

**Background**

**What is reliability?**

- Consistency of a “biomarker”: how trustworthy is the value-tracking signal in the ventral striatum?
- True score theory: scaling of measurement error (psychometric evaluation)
  - The smaller the measurement error, the more reliable is the measure

**Why is it important?**

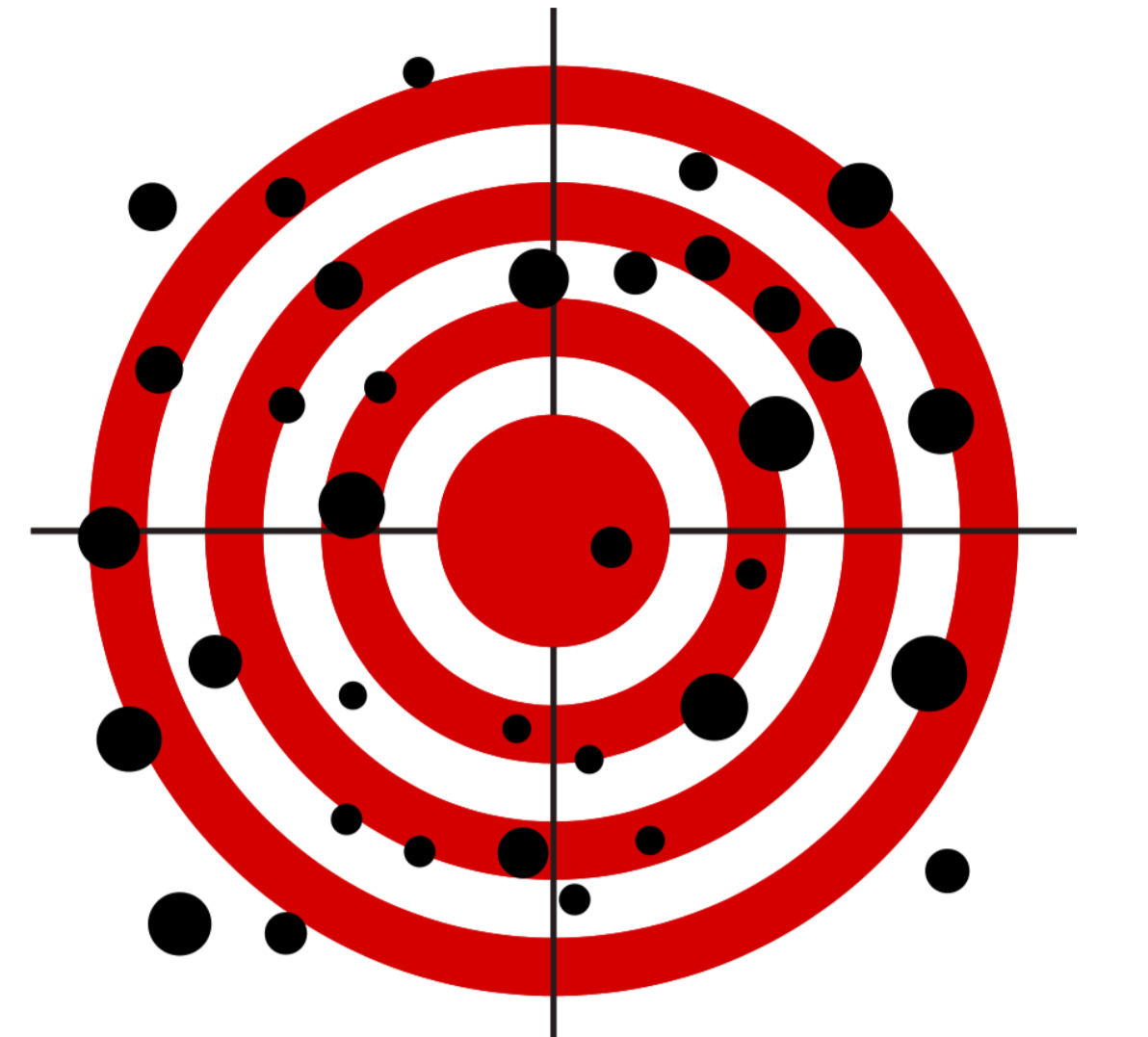
- Helps us to judge whether our measure captures inter- and intra-individual differences well

**What do we already know about reliability in fMRI [1]?**

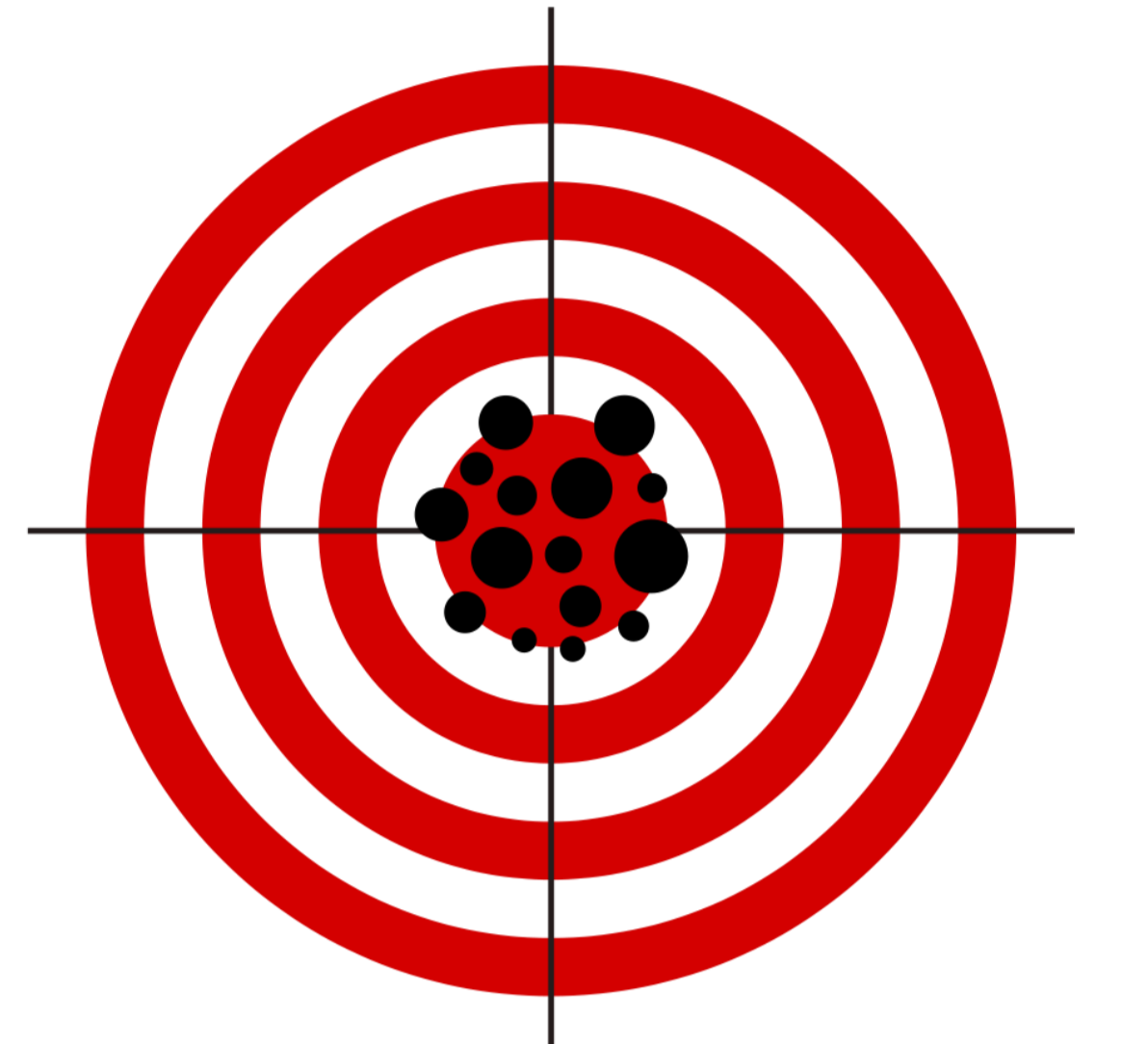
- Reliability is dependent on tasks and the variability in their underlying cognitive functions [e.g. 2,3]
- Range from low to high reliability
- *Reliability is dependent on the level of analysis*: group vs. individual; global vs. local response [4,5]

**How can we make reliability assessment more accessible to the brain mapping community?**

- fmreli is a collection of key indices to facilitate reliability analyses as an integral part of task development
- Novel feature: split-half reliability for single runs/sessions



Unreliable, But Valid



Both Reliable & Valid

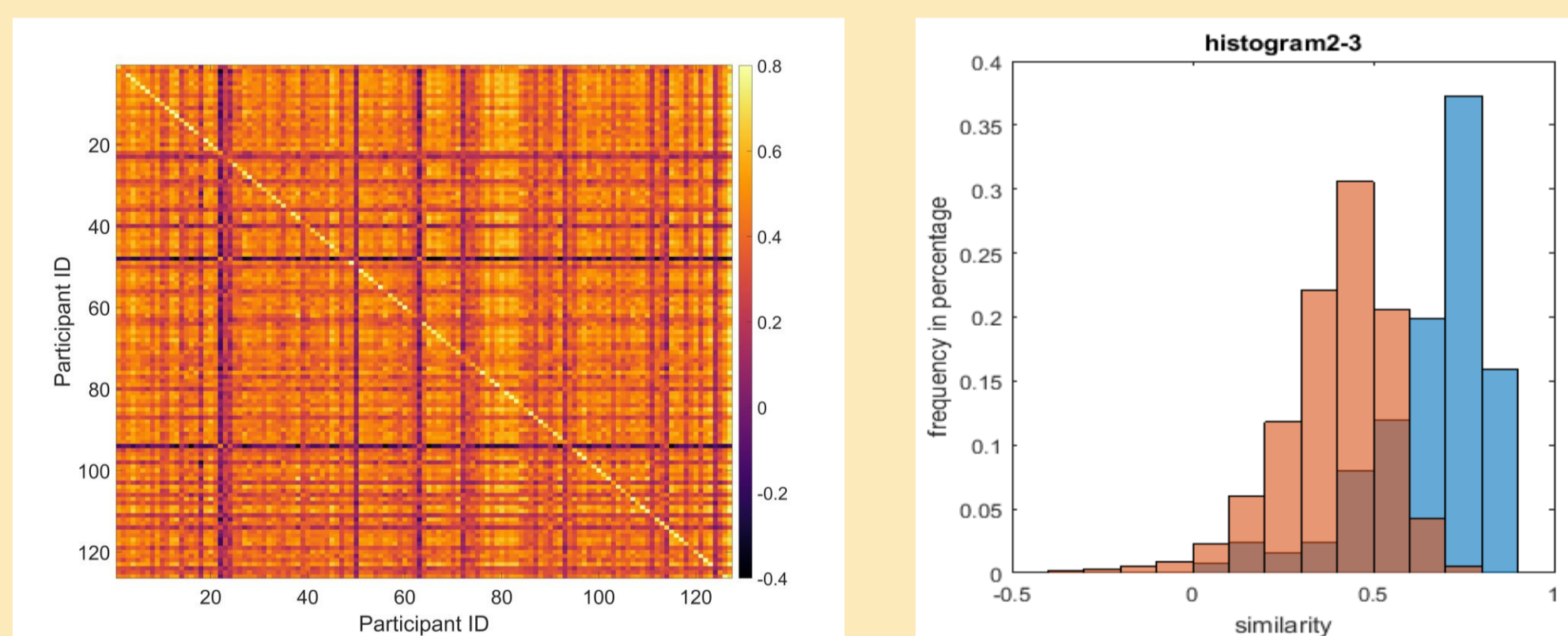
**The toolbox**

**Study Characteristics**

- **Design:**
  - Where is the data?
  - Which subjects? Which session(s)?
- **Contrast(s) of interest**
  - One or two contrasts?
  - Which one?

**Global reliability**

- Whole-brain/within region-of-interest
- **Similarity**
  - Vectorize 3D-matrix for each subject → correlate within (diagonal/blue) and between (off-diagonal/red) subject

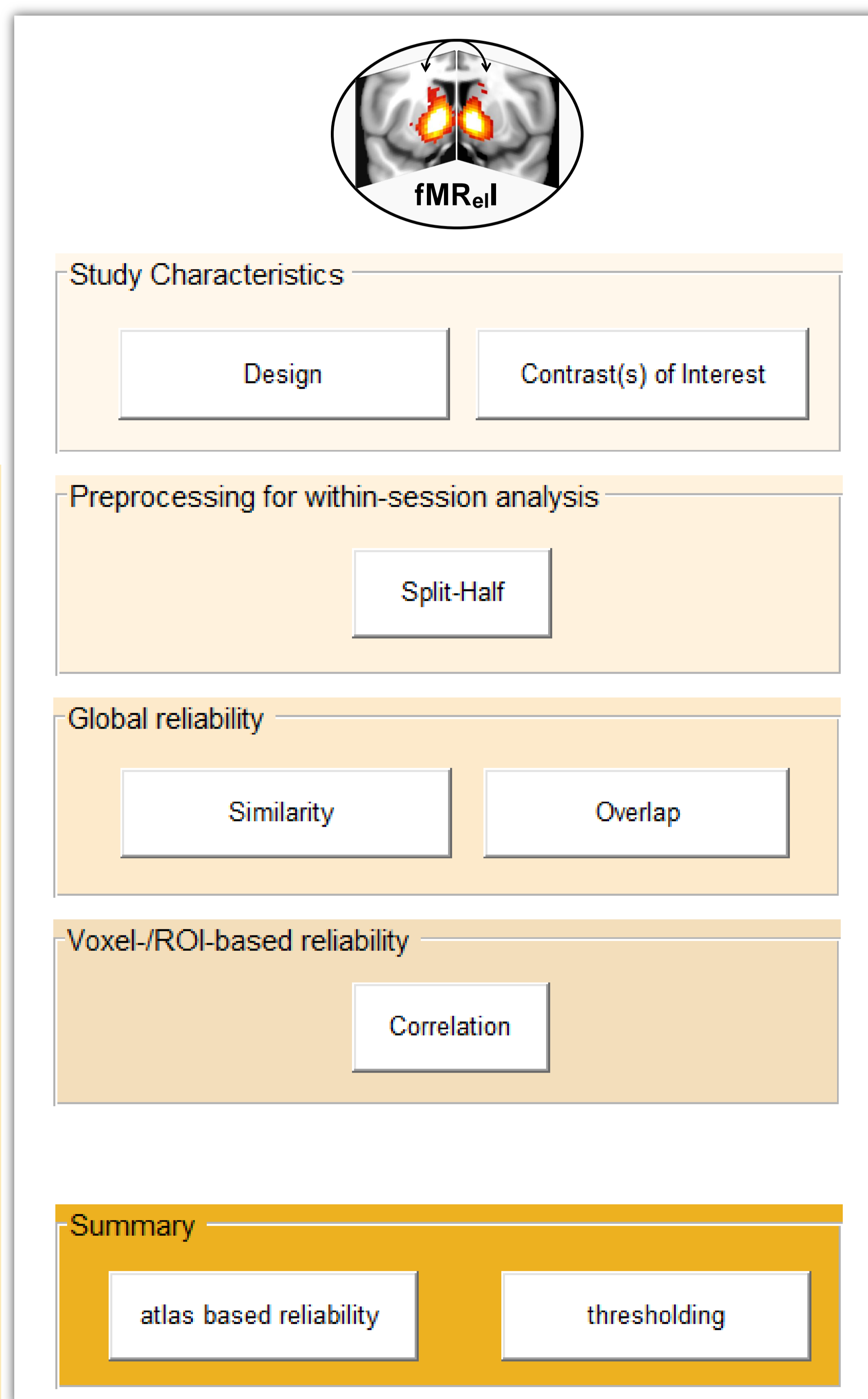


☺ easy comparison of between- and within-subject reliability

• **Overlap**

- Dice and Jaccard coefficient
- Percentage of overlapping significant voxels

Toolbox is available on GitHub  
<https://github.com/nkroemer/reliability>  
 see [6] for methodological details



**Split-half for within-session analysis**

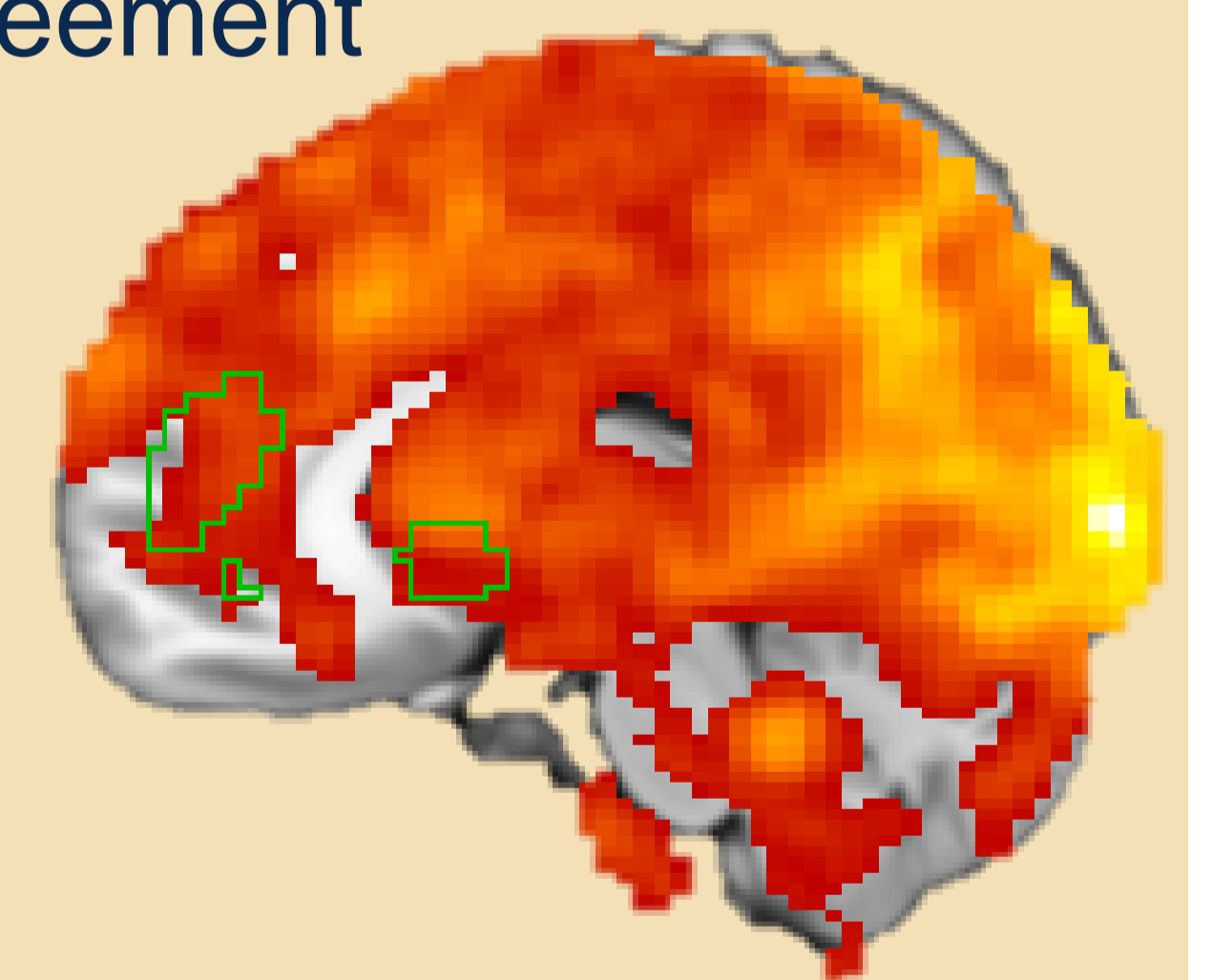
- Definition of contrast/regressor to split
- Split events and corresponding parametric modulators randomly into two parts
- Re-estimation of first-level stats

**Local/voxel-based reliability**

**Correlation coefficients**

- Pearson
- Spearman
- Intraclass Correlation Coefficient (ICC)
  - Absolute agreement → More conservative → Takes into account between-session variance
  - Consistency agreement → More liberal

ICC > 0.2 | 0.7  
 $\emptyset ICC_{ROI} = 0.22$



**Summary**

- **Atlas-based reliability**
  - Summarizes calculated reliability measures for all atlas regions
  - Default atlas: Harvard-Oxford brain atlas and the Automatic Anatomic Labeling (AAL) for the cerebellum (CONN atlas)
- **Thresholding**
  - Set threshold (e.g. 0.5) and keep only voxels with reliability  $\geq 0.5$  in a mask

**References**

1 Bennett, C. M., & Miller, M. B. (2010) <https://doi.org/10.1111/j.1749-6632.2010.05446.x>  
 2 Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., & Pernet, C. (2013) <https://doi.org/10.1016/j.neuroimage.2012.10.085>  
 3 Plichta, M. M., et al. (2012) <https://doi.org/10.1016/j.neuroimage.2012.01.129>  
 4 van den Bulk, B. G. et al. (2013) <https://doi.org/10.1016/j.dcn.2012.09.005>  
 5 Vetter, N. C., Pilhatsch, M., Weigelt, S., Ripke, S., & Smolka, M. N. (2015) <https://doi.org/10.1016/j.dcn.2015.05.001>  
 6 Fröhner, J. H., Teckentrup, V., Smolka, M. N., & Kroemer, N. B. (2017) <https://doi.org/10.1101/215053>

**Notes:**

