonlinear (De and Chaudhuri 2023; Fortunato et al. 2023; Golden et al. 2016; Whiteway et al. 2019). Additionally, interareal interactions may also be nonlinear (Lyamzin et al. 2015; Ruff and Cohen 2017). If these nonlinearities are significant, the partitioning of stimulus-related information we propose could significantly underestimate information transmitted between areas. Indeed, a recent simulation study demonstrated a two-layer spatial network connected via a nonlinear mapping with poor communication (measured through predictive performance of reduced-rank regression of the second layer on the first layer) even though activity in the target area is entrained to activity in the source area (Gozel and Doiron 2022). Although this is a concern, our method is still applicable even when the "true" mapping between areas is nonlinear, as long as we correctly interpret the meaning of linear Fisher information as (roughly) being the portion of the full Fisher information that is decodable through a linear estimator (see Section 2.2 for details) and if we consider a sufficiently small change in the stimulus value such that the interareal transformation is approximately linear. We can also consider possible extensions or alternatives to address nonlinear mappings between areas. The most obvious would be to combine the method proposed here with previously mentioned research on information propagation through a simple nonlinear network in which a linear weight matrix is composed with a static nonlinearity (Zylberberg et al. 2017). In such a scenario, the information in the image of such a linear-nonlinear mapping (equivalent to Eq 8) would be different than the information in the communication subspace (Eq 7),

30

whereas for the linear mapping they are equal (Eq 30).

Ongoing research on communication subspaces indicate other possible extensions to our work. The communication subspaces framework can be seen as a kind of latent variable model (LVM) in which there are two groups of latent variables (Fig 1A): $\mathbf{z}_{cs} = \mathbf{P}_{cs}\mathbf{x}$, which are the latent variables that generate activity in the communication subspace, and $\mathbf{z}_{priv} = \mathbf{Q}_{cs}\mathbf{x}$, which are the latent variables that generate activity in the private subspace (Murphy 2021, Section 20.2). Through this interpretation, recent work has pushed the concept of communication subspaces beyond undirected, two-area communication to allow for feedforward and feedback signaling (Semedo, Jasper, et al. 2022; Gokcen, Jasper, Semedo, et al. 2022) and communication among multiple cortical areas (Gokcen, Jasper, Xu, et al. 2024). The introduction of more complicated partitioning of activity into projections onto subspaces defined by multi-area shared latent variables should be a straightforward extension of the methods we have developed here. Another kind of LVM to consider using in place of reduced-rank regression would be sufficient dimension reduction (SDR). Unlike reduced-rank regression, SDR looks for a linear map that captures the maximal amount of statistical dependency between areas (Cook 2018). By using SDR methods (e.g., Cook et al. 2015), we could find analogs of the communication and private subspaces that have minimal overlapping information, which would have some interpretational advantages over communication subspaces defined through standard reduced-rank regression (Semedo, Gokcen, et al. 2020).

An important application of our framework is the optimization of linear Fisher Information transfer. While there has been extensive work on maximizing the representation of sensory information (e.g., Brenner et al. 2000; Quian Quiroga and Panzeri 2009; Shew et al. 2011; Wang et al. 2016), research on optimally transmitting sensory information between cortical areas is in its infancy (Zylberberg et al. 2017; Ebrahimi et al. 2022; Barreiro et al. 2024; Manning et al. 2024). Our framework allows us to distinguish between communicated or received information (i.e., information in the communication subspace, Eq 7) and how this information impacts encoding in the target population ($\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$). Finding the communication subspace that maximizes the communicated information is essentially equivalent to finding the optimal unbiased linear decoder, implying that other axes in the source population activity space are less informative and redundant. However, this axis does not necessarily optimize how the target population "reads" the information in the communication subspace, which will depend on how the transmitted signal axis ($\mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x$) aligns with the target covariance matrix. This can allow for directions in the source population activity space that are not the optimal decoding axis to be more impactful on stimulus-related information in the target area. Investigating the structures of the functional maps and communication subspaces that maximize the impactful information in the target area could provide further insights into interareal communication.

5   Acknowledgments

References

Abdeladim, L, H Shin, UK Jagadisan, MB Ogando, and H Adesnik (2023). *Probing Inter-Areal Computations with a Cellular Resolution Two-Photon Holographic Mesoscope*. Pre-published.

Anderson, TW (1999). "Asymptotic Theory for Canonical Correlation Analysis". In: *Journal of Multivariate Analysis* 70.1, pp. 1–29.

Angelaki, DE, Y Gu, and GC DeAngelis (2009). "Multisensory Integration: Psychophysics, Neurophysiology, and Computation". In: *Current Opinion in Neurobiology*. Sensory Systems 19.4, pp. 452–458.

Averbeck, BB, PE Latham, and A Pouget (2006). "Neural Correlations, Population Coding and Computation". In: *Nature Reviews Neuroscience* 7.5, p. 358.

Axler, S (2024). *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Cham: Springer International Publishing.

Barbosa, J, R Proville, CC Rodgers, MR DeWeese, S Ostojic, and Y Boubenec (2023). "Early Selection of Task-Relevant Features through Population Gating". In: *Nature Communications* 14.1, p. 6837.

Barreiro, AK, AJ Fontenele, C Ly, PC Raju, SH Gautam, and WL Shew (2024). "Sensory Input to Cortex Encoded on Low-Dimensional Periphery-Correlated Subspaces". In: *PNAS Nexus*, pgae010.

Beck, J, VR Bejjanki, and A Pouget (2011). "Insights from a Simple Expression for Linear Fisher Information in a Recurrently Connected Population of Spiking Neurons". In: *Neural Computation* 23.6, pp. 1484–1502.

Brenner, N, W Bialek, and RdR van Steveninck (2000). "Adaptive Rescaling Maximizes Information Transmission". In: *Neuron* 26.3, pp. 695–702. pmid: 10896164.

Brunel, N and JP Nadal (1998). "Mutual Information, Fisher Information, and Population Coding". In: *Neural Computation* 10.7, pp. 1731–1757.

Celotto, M et al. (2023). "An Information-Theoretic Quantification of the Content of Communication between Brain Regions". In: *Advances in Neural Information Processing Systems*. Ed. by A Oh, T Naumann, A Globerson, K Saenko, M Hardt, and S Levine. Vol. 36. Curran Associates, Inc., pp. 64213–64265.

Chung, S and LF Abbott (2021). "Neural Population Geometry: An Approach for Understanding Biological and Artificial Neural Networks". In: *Current Opinion in Neurobiology*. Computational Neuroscience 70, pp. 137–144.

Cohen, MR and A Kohn (2011). "Measuring and Interpreting Neuronal Correlations". In: *Nature Neuroscience* 14.7, pp. 811–819.

Cook, RD (2018). "Principal Components, Sufficient Dimension Reduction, and Envelopes". In: *Annual Review of Statistics and Its Application* 5.1, pp. 533–559.

Cook, RD, L Forzani, and X Zhang (2015). "Envelopes and Reduced-Rank Regression". In: *Biometrika* 102.2, pp. 439–456. JSTOR: 43908546.

Davies, PT and MKS Tso (1982). "Procedures for Reduced-Rank Regression". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31.3, pp. 244–255. JSTOR: 2347998.

De, A and R Chaudhuri (2023). "Common Population Codes Produce Extremely Nonlinear Neural Manifolds". In: *Proceedings of the National Academy of Sciences* 120.39, e2305853120.

DiCarlo, JJ, D Zoccolan, and NC Rust (2012). "How Does the Brain Solve Visual Object Recognition?" In: *Neuron* 73.3, pp. 415–434.

Ebrahimi, S, J Lecoq, O Rumyantsev, T Tasci, Y Zhang, C Irimia, J Li, S Ganguli, and MJ Schnitzer (2022). "Emergent Reliability in Sensory Cortical Coding and Inter-Area Communication". In: *Nature* 605.7911 (7911), pp. 713–721.

Ecker, AS, P Berens, AS Tolias, and M Bethge (2011). "The Effect of Noise Correlations in Populations of Diversely Tuned Neurons". In: *Journal of Neuroscience* 31.40, pp. 14272–14283.

Fortunato, C, J Bennasar-Vázquez, J Park, JC Chang, LE Miller, JT Dudman, MG Perich, and JA Gallego (2023). "Nonlinear Manifolds Underlie Neural Population Activity during Behaviour". In: *bioRxiv*, p. 2023.07.18.549575. pmid: 37503015.

Gallego, JA, MG Perich, LE Miller, and SA Solla (2017). "Neural Manifolds for the Control of Movement". In: *Neuron* 94.5, pp. 978–984.

Gallier, J (2011). *Geometric Methods and Applications: For Computer Science and Engineering*. Springer New York. 680 pp. Google Books: USC8NAEACAAJ.

Geisler, WS and DG Albrecht (1997). "Visual Cortex Neurons in Monkeys and Cats: Detection, Discrimination, and Identification". In: *Visual Neuroscience* 14.5, pp. 897–919.

Giraud, C (2011). "Low Rank Multivariate Regression". In: *Electronic Journal of Statistics* 5 (none), pp. 775–799.

Gisiger, T and M Boukadoum (2011). "Mechanisms Gating the Flow of Information in the Cortex: What They Might Look Like and What Their Uses May Be". In: *Frontiers in Computational Neuroscience* 5.

Gokcen, E, A Jasper, A Xu, A Kohn, CK Machens, and BM Yu (2024). "Uncovering Motifs of Concurrent Signaling across Multiple Neuronal Populations". In: *Advances in Neural Information Processing Systems* 36, pp. 34711–34722.

Gokcen, E, AI Jasper, JD Semedo, A Zandvakili, A Kohn, CK Machens, and BM Yu (2022). "Disentangling the Flow of Signals between Populations of Neurons". In: *Nature Computational Science* 2.8, pp. 512–525.

Golden, JR, KP Vilankar, MCK Wu, and DJ Field (2016). "Conjectures Regarding the Nonlinear Geometry of Visual Neurons". In: *Vision Research*. Vision and the Statistics of the Natural Environment 120, pp. 74–92.

Goris, RLT, JA Movshon, and EP Simoncelli (2014). "Partitioning Neuronal Variability". In: *Nature Neuroscience* 17.6 (6), pp. 858–865.

Gozel, O and B Doiron (2022). *Between-Area Communication through the Lens of within-Area Neuronal Dynamics*. Pre-published.

Hahn, G, A Ponce-Alvarez, G Deco, A Aertsen, and A Kumar (2019). "Portraits of Communication in Neuronal Networks". In: *Nature Reviews Neuroscience* 20.2 (2), pp. 117–127.

Hartwig, RE (1986). "The Reverse Order Law Revisited". In: *Linear Algebra and its Applications* 76, pp. 241–246.

Hua, Y, M Nikpour, and P Stoica (2001). "Optimal Reduced-Rank Estimation and Filtering". In: *IEEE Transactions on Signal Processing* 49.3, pp. 457–469.

Huang, C, A Pouget, and B Doiron (2022). "Internally Generated Population Activity in Cortical Networks Hinders Information Transmission". In: *Science Advances* 8.22, eabg5244.

Huang, W and K Zhang (2018). "Information-Theoretic Bounds and Approximations in Neural Population Coding". In: *Neural Computation* 30.4, pp. 885–944.

Ince, RAA, NJ van Rijsbergen, G Thut, GA Rousselet, J Gross, S Panzeri, and PG Schyns (2015). "Tracing the Flow of Perceptual Features in an Algorithmic Brain Network". In: *Scientific Reports* 5.1, p. 17681.

Iyer, R, JH Siegle, G Mahalingam, S Olsen, and S Mihalas (2021). *Geometry of Inter-Areal Interactions in Mouse Visual Cortex*. Pre-published.

Izenman, AJ (2008). *Modern Multivariate Statistical Techniques*. Red. by G Casella, S Fienberg, and I Olkin. Springer Texts in Statistics. New York, NY: Springer New York.

Izenman, AJ (1975). "Reduced-Rank Regression for the Multivariate Linear Model". In: *Journal of Multivariate Analysis* 5.2, pp. 248–264.

Javadzadeh, M and SB Hofer (2022). "Dynamic Causal Communication Channels between Neocortical Areas". In: *Neuron* 110.15, 2470–2483.e7.

Jun, NY, DA Ruff, LE Kramer, B Bowes, ST Tokdar, MR Cohen, and JM Groh (2022). "Coordinated Multiplexing of Information about Separate Objects in Visual Cortex". In: *eLife* 11. Ed. by S Ray, JI Gold, and R Coen-Cagli, e76452.

Kafashan, M, AW Jaffe, SN Chettih, R Nogueira, I Arandia-Romero, CD Harvey, R Moreno-Bote, and J Drugowitsch (2021). "Scaling of Sensory Information in Large Neural Populations Shows Signatures of Information-Limiting Correlations". In: *Nature Communications* 12.1 (1), p. 473.

Kang, B and S Druckmann (2020). "Approaches to Inferring Multi-Regional Interactions from Simultaneous Population Recordings". In: *Current Opinion in Neurobiology*. Whole-Brain Interactions between Neural Circuits 65, pp. 108–119.

Kanitscheider, I, R Coen-Cagli, A Kohn, and A Pouget (2015). "Measuring Fisher Information Accurately in Correlated Neural Populations". In: *PLOS Computational Biology* 11.6, e1004218.

Kass, RE, H Bong, M Olarinre, Q Xin, and K Urban (2023). "Identification of Interacting Neural Populations: Methods and Statistical Considerations". In: *Journal of Neurophysiology*, jn.00131.2023.

Kaufman, MT, MM Churchland, SI Ryu, and KV Shenoy (2014). "Cortical Activity in the Null Space: Permitting Preparation without Movement". In: *Nature Neuroscience* 17.3 (3), pp. 440–448.

Kohn, A, R Coen-Cagli, I Kanitscheider, and A Pouget (2016). "Correlations and Neuronal Population Information". In: *Annual Review of Neuroscience* 39.1 (1), pp. 237–256. pmid: 27145916.

Kohn, A, AI Jasper, JD Semedo, E Gokcen, CK Machens, and BM Yu (2020). "Principles of Corticocortical Communication: Proposed Schemes and Design Considerations". In: *Trends in Neurosciences* 43.9, pp. 725–737.

Kriegeskorte, N and XX Wei (2021). "Neural Tuning and Representational Geometry". In: *Nature Reviews Neuroscience* 22.11, pp. 703–718.

Lewandowski, D, D Kurowicka, and H Joe (2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". In: *Journal of Multivariate Analysis* 100.9, pp. 1989–2001.

Lyamzin, DR, SJ Barnes, R Donato, JA Garcia-Lazaro, T Keck, and NA Lesica (2015). "Nonlinear Transfer of Signal and Noise Correlations in Cortical Networks". In: *Journal of Neuroscience* 35.21 (21), pp. 8065–8080. pmid: 26019325.

MacDowell, CJ, A Libby, CI Jahn, S Tafazoli, and TJ Buschman (2023). *Multiplexed Subspaces Route Neural Activity Across Brain-wide Networks*. preprint. Neuroscience.

Manning, TS, E Alexander, BG Cumming, GC DeAngelis, X Huang, and EA Cooper (2024). "Transformations of Sensory Information in the Brain Suggest Changing Criteria for Optimality". In: *PLOS Computational Biology* 20.1, e1011783.

Mastrogiuseppe, F and S Ostojic (2018). "Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks". In: *Neuron* 99.3, 609–623.e29.

Moreno-Bote, R, J Beck, I Kanitscheider, X Pitkow, P Latham, and A Pouget (2014). "Information-Limiting Correlations". In: *Nature Neuroscience* 17.10 (10), pp. 1410–1417.

Murphy, KP (2021). *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning Series. Cambridge, MA: MIT Press. 1067 pp.

Negahban, S and MJ Wainwright (2011). "Estimation of (near) Low-Rank Matrices with Noise and High-Dimensional Scaling". In: *The Annals of Statistics* 39.2.

Panzeri, S, M Moroni, H Safaai, and CD Harvey (2022). "The Structures and Functions of Correlations in Neural Population Codes". In: *Nature Reviews Neuroscience*, pp. 1–17.

Penrose, R (1955). "A Generalized Inverse for Matrices". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3, pp. 406–413.

Perich, MG and K Rajan (2020). "Rethinking Brain-Wide Interactions through Multi-Region 'Network of Networks' Models". In: *Current Opinion in Neurobiology*. Whole-Brain Interactions between Neural Circuits 65, pp. 146–151.

Perich, MG, C Arlt, et al. (2021). *Inferring Brain-Wide Interactions Using Data-Constrained Recurrent Neural Network Models*. Pre-published.

Pesaran, B, M Hagan, S Qiao, and R Shewcraft (2021). "Multiregional Communication and the Channel Modulation Hypothesis". In: *Current Opinion in Neurobiology*. Developmental Neuroscience 66, pp. 250–257.

Pica, G, M Soltanipour, and S Panzeri (2019). "Using Intersection Information to Map Stimulus Information Transfer within Neural Networks". In: *Biosystems* 185, p. 104028.

Quian Quiroga, R and S Panzeri (2009). "Extracting Information from Neuronal Populations: Information Theory and Decoding Approaches". In: *Nature Reviews Neuroscience* 10.3 (3), pp. 173–185.

Renart, A and MCW van Rossum (2012). "Transmission of Population-Coded Information". In: *Neural Computation* 24.2, pp. 391–407.

Riesenhuber, M and T Poggio (1999). "Hierarchical Models of Object Recognition in Cortex". In: *Nature Neuroscience* 2.11, pp. 1019–1025.

Ritz, H and A Shenhav (2024). "Orthogonal Neural Encoding of Targets and Distractors Supports Multivariate Cognitive Control". In: *Nature Human Behaviour*, pp. 1–17.

Rowland, JM, TL van der Plas, M Loidolt, RM Lees, J Keeling, J Dehning, T Akam, V Priesemann, and AM Packer (2023). "Propagation of Activity through the Cortical Hier-

archy and Perception Are Determined by Neural Variability". In: *Nature Neuroscience* 26.9 (9), pp. 1584–1594.

Ruff, DA and MR Cohen (2017). "A Normalization Model Suggests That Attention Changes the Weighting of Inputs between Visual Areas". In: *Proceedings of the National Academy of Sciences* 114.20, E4085–E4094.

Rumyantsev, OI et al. (2020). "Fundamental Bounds on the Fidelity of Sensory Cortical Coding". In: *Nature*, pp. 1–6.

Seguin, C, O Sporns, and A Zalesky (2023). "Brain Network Communication: Concepts, Models and Applications". In: *Nature Reviews Neuroscience* 24.9 (9), pp. 557–574.

Semedo, JD, E Gokcen, CK Machens, A Kohn, and BM Yu (2020). "Statistical Methods for Dissecting Interactions between Brain Areas". In: *Current Opinion in Neurobiology*. Whole-Brain Interactions between Neural Circuits 65, pp. 59–69.

Semedo, JD, AI Jasper, A Zandvakili, A Krishna, A Aschner, CK Machens, A Kohn, and BM Yu (2022). "Feedforward and Feedback Interactions between Visual Cortical Areas Use Different Population Activity Patterns". In: *Nature Communications* 13.1 (1), p. 1099.

Semedo, JD, A Zandvakili, CK Machens, BM Yu, and A Kohn (2019). "Cortical Areas Interact through a Communication Subspace". In: *Neuron* 102.1, 249–259.e4.

Seriès, P, PE Latham, and A Pouget (2004). "Tuning Curve Sharpening for Orientation Selectivity: Coding Efficiency and the Impact of Correlations". In: *Nature Neuroscience* 7.10, pp. 1129–1135.

Shadlen, MN and WT Newsome (1998). "The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding". In: *Journal of Neuroscience* 18.10 (10), pp. 3870–3896. pmid: 9570816.

Shew, WL, H Yang, S Yu, R Roy, and D Plenz (2011). "Information Capacity and Transmission Are Maximized in Balanced Cortical Networks with Neuronal Avalanches". In: *Journal of Neuroscience* 31.1, pp. 55–63. pmid: 21209189.

Srinath, R, AM Ni, C Marucci, MR Cohen, and DH Brainard (2024). *Orthogonal Neural Representations Support Perceptual Judgements of Natural Stimuli*. Pre-published.

Srinath, R, DA Ruff, and MR Cohen (2021). "Attention Improves Information Flow between Neuronal Populations without Changing the Communication Subspace". In: *Current Biology* 31.23, 5299–5313.e4.

Stein, RB, ER Gossen, and KE Jones (2005). "Neuronal Variability: Noise or Part of the Signal?" In: *Nature Reviews Neuroscience* 6.5 (5), pp. 389–397.

Timme, NM and C Lapish (2018). "A Tutorial for Information Theory in Neuroscience". In: *eNeuro* 5.3. pmid: 30211307.

Tishby, N, FC Pereira, and W Bialek (2000). *The Information Bottleneck Method*. arXiv: physics/0004057. Pre-published.

Tolhurst, DJ, JA Movshon, and AF Dean (1983). "The Statistical Reliability of Signals in Single Neurons in Cat and Monkey Visual Cortex". In: *Vision research* 23.8 (8), pp. 775–785.

Tso, MKS (1981). "Reduced-Rank Regression and Canonical Analysis". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 43.2, pp. 183–189. JSTOR: 2984847.

Urai, AE, B Doiron, AM Leifer, and AK Churchland (2022). "Large-Scale Neural Recordings Call for New Insights to Link Brain and Behavior". In: *Nature Neuroscience* 25.1, pp. 11–19.

user1551 (n.d.). *Positive Semidefinite Inequality: $(AXA^T)^\square \geq (AXA^T + Y)^{\square_1}$ on Im(A)*. Mathematics Stack Exchange. eprint: https://math.stackexchange.com/q/4853673.

Wang, Z, AA Stocker, and DD Lee (2016). "Efficient Neural Codes That Minimize Lp Reconstruction Error". In: *Neural Computation* 28.12, pp. 2656–2686.

Wei, XX and AA Stocker (2016). "Mutual Information, Fisher Information, and Efficient Coding". In: *Neural Computation* 28.2, pp. 305–326.

Whiteway, MR, K Socha, V Bonin, and DA Butts (2019). "Characterizing the Nonlinear Structure of Shared Variability in Cortical Neuron Populations Using Latent Variable Models". arXiv: 1904.10441 [q-bio].

Zamir, R (1998). "A Proof of the Fisher Information Inequality via a Data Processing Argument". In: *IEEE Transactions on Information Theory* 44.3, pp. 1246–1250.

Zhang, F (2011). *Matrix Theory: Basic Results and Techniques*. Universitext. New York, NY: Springer New York.

Zylberberg, J, A Pouget, PE Latham, and E Shea-Brown (2017). "Robust Information Propagation through Noisy Neural Circuits". In: *PLOS Computational Biology* 13.4, e1005497.

A   Appendix 1

To define the projections onto the communication subspace $\mathbf{P}_{cs}$ and onto the private subspace $\mathbf{Q}_{cs}$, we will use the (Moore-Penrose) pseudoinverse (written $\mathbf{A}^+$) (Penrose 1955). The pseudoinverse is the unique matrix satisfying the following equations:

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \tag{24a}$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \tag{24b}$$

$$(\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+ \tag{24c}$$

$$(\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A} \tag{24d}$$

Concretely, if $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ is the singular value decomposition (SVD), then $\mathbf{A}^+ = \mathbf{V}\mathbf{S}^+\mathbf{U}^\top$, where $\mathbf{S}^+$ is a diagonal matrix whose nonzero entries are the reciprocals of the nonzero diagonal entries of $\mathbf{S}$ (i.e., the reciprocals of the nonzero singular values of $\mathbf{A}$; the singular values equal to 0 are retained). For any linear map $\mathbf{A}$, the map $\mathbf{A}^+\mathbf{A}$ is equivalent to the projection onto $\ker(\mathbf{A})^\perp$ (Axler 2024, Propositions 6.69 and 7.75). As this is how we defined the communication subspace previously, we can define the projection onto the communication subspace as

$$\mathbf{P}_{cs} := \mathbf{B}_{cs}^+\mathbf{B}_{cs} \tag{25}$$

B   Appendix 2

Here we prove that the two expressions for Fisher information in the communication subspace, introduced in Section 2.2.1), are equivalent, using the following proposition:

Lemma 1. *Let $\mathbf{A}, \mathbf{B}$ be injective matrices (i.e., matrices with full column rank) of size $n \times k$, and let $\mathbf{C}$ be a $k \times k$ square invertible matrix. For any k-dimensional vectors a, b, the following identity holds:*

$$b^\top \mathbf{B}^\top (\mathbf{A}\mathbf{C}\mathbf{B}^\top)^+ \mathbf{A}a = b^\top \mathbf{C}^+ a$$

Proof. It suffices to show that $(\mathbf{A}\mathbf{C}\mathbf{B}^\top)^+ = (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ since

$$\begin{aligned}
b^\top \mathbf{B}^\top (\mathbf{A}\mathbf{C}\mathbf{B}^\top)^+ \mathbf{A}a &= b^\top \mathbf{B}^\top (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\mathbf{A}a \\
&= b^\top \mathbf{B}^\top (\mathbf{B}^+)^\top \mathbf{C}^+\mathbf{A}^+\mathbf{A}a \\
&= b^\top (\mathbf{B}^+\mathbf{B})^\top \mathbf{C}^+(\mathbf{A}^+\mathbf{A})a \\
&= b^\top (\mathbf{B}^+\mathbf{B})\mathbf{C}^+(\mathbf{A}^+\mathbf{A})a \\
&= b^\top \mathbf{C}^+ a
\end{aligned}$$

Where we have used the facts that $\mathbf{B}^+\mathbf{B}$ is symmetric (Eq 24d) and the pseudoinverse of an injective matrix is a left inverse of $\mathbf{A}$ (i.e., $\mathbf{A}^+\mathbf{A} = \mathbf{I}$ for $\mathbf{A}$ injective) because $\mathbf{A}^+\mathbf{A} = \mathbf{P}_{\ker(\mathbf{A})^\perp}$ and $\ker(\mathbf{A}) = \{0\}$ since $\mathbf{A}$ is injective; thus, its orthogonal complement is the whole space. In order to show that $(\mathbf{A}\mathbf{C}\mathbf{B}^\top)^+ = (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ we simply check that $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ satisfies

Eqs 24a-d for $\mathbf{ACB}^\top$; by uniqueness of the pseudoinverse, the result follows. To see this, we first note that:

$$\mathbf{ACB}^\top(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+ = \mathbf{AC}\left(\mathbf{B}^+\mathbf{B}\right)^\top\mathbf{C}^+\mathbf{A}^+$$
$$= \mathbf{ACC}^+\mathbf{A}^+ \qquad (26)$$
$$= \mathbf{AA}^+$$

Because $\mathbf{B}$ is injective and $\mathbf{C}$ is invertible. This proves that $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ satisfies Eq 24c, because $\mathbf{AA}^+$ is symmetric by the same requirement. A similar computation shows that $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\mathbf{ACB}^\top = \mathbf{BB}^+$, which shows that $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ satisfies Eq 24d because $\mathbf{BB}^+$ is symmetric (Eq 24c). Using these facts, we can calculate:

$$\mathbf{ACB}^\top(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\mathbf{ACB}^\top = \left(\mathbf{ACB}^\top(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\right)\mathbf{ACB}^\top$$
$$= \mathbf{AA}^+\mathbf{ACB}^\top \qquad \text{(by Eq 26)}$$
$$= (\mathbf{AA}^+\mathbf{A})\mathbf{CB}^\top$$
$$= \mathbf{ACB}^\top \qquad \text{(by Eq 24a)}$$

Which shows that $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ satisfies Eq 24a. Similarly, to see that this meets Eq 24b,

$$(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\mathbf{ACB}^\top(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+ = (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\left(\mathbf{ACB}^\top(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\right)$$
$$= (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+\mathbf{AA}^+ \qquad \text{(by Eq 26)}$$
$$= (\mathbf{B}^\top)^+\mathbf{C}^+(\mathbf{A}^+\mathbf{AA}^+)$$
$$= (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+ \qquad \text{(by Eq 24b)}$$

Therefore, $(\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$ meets all of the requirements to be a pseudoinverse of $\mathbf{ACB}^\top$, so $(\mathbf{ACB}^\top)^+ = (\mathbf{B}^\top)^+\mathbf{C}^+\mathbf{A}^+$. □

We additionally need the following three facts: first, $\mathbf{\Sigma}$ is symmetric positive semi-definite (that is, if $x^\top\mathbf{\Sigma}x \geq 0$ for all vectors x) *if and only if*

$$\mathbf{\Sigma} = \mathbf{M}^\top\mathbf{M} \qquad (27)$$

If $\mathbf{\Sigma}$ is also invertible (i.e., symmetric positive-*definite*, $x^\top\mathbf{\Sigma}x > 0$ for all vectors $x \neq 0$), then $\mathbf{M}$ in Eq 27 can be taken to be a symmetric and invertible matrix (Zhang 2011, Theorem 7.3).

Second, for any matrix $\mathbf{C}$,

$$\ker(\mathbf{C}^\top\mathbf{C}) = \ker(\mathbf{C}) \qquad (28)$$

To see this, suppose $x \in \ker(\mathbf{C}^\top\mathbf{C})$. Consider $x^\top\mathbf{C}^\top\mathbf{C}x = x^\top 0 = 0 = (\mathbf{C}x)^\top(\mathbf{C}x)$. By definition of the inner product, this implies $\mathbf{C}x = 0$, or $x \in \ker(\mathbf{C})$, so $\ker(\mathbf{C}^\top\mathbf{C}) \subseteq \ker(\mathbf{C})$, and the other inclusion is trivial.

Finally, if $\mathbf{\Sigma}$ is a positive-definite matrix and $\mathbf{V}$ is an injective matrix, then $\mathbf{A} = \mathbf{V}^\top\mathbf{\Sigma}\mathbf{V}$ is positive-definite. To see this, let $\mathbf{\Sigma} = \mathbf{M}^2$ by Eq 27. Then

$$\mathbf{A} = \mathbf{V}^\top\mathbf{MMV} = (\mathbf{MV})^\top\mathbf{MV} \qquad (29)$$

so $\mathbf{A}$ satisfies Eq 27, which means that $\mathbf{A}$ is positive semi-definite. To show that $\mathbf{A}$ is

positive-definite, we need to show that $\mathbf{A}$ is invertible. To do so, it suffices to show that $\ker(\mathbf{A}) = \{0\}$, since $\mathbf{A}$ is a square matrix and by the rank-nullity theorem, $\ker(\mathbf{A}) = \{0\}$ implies $\mathbf{A}$ has full column-rank. Suppose $x \in \ker(\mathbf{A}) = \ker(\mathbf{V}^\top \mathbf{\Sigma} \mathbf{V})$. This implies that $x \in \ker(\mathbf{MV})$ because $\mathbf{A}$ is positive semi-definite (Eq 29) and by Eq 28. In other words, this means $\mathbf{MV}x = 0$. Because $\mathbf{M}$ is invertible, this implies that $\mathbf{V}x = 0$. However, this means $x = 0$ by the injectivity of $\mathbf{V}$, so $\ker(\mathbf{A}) = \{0\}$.

Now, we can prove that

$$\mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta) = \mathbf{J}(\mathbf{B}_{cs}\mathbf{x}; \theta) \tag{30}$$

Proof. For $\mathbf{B}_{cs}$, we define the compact SVD by $\mathbf{B}_{cs} = \mathbf{V}_k \mathbf{S}_k \mathbf{W}_k^\top$, for $\mathbf{S}_k$ a diagonal $k \times k$ matrix corresponding to the non-zero singular values of $\mathbf{B}_{cs}$. In this notation, $\mathbf{P}_{cs} = \mathbf{W}_k \mathbf{W}_k^\top$ by direct computation.

We will first show that $\mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta) = \mathbf{J}(\mathbf{W}_k^\top \mathbf{x}; \theta)$. Define $a = \mathbf{W}_k^\top d\mathbf{f}_x$ and $\mathbf{A} = \mathbf{W}_k^\top \mathbf{\Sigma}_x \mathbf{W}_k$. From the above, $\mathbf{A}$ is positive-definite and, in particular, invertible. Then

$$
\begin{aligned}
\mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta) &= d\mathbf{f}_x^\top \mathbf{W}_k \mathbf{W}_k^\top (\mathbf{W}_k \mathbf{W}_k^\top \mathbf{\Sigma}_x \mathbf{W}_k \mathbf{W}_k^\top)^+ \mathbf{W}_k \mathbf{W}_k^\top d\mathbf{f}_x \\
&= a^\top \mathbf{W}_k^\top (\mathbf{W}_k \mathbf{A} \mathbf{W}_k^\top)^+ \mathbf{W}_k a \\
&= a^\top \mathbf{A}^+ a && \text{(by Lemma 1)} \\
&= d\mathbf{f}_x^\top \mathbf{W}_k (\mathbf{W}_k^\top \mathbf{\Sigma}_x \mathbf{W}_k)^+ \mathbf{W}_k^\top d\mathbf{f}_x \\
&= \mathbf{J}(\mathbf{W}_k^\top \mathbf{x}; \theta)
\end{aligned}
$$

Now, it suffices to show that

$$\mathbf{J}(\mathbf{B}_{cs}\mathbf{x}; \theta) = \mathbf{J}(\mathbf{W}_k^\top \mathbf{x}; \theta)$$

We can do so using the same logic as above. Define $a, \mathbf{A}$ as before, and let $\widehat{\mathbf{V}} = \mathbf{V}_k \mathbf{S}_k$. Note that $\widehat{\mathbf{V}}$ is an injective matrix: if $\mathbf{V}_k \mathbf{S}_k x = 0$, by the injectivity of $\mathbf{V}_k$ this implies $\mathbf{S}_k x = 0$. However, $\mathbf{S}_k$ is invertible, so this is only true if $x = 0$. Then, using the SVD of $\mathbf{B}_{cs}$:

$$
\begin{aligned}
\mathbf{J}(\mathbf{B}_{cs}\mathbf{x}; \theta) &= d\mathbf{f}_x^\top \mathbf{B}_{cs}^\top (\mathbf{B}_{cs} \mathbf{\Sigma}_x \mathbf{B}_{cs}^\top)^+ \mathbf{B}_{cs} d\mathbf{f}_x \\
&= d\mathbf{f}_x^\top \mathbf{W}_k \mathbf{S}_k \mathbf{V}_k^\top (\mathbf{V}_k \mathbf{S}_k \mathbf{W}_k^\top \mathbf{\Sigma}_x \mathbf{W}_k \mathbf{S}_k \mathbf{V}_k^\top)^+ \mathbf{V}_k \mathbf{S}_k \mathbf{W}_k^\top d\mathbf{f}_x \\
&= a^\top \widehat{\mathbf{V}}^\top (\widehat{\mathbf{V}} \mathbf{A} \widehat{\mathbf{V}}^\top)^+ \widehat{\mathbf{V}} a \\
&= a \mathbf{A}^+ a && \text{(by Lemma 1)} \\
&= d\mathbf{f}_x^\top \mathbf{W}_k (\mathbf{W}_k^\top \mathbf{\Sigma}_x \mathbf{W}_k)^+ \mathbf{W}_k^\top d\mathbf{f}_x \\
&= \mathbf{J}(\mathbf{W}_k^\top \mathbf{x}; \theta)
\end{aligned}
$$

$\square$

C   Appendix 3

Lemma 2. *Suppose Σ is a symmetric positive-definite matrix and let $\mathbf{P}$ be a orthogonal projection ($\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}^\top = \mathbf{P}$). Then*

$$\Sigma^{-1} - (\mathbf{P}\Sigma\mathbf{P})^+ \text{ is positive semi-definite} \tag{31}$$

Proof. Let n be the rank of Σ and let $k < n$ be the rank of $\mathbf{P}$. There exists a basis for $\mathbb{R}^n$ such that $\mathbf{P}\Sigma\mathbf{P} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ for $\mathbf{A}$ a symmetric positive-definite $k \times k$ matrix: create an orthonormal basis $\{u_1, \cdots, u_k\}$ for the column-space of $\mathbf{P}$ and complete to a basis $\mathbf{U}$ of $\mathbb{R}^n$; the matrix $\mathbf{U}^{-1}\mathbf{P}\mathbf{U}$ is now of the form $\begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, so $\mathbf{P}\Sigma\mathbf{P}$ has the desired form in this basis. Similarly, in this basis $\Sigma = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{pmatrix}$, for $\mathbf{B}$ a $k \times n - k$ matrix and $\mathbf{C}$ an $n - k \times n - k$ matrix. We can write Σ^{-1} as a function of the matrices in this partition using the Schur complement $\Sigma/\mathbf{A} = \mathbf{C} - \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B}$ (Zhang 2011, Theorems 2.3 and 2.4) and compute $(\mathbf{P}\Sigma\mathbf{P})^+ = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ directly by checking the defining properties of the pseudoinverse (Eq 24). With this notation:

$$\Sigma^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1} \\ -(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & (\Sigma/\mathbf{A})^{-1} \end{pmatrix}$$

$$= (\mathbf{P}\Sigma\mathbf{P})^+ + \underbrace{\begin{pmatrix} \mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1} \\ -(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & (\Sigma/\mathbf{A})^{-1} \end{pmatrix}}_{\mathbf{X}}$$

It suffices to show that $\mathbf{X}$ defined above is positive semi-definite. To show this, we need to find a matrix $\mathbf{Y}$ such that $\mathbf{X} = \mathbf{Y}^\top\mathbf{Y}$ (Eq 27). Because Σ is positive-definite, $\Sigma/\mathbf{A}$ is also positive-definite (Gallier 2011, Proposition 16.2), which implies $(\Sigma/\mathbf{A})^{-1}$ is positive-definite and there is a symmetric matrix $\mathbf{M}$ such that $(\Sigma/\mathbf{A})^{-1} = \mathbf{M}^2$. Define $\mathbf{Y} = (-\mathbf{M}\mathbf{B}^\top\mathbf{A}^{-1} \quad \mathbf{M})$, which is an $n - k \times n$ matrix. Then

$$\mathbf{Y}^\top\mathbf{Y} = \begin{pmatrix} -\mathbf{A}^{-1}\mathbf{B}\mathbf{M} \\ \mathbf{M} \end{pmatrix}(-\mathbf{M}\mathbf{B}^\top\mathbf{A}^{-1} \quad \mathbf{M}) = \begin{pmatrix} \mathbf{A}^{-1}\mathbf{B}\mathbf{M}^2\mathbf{B}^\top\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{M}^2 \\ -\mathbf{M}^2\mathbf{B}^\top\mathbf{A}^{-1} & \mathbf{M}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\Sigma/\mathbf{A})^{-1} \\ -(\Sigma/\mathbf{A})^{-1}\mathbf{B}^\top\mathbf{A}^{-1} & (\Sigma/\mathbf{A})^{-1} \end{pmatrix}$$

$$= \mathbf{X}$$

Therefore, $\mathbf{X} = \Sigma^{-1} - (\mathbf{P}\Sigma\mathbf{P})^+$ is positive semi-definite. $\qquad\square$

To see how this proves Eq 16, write:

$$\mathbf{CI}_{cs} - \mathbf{J}(\mathbf{P}_{cs}\mathbf{x}; \theta) = d\mathbf{f}_x^\top\mathbf{P}_{cs}\Sigma_x^{-1}\mathbf{P}_{cs}d\mathbf{f}_x - d\mathbf{f}_x^\top\mathbf{P}_{cs}\left(\mathbf{P}_{cs}\Sigma_x^{-1}\mathbf{P}_{cs}\right)^+\mathbf{P}_{cs}d\mathbf{f}_x$$

$$= d\mathbf{f}_x^\top\mathbf{P}_{cs}\left(\Sigma_x^{-1} - \left(\mathbf{P}_{cs}\Sigma_x^{-1}\mathbf{P}_{cs}\right)^+\right)\mathbf{P}_{cs}d\mathbf{f}_x \geq 0 \qquad \text{(by Lemma 2)}$$

and similarly for Eq 17. As noted, this also shows that $\mathbf{J}(\mathbf{x}; \theta) \geq \mathbf{J}_{\mathrm{cs}}$ and $\mathbf{J}_{\mathrm{priv}}$, as similar logic to the above proof would demonstrate that $\mathbf{\Sigma}^{-1} - \mathbf{P}(\mathbf{P}\mathbf{\Sigma}\mathbf{P})^+ \mathbf{P}$ is positive semi-definite.

D   Appendix 4

Let $v = \mathbf{P}_{\mathrm{cs}}\mathbf{df}_x$ and $w = \mathbf{Q}_{\mathrm{cs}}\mathbf{df}_x$. Then

$$\mathbf{J}(\mathbf{x}; \theta) - \mathbf{CI}_{\mathrm{cs}} = w^\top \mathbf{\Sigma}_x^{-1} w + 2v^\top \mathbf{\Sigma}_x^{-1} w$$

There must be some v such that $v^\top \mathbf{\Sigma}_x^{-1} w < 0$ (for example, $-w$ is one such vector because $\mathbf{\Sigma}_x^{-1}$ is positive-definite, so $(-w)^\top \mathbf{\Sigma}_x^{-1} w = -(w^\top \mathbf{\Sigma}_x^{-1} w) < 0$). Let $z = (1+\epsilon) \cdot \frac{w^\top \mathbf{\Sigma}_x^{-1} w}{2|v^\top \mathbf{\Sigma}_x^{-1} w|} v$ for $\epsilon > 0$ some constant. Then since $\mathbf{\Sigma}_x$ is positive-definite, we have

$$\mathbf{J}(\mathbf{x}; \theta) - \mathbf{CI}_{\mathrm{cs}} = -\epsilon(w^\top \mathbf{\Sigma}_x^{-1} w) < 0$$

This calculation shows that there will be some tuning curve and some population covariance matrix such that $\mathbf{J}(\mathbf{x}; \theta) < \mathbf{CI}_{\mathrm{cs}}$.

E   Appendix 5

The more direct argument to prove that the impactful information is bounded by the information in the communication subspace, would be to show, similar to Lemma 2, that

$$\mathbf{B}_{\mathrm{cs}}^\top (\mathbf{B}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{B}_{\mathrm{cs}}^\top)^+ \mathbf{B}_{\mathrm{cs}} - \mathbf{B}_{\mathrm{cs}}^\top (\mathbf{B}_{\mathrm{cs}}\mathbf{\Sigma}_x\mathbf{B}_{\mathrm{cs}}^\top + \mathbf{\Sigma}_r)^{-1} \mathbf{B}_{\mathrm{cs}} \text{ is positive semi-definite} \quad (32)$$

To do so, we prove the following lemma (this lemma and the argument that this proves Eq 32 are due to user1551 n.d.).

Lemma 3. *Let $\mathbf{A}$ be an $n \times n$ symmetric positive semi-definite matrix, let $\mathbf{B}$ be an $n \times n$ symmetric positive-definite matrix, and assume that $\mathbf{A} + \mathbf{B}$ is invertible. Then*

$$\mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} \text{ is positive semi-definite} \quad (33)$$

Proof. By assumption, $\mathbf{A}$ is positive semi-definite, so it trivially follows that $\mathbf{A} + \mathbf{B} - \mathbf{B}$ is positive semi-definite. Because $\mathbf{A} + \mathbf{B}$ and $\mathbf{B}$ are positive-definite and invertible, we have that $\mathbf{B}^{-1} - (\mathbf{A} + \mathbf{B})^{-1}$ is positive semi-definite (Zhang 2011, Theorem 7.8). For any positive semi-definite matrix $\mathbf{X}$ and any matrix $\mathbf{Y}$, $\mathbf{Y}^\top \mathbf{X}\mathbf{Y}$ is positive semi-definite (Eq 29), and $\mathbf{B}$ is symmetric. This means that

$$\mathbf{B}(\mathbf{B}^{-1} - (\mathbf{A} + \mathbf{B})^{-1})\mathbf{B} = \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} \text{ is positive semi-definite}$$

By defining $\mathbf{C} = \mathbf{A} + \mathbf{B}$, we can rewrite this in terms of $\mathbf{A}$ and $\mathbf{C}$:

$$
\begin{aligned}
\mathbf{B} - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} &= (\mathbf{A}+\mathbf{B}) - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} - \mathbf{A} \\
&= \mathbf{C} - (\mathbf{C}-\mathbf{A})\mathbf{C}^{-1}(\mathbf{C}-\mathbf{A}) - \mathbf{A} \\
&= \mathbf{C} - (\mathbf{C}-\mathbf{A})(\mathbf{I} - \mathbf{C}^{-1}\mathbf{A}) - \mathbf{A} \\
&= \mathbf{C} - \mathbf{C} + \mathbf{A} + \mathbf{A} - \mathbf{A}\mathbf{C}^{-1}\mathbf{A} - \mathbf{A} \\
&= \mathbf{A} - \mathbf{A}\mathbf{C}^{-1}\mathbf{A} = \mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}
\end{aligned}
$$

Therefore, $\mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$ positive semi-definite. $\qquad\square$

To see how this generalizes Eq 32, we prove that $\mathrm{Im}(\mathbf{B}_{cs}) = \mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)$. The inclusion $\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top) \subseteq \mathrm{Im}(\mathbf{B}_{cs})$ is trivial. To prove that they are equivalent, we show that the images have the same dimensionality, which implies that they are equal because the only n dimensional subspace of an n dimensional vector space is itself. For this, we consider $\ker(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)$. Because Σ_x is symmetric positive-definite, $\Sigma_x = \mathbf{M}^2$ for $\mathbf{M}$ symmetric and invertible (Eq 27). Using this and Eq 28, we find that

$$
\ker(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top) = \ker(\mathbf{B}_{cs}\mathbf{M}\mathbf{M}\mathbf{B}_{cs}^\top) = \ker((\mathbf{M}\mathbf{B}_{cs}^\top)^\top(\mathbf{M}\mathbf{B}_{cs}^\top)) = \ker(\mathbf{M}\mathbf{B}_{cs}^\top)
$$

Since $\mathbf{M}$ is invertible, $\ker(\mathbf{M}\mathbf{B}_{cs}^\top) = \ker(\mathbf{B}_{cs}^\top)$, so we have $\ker(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top) = \ker(\mathbf{B}_{cs}^\top)$.

By using the rank-nullity theorem and the fact that row-rank is equivalent to column-rank (namely $\dim(\mathrm{Im}(\mathbf{B}_{cs}^\top)) = \dim(\mathrm{Im}(\mathbf{B}_{cs}))$),

$$
\begin{aligned}
q_y &= \dim(\ker(\mathbf{B}_{cs}^\top)) + \dim(\mathrm{Im}(\mathbf{B}_{cs}^\top)) = \dim(\ker(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)) + \dim(\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)) \\
\dim(\mathrm{Im}(\mathbf{B}_{cs})) &= \underbrace{\dim(\ker(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)) - \dim(\ker(\mathbf{B}_{cs}^\top))}_{0} + \dim(\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)) \\
\dim(\mathrm{Im}(\mathbf{B}_{cs})) &= \dim(\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top))
\end{aligned}
$$

Therefore, because $\mathrm{Im}(\mathbf{B}_{cs})$ and $\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)$ have the same dimensionality and $\mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top) \subseteq \mathrm{Im}(\mathbf{B}_{cs}) \Rightarrow \mathrm{Im}(\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top) = \mathrm{Im}(\mathbf{B}_{cs})$.

Now, since $\mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x \in \mathrm{Im}(\mathbf{B}_{cs})$ trivially, there is some vector v in the target population activity space such that $\mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x = \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top v$. Then

$$
\begin{aligned}
\mathbf{J}(\mathbf{B}_{cs}\mathbf{x};\theta) &= \mathrm{d}\mathbf{f}_x^\top \mathbf{B}_{cs}^\top (\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)^+ \mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x \\
&= v^\top \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top (\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top)^+ \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top v \\
&= v^\top \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top v
\end{aligned}
\qquad (34)
$$

Where the last line of Eq 34 follows from a defining property of the pseudoinverse (Eq 24a).

Similarly, we expand $\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \eta_y;\theta)$ as

$$
\begin{aligned}
\mathbf{J}(\mathbf{B}_{cs}\mathbf{x} + \eta_y;\theta) &= \mathrm{d}\mathbf{f}_x^\top \mathbf{B}_{cs}^\top \Sigma_y^{-1} \mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x \\
&= \mathrm{d}\mathbf{f}_x^\top \mathbf{B}_{cs}^\top (\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top + \Sigma_r)^{-1} \mathbf{B}_{cs}\mathrm{d}\mathbf{f}_x \\
&= v^\top \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top (\mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top + \Sigma_r)^{-1} \mathbf{B}_{cs}\Sigma_x\mathbf{B}_{cs}^\top v
\end{aligned}
\qquad (35)
$$

Using Eqs 34 and 35, we can restate Eq 19 as

$$v^\top \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top v \geq v^\top \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top \left(\mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top + \mathbf{\Sigma}_r \right)^{-1} \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top v$$

After rearranging this, we have

$$v^\top \left(\mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top - \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top \left(\mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top + \mathbf{\Sigma}_r \right)^{-1} \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top \right) v \geq 0$$

In other words, Eq 19 is equivalent to the claim that

$$\mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top - \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top \left(\mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top + \mathbf{\Sigma}_r \right)^{-1} \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top \text{ is positive semi-definite} \quad (36)$$

Now defining $\mathbf{A} = \mathbf{B}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{B}_{\mathrm{cs}}^\top$ and $\mathbf{B} = \mathbf{\Sigma}_r$ in Eq 36, we get Eq 33. Therefore, as a corollary of Lemma 3, we have proved Eq 32 and therefore Eq 19, namely that the information in the communication subspace $\mathbf{J}_{\mathrm{cs}} = \mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x}; \theta)$ is an upper bound on the impactful information $\mathbf{J}(\mathbf{B}_{\mathrm{cs}}\mathbf{x} + \boldsymbol{\eta}_y; \theta)$.

F Appendix 6

Here we derive mathematically the results illustrated by the simulations of Fig 4. Central to these derivations will be the following identities:

$$\begin{aligned}
\mathbf{P}_{\mathrm{cs}} \mathbf{f}_x^{\uparrow \mathrm{priv}} &= \mathbf{P}_{\mathrm{cs}} \mathbf{f}_x \\
\mathbf{Q}_{\mathrm{cs}} \mathbf{f}_x^{\uparrow \mathrm{priv}} &= C \mathbf{Q}_{\mathrm{cs}} \mathbf{f}_x \\
\mathbf{B}_{\mathrm{cs}} \mathbf{f}_x^{\uparrow \mathrm{priv}} &= \mathbf{B}_{\mathrm{cs}} \mathbf{f}_x
\end{aligned} \quad (37)$$

These follow immediately from the basic properties of orthogonal projections applied to Eq 22. First, we will use the source Fisher information decomposition (Eq 12) for $\mathbf{J}(\mathbf{x}^{\uparrow \mathrm{priv}}; \theta)$ and simplify using these identities to obtain:

$$\mathbf{J}(\mathbf{x}^{\uparrow \mathrm{priv}}; \theta) = \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{P}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x + C^2 \mathrm{d}\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x + 2C \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x \quad (38)$$

If we examine the difference between the source Fisher information before and after modifying the private tuning

$$\mathbf{J}(\mathbf{x}^{\uparrow \mathrm{priv}}; \theta) - \mathbf{J}(\mathbf{x}; \theta) = (C^2 - 1) \mathrm{d}\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x + (C - 1) 2 \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x \quad (39)$$

Defining $a = \mathrm{d}\mathbf{f}_x^\top \mathbf{Q}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x$, $b = \mathrm{d}\mathbf{f}_x^\top \mathbf{P}_{\mathrm{cs}} \mathbf{\Sigma}_x \mathbf{Q}_{\mathrm{cs}} \mathrm{d}\mathbf{f}_x$, simple algebra will show that this difference is positive as long as $C \geq \dfrac{|a + b| - b}{a}$[3]. This shows that the source Fisher information will increase after modifying the private tuning in this manner (Fig 4B).

---

[3] If $-b > a$, then $C \geq \dfrac{-2b}{a} - 1 > 1$; otherwise, $C \geq 1$. Note that we ignore the smaller root of Eq 39, but in general, if we choose C to be less than this smaller root, then Eq 39 will also be positive.

Next, if we look at the target Fisher information decomposition (Eq 13 with $\mathbf{r}_y = 0$) and using the identities (Eq 37), we obtain

$$\mathbf{J}(\mathbf{y}^{\uparrow\text{priv}}; \theta) = \mathbf{J}(\mathbf{B}_{\text{cs}}\mathbf{x}^{\uparrow\text{priv}} + \boldsymbol{\eta}_y; \theta) = \text{d}\mathbf{f}_x^\top \mathbf{B}_{\text{cs}}^\top \left(\mathbf{B}_{\text{cs}}\boldsymbol{\Sigma}_x\mathbf{B}_{\text{cs}}^\top + \sigma_2^2\mathbf{I}\right)^{-1} \mathbf{B}_{\text{cs}}\text{d}\mathbf{f}_x = \mathbf{J}(\mathbf{B}_{\text{cs}}\mathbf{x}; \theta)$$
(40)

This shows that there is no change in the amount of Fisher information in the target area. The same derivation applies to the information in the communication subspace (using Eq 40 with $\sigma_2 = 0$ and using the pseudoinverse), also demonstrating that the Fisher information in the communication subspace will remain invariant to changes of tuning in the private subspace (Fig 4E).

Lastly, if we compare the Fisher information in the private subspace for $\mathbf{x}$ and $\mathbf{x}^{\uparrow\text{priv}}$, $\mathbf{J}(\mathbf{Q}_{\text{cs}}\mathbf{x}^{\uparrow\text{priv}}; \theta) - \mathbf{J}(\mathbf{Q}_{\text{cs}}\mathbf{x}; \theta)$, we can write this difference as

$$C^2\text{d}\mathbf{f}_x^\top \mathbf{Q}_{\text{cs}} \left(\mathbf{Q}_{\text{cs}}\boldsymbol{\Sigma}_x\mathbf{Q}_{\text{cs}}\right)^+ \mathbf{Q}_{\text{cs}}\text{d}\mathbf{f}_x - \text{d}\mathbf{f}_x^\top \mathbf{Q}_{\text{cs}} \left(\mathbf{Q}_{\text{cs}}\boldsymbol{\Sigma}_x\mathbf{Q}_{\text{cs}}\right)^+ \mathbf{Q}_{\text{cs}}\text{d}\mathbf{f}_x$$
$$= (C^2 - 1)\text{d}\mathbf{f}_x^\top \mathbf{Q}_{\text{cs}} \left(\mathbf{Q}_{\text{cs}}\boldsymbol{\Sigma}_x\mathbf{Q}_{\text{cs}}\right)^+ \mathbf{Q}_{\text{cs}}\text{d}\mathbf{f}_x$$
(41)

Thus, as long as $C > 1$, then $\mathbf{J}(\mathbf{Q}_{\text{cs}}\mathbf{x}^{\uparrow\text{priv}}; \theta) \geq \mathbf{J}(\mathbf{Q}_{\text{cs}}\mathbf{x}; \theta)$, or the Fisher information in the private subspace will increase (Fig 4D).