COGS138: Neural Data Science

Lecture 18

C. Alex Simpkins, PhD
UCSD Dept. of Cognitive Science, Spring 2023
RDPRobotics, LLC
http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23
rdprobotics@gmail.com | csimpkinsjr@ucsd.edu

Plan for today

- Announcements
- In class hypothesis discussion and refinement
- In class paper
- Single trial analysis, examples and readings
- PCA in neural data science, practical examples

Announcements

- ***<u>Deadlines upcoming this week:</u>***
- **Tuesday**:
- **Wednesday:**
- **Saturday**:
    - Reading Quiz 4 11:59pm
    - In-class paper completion and submission
    - Lecture quiz
- **Friday**:
    - Data checkpoint 11:59pm

Announcements

- Extra office hours for Dr. Simpkins tomorrow (Friday at 12-1pm)
    - Will be available all day either over zoom via quick appointment or piazza/email as always
- We are working on getting the canvas grades updated with the weights, so this week you will be able to check in on that
- Project feedback for the proposal released in your repos as an ‘issue’
- About final presentations
    - Please take the survey about the presentations by Saturday for your input to be considered as part of the final decision for presentation style

Project schedule

Task due	Date due	Description
Previous project review	5/23/2023 at 11:59pm (Tuesday)	Select 2 of the 3 available, review as individuals and then come together as a group to submit your responses to the questions after a discussion. This will orient you to the class project
Project proposal	5/26/2023 at 11:59pm (Friday wk8)	Generate your question, hypothesis, initial data sets you'll be working with, etc., describe your plan, schedule, who is doing what, potential issues, suggested analysis and how it will answer your question
Data checkpoint	*6/2/2023 at 11:59pm (Friday wk9)*	*Builds on the proposal by taking the feedback from PP above and actually getting, loading, describing your data,*
EDA checkpoint	6/9/2023 at 11:59pm (Friday wk10)	Builds on the previous checkpoint, essentially most of your analysis should be done by this point
Final report	6/15/2023 at 11:59pm (Thursday Fin wk)	Due Thursday of finals week so we can grade before the Tuesday deadline, otherwise your grade may be delayed
Group evaluations	6/15/2023 at 11:59pm (Thursday Fin wk)	You will evaluate each other based on participation and performance, this will contribute to your overall final project grade 5%)

Remaining assignments schedule

- A4 wk9-10, A5 extra credit
- R4 wk9
- LQquiz wk 9, 10
- Paper completion this week, mostly in class or via appointment

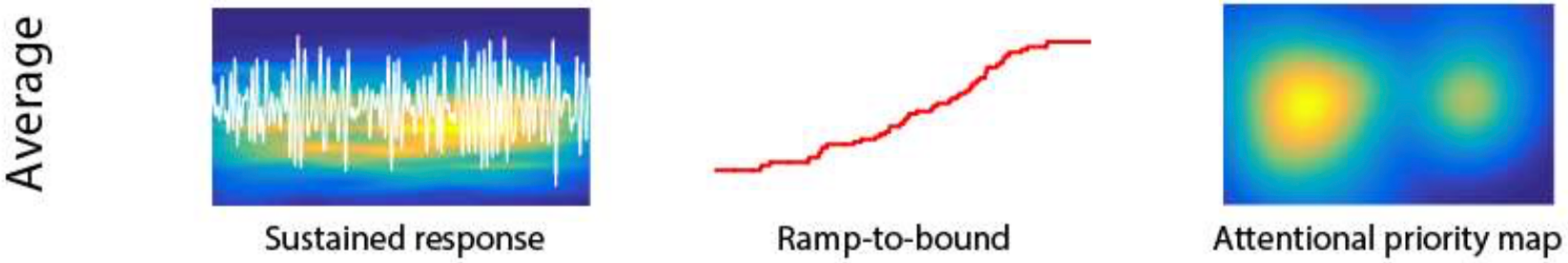
Last time…

Motivation for single trial analysis

- Traditionally neuroimaging techniques are used to compute differences between means over many trials/subjects/studies
    - e.g. in classical cognitive neuroscience, theories of working memory assumed that task-relevant info. is maintained by persistent neural activity
        - Representations kept online by persistent activity patterns, evidence based on averaging massive numbers of subject trials
        - Assumption is if true distribution is contained in noisy measures, measure many times, average, you recover the noise-free representative pattern

So it should look like the following…

- Many single trial measures averaged…
- (Stokes and Spaak 2016)
- Lundqvist,M.etal.(2016) Gamma and Beta Bursts Underlie Working Memory. Neuron



How does the brain really work? Why?

How does the brain really work? Why?

- Brain doesn’t operate according to average response
  - Differences in perception, conscious and unconscious processes, real-world embodied, embedded, situated issues (active perception), encoding, decision making
- Strong evidence for high dimensionality of encoding especially in pre-frontal cortex
  - Rigotti, M. et al (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590
- ***Must*** understand neural dynamics within a single trial
- Consider (Stokes and Spaak 2016)

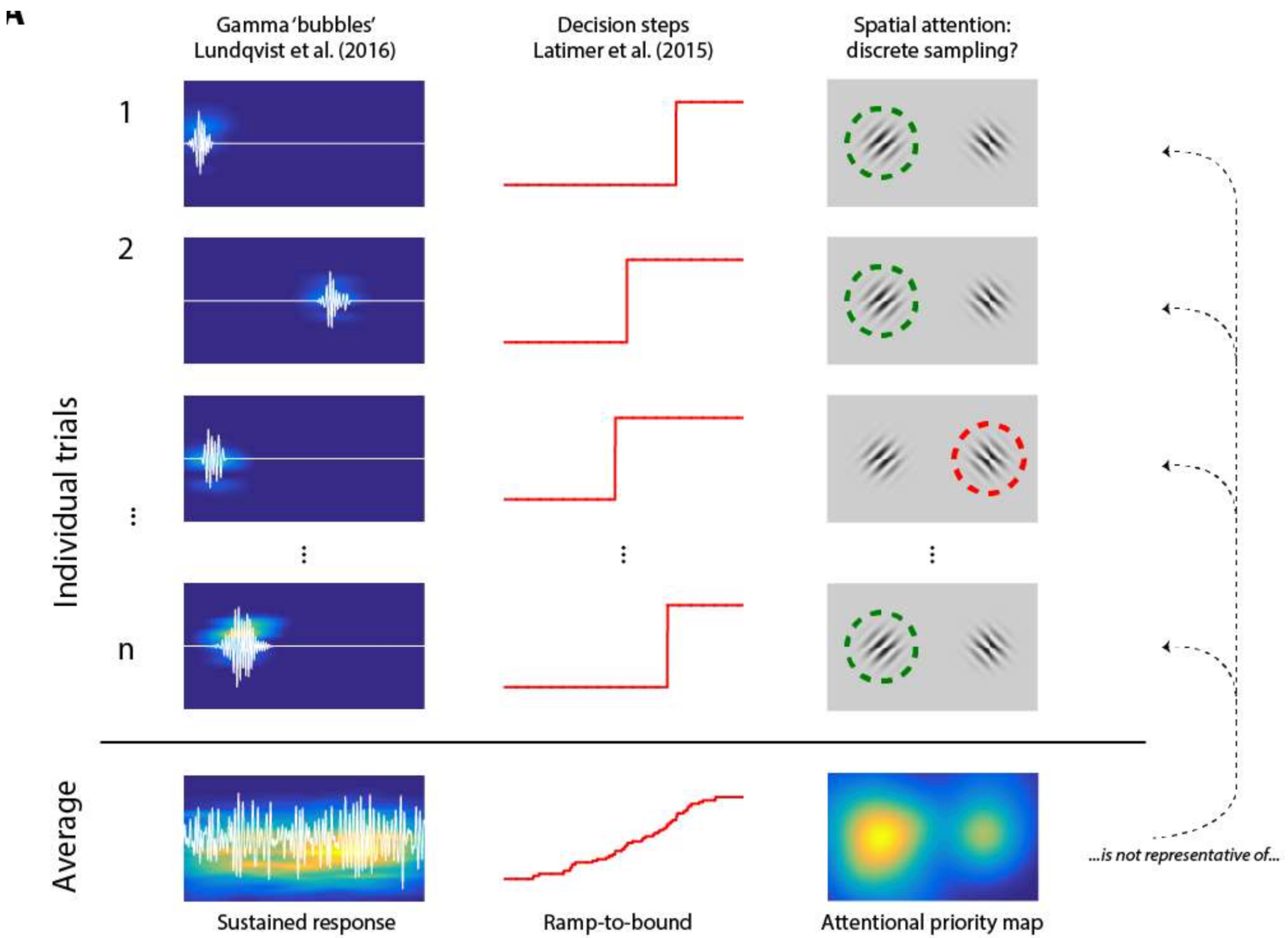
What they tested

- Lunqvist et al. developed novel method to characterize trial-wise dynamics in working memory tasks for primates
- Do we see sustained activity, as previously concluded as recorded from LFP in primate PFC at the single trial?
- Or is it different dynamically at single trial level from ‘average response?’

How did they test it?

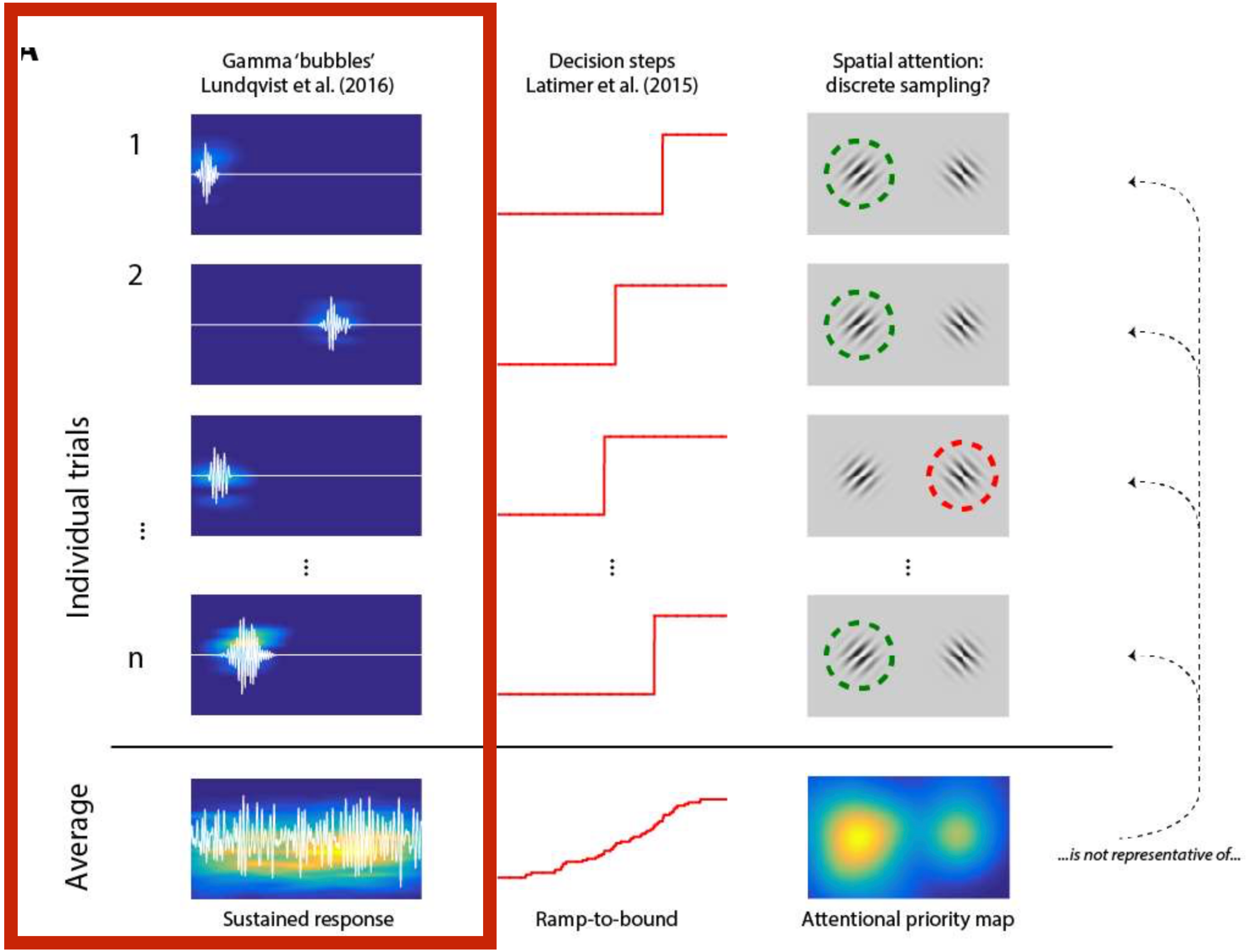
- Developed a novel metric they refer to as ‘burstiness’ to quantify temporal gamma activity for single trials before averaging
- By using 2nd order average of this metric, found persistent activity consists of bursts of activity - not an unbroken chain of firing
- Memories stored in hidden neural states

Combining by computing mean doesn't necessarily create a good representation



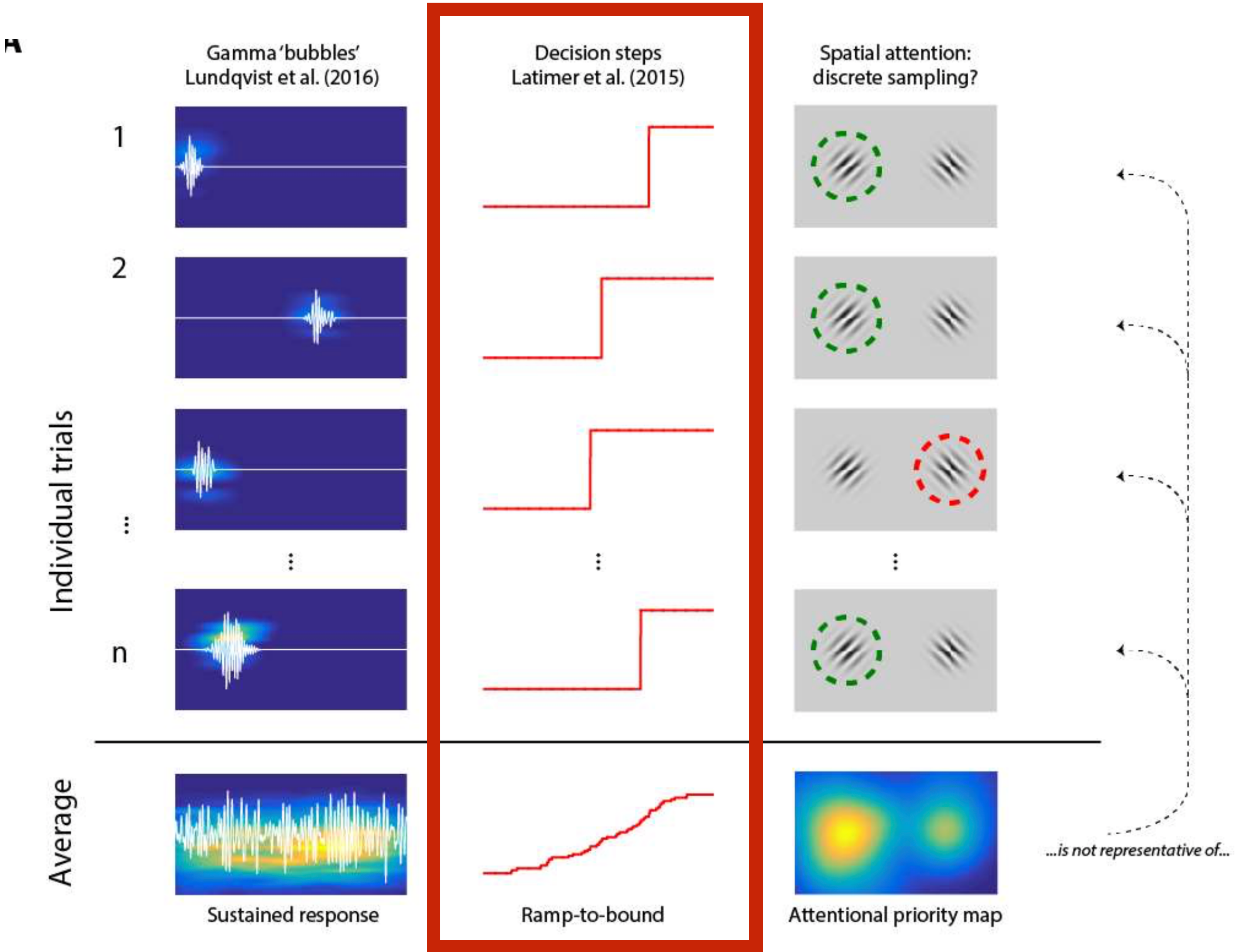
Combining by computing mean doesn't necessarily create a good representation

- Left: (time-frequency representations of power, with traces superimposed) in the prefrontal cortex during individual trials of working memory maintenance activity
- The average shows a familiar sustained gamma response, but qualitatively misrepresents the single trial dynamics.



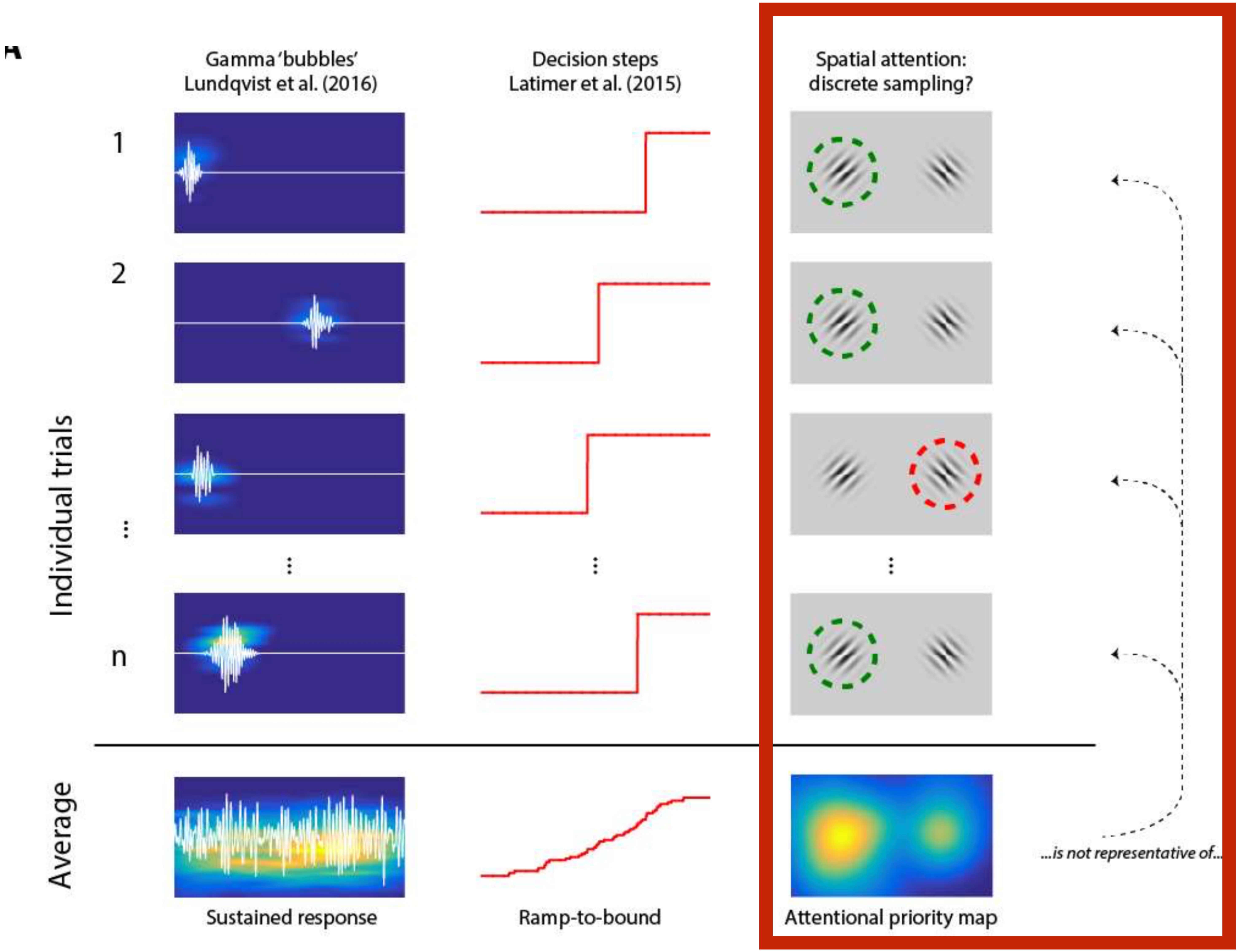
Combining by computing mean doesn't necessarily create a good representation

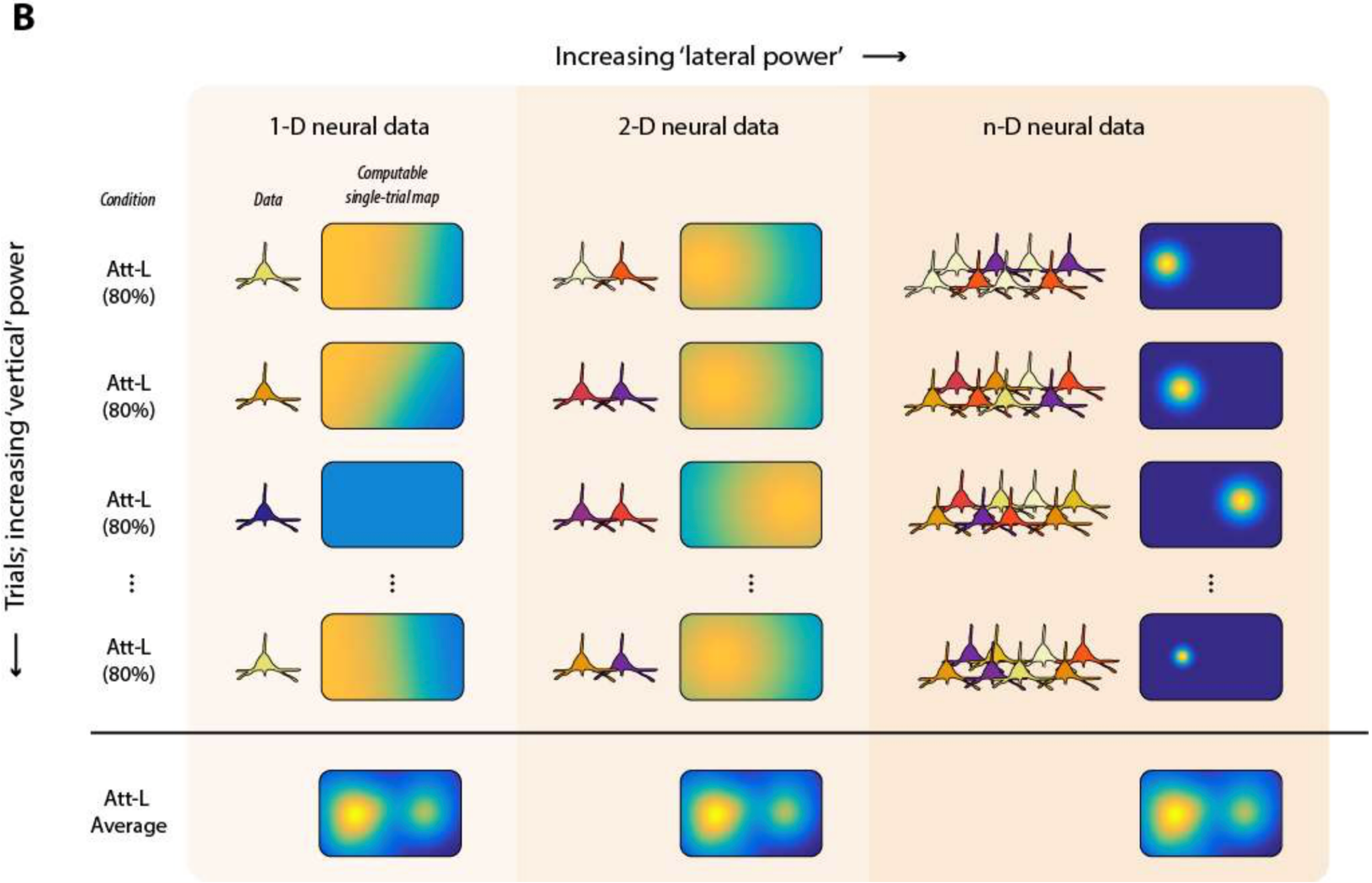
- Middle: neurons display discrete steps reflecting the time of sensory decisions
- The average response shows a classic ramp-to-bound process for the decision.
- Again not representing what's happening



Combining by computing mean doesn't necessarily create a good representation

- Right: spatial attention ***might*** be distributed in a continuous fashion throughout the visual field (as in the average, bottom),
- but such an average profile ***could also be caused by*** individual trials sampling discretely from visual space (80% of trials on the left, 20% on the right).

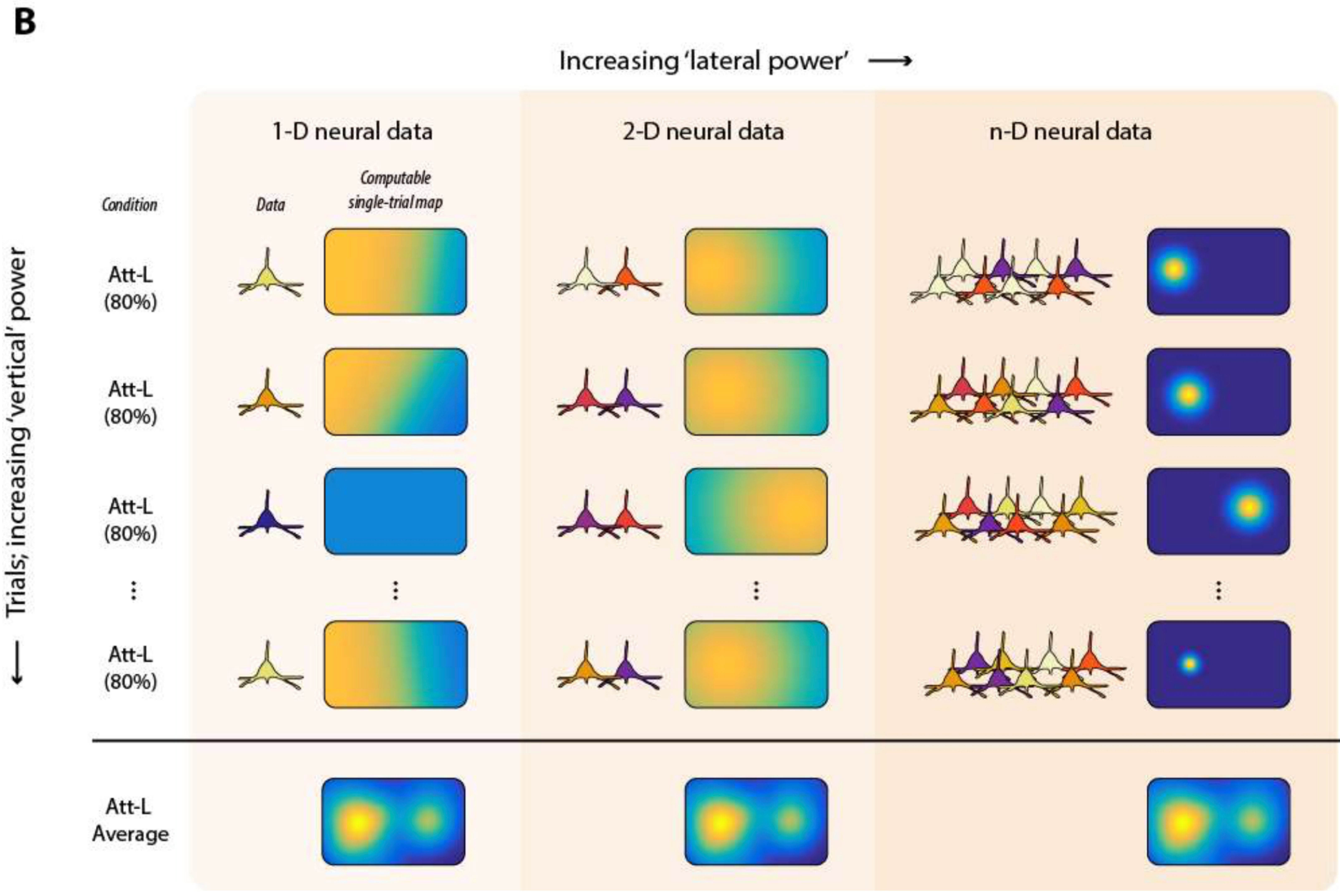



B
Increasing 'lateral power'
1-D neural data
2-D neural data
n-D neural data
Condition
Data
Computable single-trial map
Trials; increasing 'vertical' power
Att-L (80%)
Att-L (80%)
Att-L (80%)
Att-L (80%)
Att-L Average

Combining by computing mean doesn't necessarily create a good representation

- Traditionally **statistical power** = more observations (i.e., trials) to average data
  - "Vertical power"
- **Lateral power:** adding more measurement density (spatial dimension).

Larger lateral power-> Necessary for characterizing neural dynamics in single trial (we'll come back to this)



Motivation for single trial analysis

- We can thus miss important variability patterns by collapsing to the mean
  - Consider a large classroom with 500 students all talking while waiting for lecture.
  - If we record their conversations even for the same class each day all quarter then average the recording, will we recover what the individuals said?
  - No it doesn’t ‘average’ to the conversation as each day has subtle differences
- <u>Solution</u> -**Single trial analysis** studies variability across trials

Single Trial Analysis allows for systematic mapping between

- Brain activity and stimulus information space
- Brain activity and subject behavioral variability
- Brain activity measured using multiple imaging techniques (EEG, fMRI etc)

Single Trial Analysis definition and classes

- All methods that consider ***variance within subjects***
    - 2 classes of methods

        1. Univariate methods

        2. Multivariate methods

- Applications - *Useful for both behavioral and neuroimaging experiments*

STA - Univariate methods

- **Regression over all trials in single subjects measures the relationship between parameterized stimulus space and signal amplitude**

- fMRI - “Parametric design”

- EEG - using same type of approach

    - Neural response to stimulus in individual subjects

    - Probabilistic mapping between stimulus information and EEG amplitude
        - Rousselet G. A., Gaspar C. M., Wieczorek K. P., Pernet C. R. (2011). Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Front. Psychol.* 2:137. 10.3389/fpsyg.2011.00107

    - Time-frequency decomposition of power and phase
        - Cohen M. X., Cavanagh J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Front. Psychol.* 2:30. 10.3389/fpsyg.2011.00030

STA - Univariate methods

- **Variance among trials contains info regarding subjects and their cognitive states**
  - e.g. study establishing increased variance in latency of P1 response to Gabor patches than controls for children with autism
    - Milne E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Front. Psychol.* 2:51. 10.3389/fpsyg.2011.00051
  - e.g. pre-stimulus alpha power correlated with subject judgment of state of attention
    - Macdonald J. S. P., Mathan S., Yeung N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Front. Psychol.* 2:82. 10.3389/fpsyg.2011.00082
    - *(review)* VanRullen R., Busch N. A., Drewes J., Dubois J. (2011). Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Front. Psychol.* 2:60. 10.3389/fpsyg.2011.00060

STA - Multivariate methods

- **Often derive pattern classifiers to characterize the spatial-temporal variance in each trial**
- e.g. Touryan J., Gibson L., Horne J. H., Weber P. (2011). Real-time measurement of face recognition in rapid serial visual presentation. Front. Psychol. 2:42. 10.3389/fpsyg.2011.00042
  - Used variance in time/space to train a discriminant function that could classify brain activity associated with familiar/unfamiliar faces in real-time
  - Group ERPs could be used to differentiate over frontal AND parietal electrodes, but with the above methods, ONLY parietal response allowed categorical discrimination on single-trial basis
  - So averaging can actually create misleading illusory signals that are not actually present in individual subjects!
    - Gaspar C. M., Rousselet G. A., Pernet C. R. (2011). Reliability of ERP and single-trial analyses. *Neuroimage* 58, 620–629 10.1016/j.neuroimage.2011.06.052

STA - An additional dimension

- Allows for interpretation of individual differences to quantify effects within and between subjects
- An additional window into brain function
- Rich data description can help expose subtle brain mechanisms that may be hidden when looking at traditionally pooled data (averaged)

STA - caveats -> Requirements

- <u>Many trials (“Vertical power”)</u>
  - To reduce Signal to Noise Ratio - regression over trials
  - Have to be careful not to smooth over important heterogeneity
  - Metrics were developed of ‘burstiness’) - needs a priori model
- <u>Dense coverage (“Lateral power”)</u>
  - For good patterns - time/frequency intervals, localization, avoiding missing spikes in activity, sparse behaviors, etc
- e.g. Rousselet G. A., Husk J. S., Bennett P. J., Sekuler A. B. (2008). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 3, 1–18 10.1167/8.12.3

STA - toolboxes

- **Recipes**
  - Parra L. C., Spence C. D., Gerson A. D., Sajda P. (2005). Recipes for the linear analysis of EEG. *Neuroimage* 28, 326–341 10.1016/j.neuroimage.2005.05.032
- **PyMVPA (http://www.pymvpa.org)**
  - Hanke M., Halchenko Y. O., Sederberg P. B., Hanson S. J., Haxby J. V., Pollmann S. (2009). PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7, 37–53 10.1007/s12021-008-9041-y
- **EEGLAB, SIFT, NFT, BCILAB, and ERICA**
  - Delorme A., Mullen T., Kothe C., AkalinAcar Z., Bigdely-Shamlo N., Vankov A., Makeig S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput. Intell. Neurosci.* 2011, 130714.

STA - more toolboxes

- Hartmann T., Schulz H., Weisz N. (2011). Probing of brain states in real-time: introducing the console environment. *Front. Psychol.* 2:36. 10.3389/fpsyg.2011.00036
- FieldTrip
  - Oostenveld R., Fries P., Maris E., Schoffelen J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell Neurosci.* 2011, 156869.
- Pernet C. R., Chauveau N., Gaspar C., Rousselet G. A. (2011). LIMO EEG: a toolbox for hierarchical linearmodeling of electroencephalographiuc data. *Comput. Intell. Neurosci.* 2011, 831409.

In class report development (~30m)

- Define this course's intent
- Draw comparisons between this course and requirements
- How does this course build upon what came before?
- How can you use your starting point in this course to expand your understanding?

On to today…

Refining the hypotheses

A hypothesis should be

- Narrow
- Very specific
- ***<u>Not</u>*** include a conclusion or interpretation
- Consist of a research and null hypothesis
- Remember we are trying to reject or fail to reject the null, which basically says we either
  - ***'didn't find anything' or***
  - ***'failed to <u>not</u> find anything'***

Developing a hypothesis - overview readings to review

- https://www.scribbr.com/statistics/hypothesis-testing/

- https://opentext.wsu.edu/carriecuttler/chapter/developing-a-hypothesis/#:~:text=A%20researcher%20begins%20with%20a,prediction%20is%20called%20a%20hypothesis.

- https://www.skillsyouneed.com/num/hypotheses-testing.html

- https://www.nedarc.org/statisticalhelp/advancedstatisticaltopics/hypothesisTesting.html

- https://www.youtube.com/watch?v=joNb67F1UbY

Hypothesis : Simplicity, narrowness

- KISS principle
- Boiled down to the essence of the relationship you are testing
- Research/Alternative and Null are opposites

Hypothesis testing

-Cannot prove hypothesis

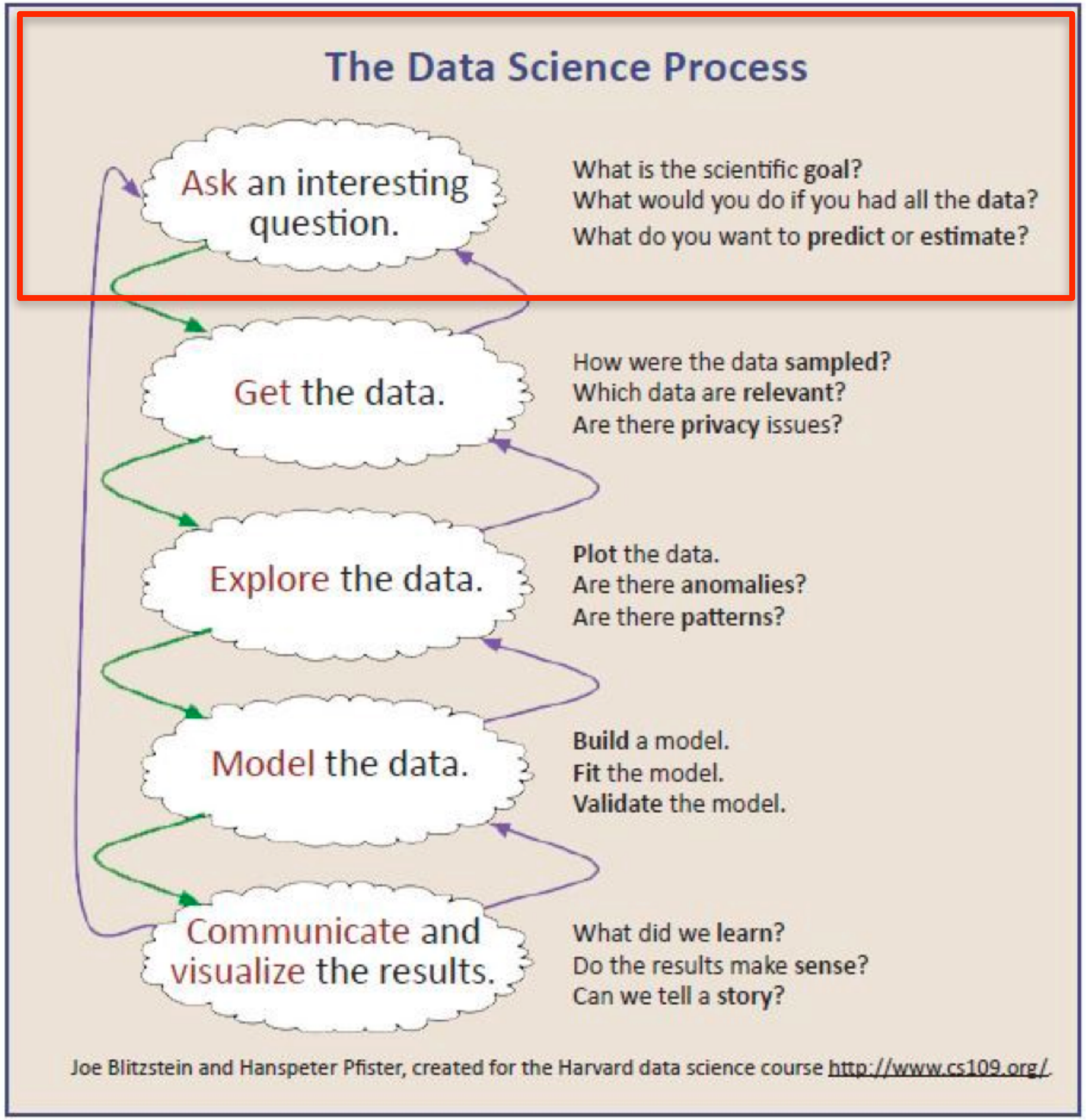
-Can only reject or fail to reject null hypothesis

-Why?

Data Science questions should...

- Be specific
- Be answerable with data
- Specify what's being measured

What makes a question a good question?



adapted from Chris Keown

Working toward a strong data science question

Vague: How does the brain change when you have a brain injury?

Better: What neurological changes are there after a stroke?

Even better: What neurological and behavioral changes can be measured with EEG and motion capture between an average normal subject and a stroke patient who had a recent stroke that impaired motor function?

Best?