We would like to thank the reviewers for their insightful comments which will help us improve the article. We begin by addressing concerns shared by several reviewers:

Impact of the noise parameter: We will add Fig.(a) below in appendix. We compute the accuracy of MultiviewICA on the time-segment matching experiment (Fig.2 in the paper) on the sherlock dataset with 10 components when the noise parameter varies. MultiviewICA performs consistently well for a wide range of noise parameter values, and only breaks at very high values. It supports the theoretical claim of Sec.2.3 that the noise parameter is of little importance.

Model 1. is unrealistic: A more natural model is $\mathbf{x}^i = A^i\mathbf{s} + \mathbf{n}^i$ (*) where $\mathbf{n}^i$ is iid. Unfortunately, this model can not be fit in a reasonable amount of time, as reported many times in the literature (see paragraph L.179). For model (1), $\mathbf{x}^i = A^i(\mathbf{s} + \mathbf{n}^i)$, noise can be interpreted as individual variability rather than sensor noise. It offers a way to capture more structured noise, as is often the case in brain signals (cf. ICA based solutions for cleaning EEG or fMRI data). To test its robustness to model mis-specification, we generate data following model (*), and report the reconstruction error in Fig.(b). The difference between algorithms is small. Finally, we argue that whether realistic or not, MultiviewICA is a robust algorithm which in practice improves upon the state-of-the-art on multiple experiments. This discussion will be added in Sec.3.

Model expressivity, comparison to deep methods Our method indeed lacks the expressivity of non-linear models. However, linearity is still widespread for brain signal analysis, and non-linear methods are not necessarily better in this setting where the number of samples is limited ([Deep learning for brains?: Different linear and nonlinear scaling in UK Biobank brain images vs. machine-learning datasets, Schulz et al., 2019]). In appendix E.4, we obtain better results than some non-linear deep-learning methods [21]. We will discuss this more thoroughly in Sec.3.

Rev.1: *Difference in estimated components* We did not perform qualitative comparison of the estimated components, which is usually difficult due to the intrinsic randomness/non-convexity of ICA. We can only say that the *reconstruction* experiment (Fig.2) and the localization on *MEG phantom data* (Fig.3a) rely only on components, which are hence different. We will add some component maps in the appendix, but leave a more qualitative comparison for future work.

Unclear experiment/metric: We will describe the experiments and metric at greater length, and give more intuitions.

Rev.2: *Method section is very dense*: We will do our best to unclutter it, and spend more time on important concepts.

Add CanICA: We will add CanICA in the synthetic and Phantom experiments (see Figs.(b), (c), (d)).

Rev.3: *Shared response?* We will do our best to introduce this key concept more clearly: subjects exposed to the same stimulus should share a common response to it. Many methods have been proposed to reconstruct it (e.g [20]).

Show data: We will add in appendix figures showing data, mixing matrices and shared response

Quantitative and qualitative results: We will add a table summarizing our quantitative results. The only qualitative result in the paper is the Cam-CAN experiment, which uncovers relevant brain regions and clean evoked potentials. While this is not a discovery, it illustrates the potential of MultiviewICA for unsupervised brain data exploration.

Rev.4 *Robustness to kurtosis:* The proposed algorithm follows the route of infomax, and therefore only separates super-Gaussian sources. Arguably, most brain sources are super-Gaussian [24]. We plan to develop an extended version which switches between sub-and super-Gaussian density, like extended-Infomax, in a future work.

Response robustness: We will add a synthetic experiment where we monitor reconstruction error (Fig.(c)).

Link with multiview CCA Thank you for this relevant works which we did not know about. We will mention these methods in Sec.3. We have added Bayesian Multiview CCA (BCorrCA) in the Phantom experiment (Fig.(d) below).

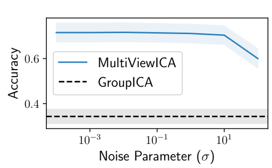
Dimension reduction: All algorithms reported here rely on dimension reduction as preprocessing, hence the concerns of Artoni et. al. apply, but we believe that this problem is shared by all methods considered.

ICA of Lukic et al: We were not aware of this method. Unlike most ICA methods, it leverages non-stationarity of the sources rather than non-Gaussianity, which might result in very different decompositions [24]. We will mention this work in Sec.3.

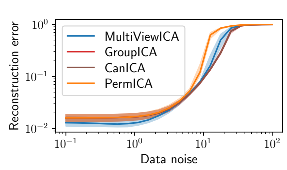
Hyperparameters: The only hyperparameters of competing algorithms are those of the ICA solver, Picard, that achieves robust convergence by default.

CV in Fig 2: The confidence interval is computed over runs and subjects. We will detail this.

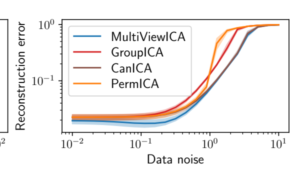
Fig 2 "sherlock" and "forrest" high performance: Possible explanation: Sherlock data undergo a 6mm spatial smoothing and Forrest data are acquired at a higher resolution (7T vs. 3T for other data). This affects SNR.



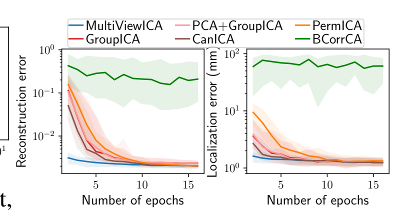
(a) Time segment matching experiment with different noise parameter values



(b) Synthetic experiment, model: $x^i = A^i s + n^i$



(c) Synthetic experiment, model: $x^i = A^i(s + n^i)$



(d) Phantom experiment