Using Correspondence Theory of Semantic Information to Understand Neural Representational Similarity

Wojciech Mamak

PhD Candidate
Section of Logic and Cognitive Science, Polish Academy of Sciences

Bulgarian-Polish Workshop
'Significance of Logical Models to Scientific Practice'
Institute of Philosophy and Sociology, Bulgarian Academy of Sciences
Sofia, 23rd Nov 2022

Quick overview

1. Intro: informational explanations in cognitive science
2. Information as correspondence - Correspondence Theory of Semantic Information (CTSI)
   a. Swoyer's surrogative inference
   b. Barwise and Seligman's formalization of information-flow
3. Representational Similarity Analysis (RSA) in neuroscience
4. Does it fit?
5. Closing remarks

Why bother?

- Dretske's (1981) recipe for naturalistic theory of representation - start with understanding information
    - **"Information is a rung on a ladder that gets one to representation. Representation is information, or representation is information plus some other ingredient" (Sprevak 2020: 591)**
- Traditionally and still most commonly interpreted as Shannonian information.
- **Run-of-the-mill working view in neuroscience: information is assumed to be correlational yet semantic (operationalized as selective responding to a particular set of stimuli).**
- The ongoing philosophical quest for content
    - Teleosemantic theories [Neander 2004 & 2012; Cao 2012]
- Purported possibility of readout and prediction ('mindreading') [Ritchie et al, 2019; Poldrack 2018]
- Rathkopf's '**What kind of information is brain information?** [2020a]
    - how objective can it be if it is receiver-relative [2020b]
- Applications in neural representation theory (Bechtel 2016, Thomson & Piccinini 2018)

Some worries for traditional accounts

- Theoretical landscape is competitive: Shannonian, algorithmic, teleosemantic, transmission, effective etc.
  - Not clear which are in play in empirical practice (in psychology, cognitive science, biology, robotics, etc.)
- Assumed to be roughly Shannonian (quantifiable), but oftentimes simultaneously treated without further examination as ‘cognitive subject matter’. But how do you then get semantics?
- In fact, more strictly just a correlational or covariational relation
  - Informational as long as it shifts probabilities about target
  - E.g. *mutual information*
- Binding is (more-or-less) arbitrary - insofar there are degrees of freedom it suffices
- Coding does not depend on similarity between the relata (vehicle and target)

CTSI (Miłkowski 2021)

- Contrary to received views, CTSI is not primarily reliant on correlation, covariation, causation, natural laws, or logical inference
- Formalizes a somewhat paradoxical aspect of information-as-similarity that it is both omnipresent in common usage but theoretically unanalyzed
- Recipe: take Tarskian correspondence with structural similarity *and shake vigorously*
  - Information is true iff corresponds to what it is about,
  - Structural similarity is a crucial resemblance between the structure of representation vehicles and representation targets that affords representation vehicles' their content
- CTSI – in search for measurement.
  - Representational geometry possibly has some ideas!

Swoyer's surrogative inference (1991)

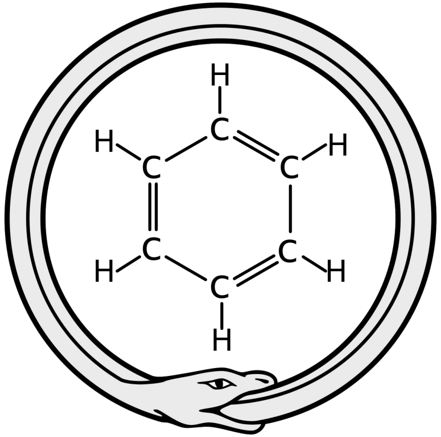
- Licensed by similarity between A and B
- Allows to draw inferences about B when only A is available
- A carries information about B - similarity is conceptualized as an *information-flow*
- Relations are preserved and counter-preserved

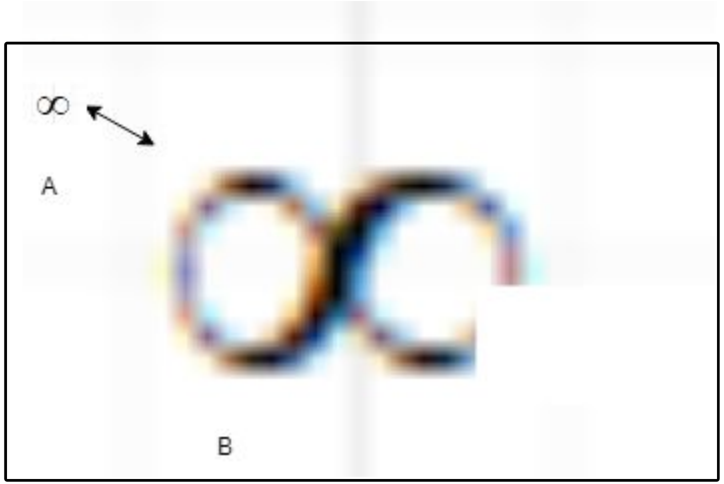
- $CTSI{:}\ r\text{-}sim\ (A, B) \equiv [r(A) \twoheadrightarrow r(B)]$
    - Note: unidirectionality (implication instead of a biconditional)

Swoyer, C. [1991]: 'Structural Representation and Surrogative Reasoning', *Synthese*, **87**, pp. 449–508.

Extraordinary examples!

- The dictator’s doppelganger!
- The cyclical snake!
- Even infinity itself!



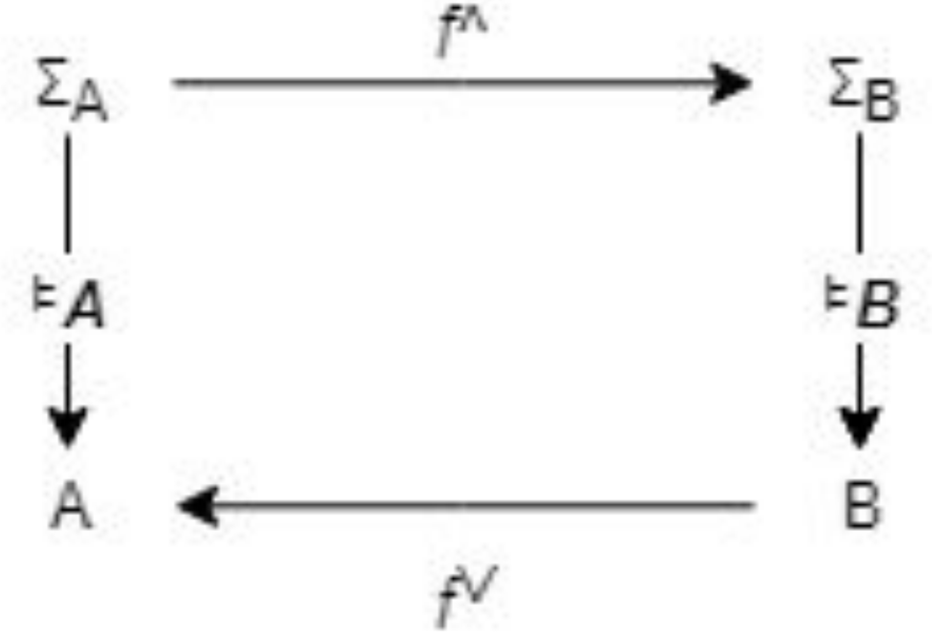


Infinity sign taken from Miłkowski (2020),
Others mined from the depths of Google search engine

Barwise and Seligman's (1997) logic of information flow

Infomorphism is a pair of *f* functions $\langle f^\wedge, f^\vee \rangle$. For two classifications, $\boldsymbol{A} = \langle A, \Sigma_A, \models_A \rangle$ and $\boldsymbol{B} = \langle B, \Sigma_B, \models_B \rangle$ an informophism *f* obtains if and only if $f^\vee(b) \models_A a \equiv b \models_B f^\wedge(a)$ for all tokens *b* of *B* and all types *a* of *A* (1997:28)

This definition is weakened for practical reasons (e.g. gradability) into the relation of **infocorrespondence** defined over fuzzy sets.



Barwise, J. and Seligman, J. [1997]: 'Information Flow: The Logic of Distributed Systems', Cambridge; NY: Cambridge U. Press.

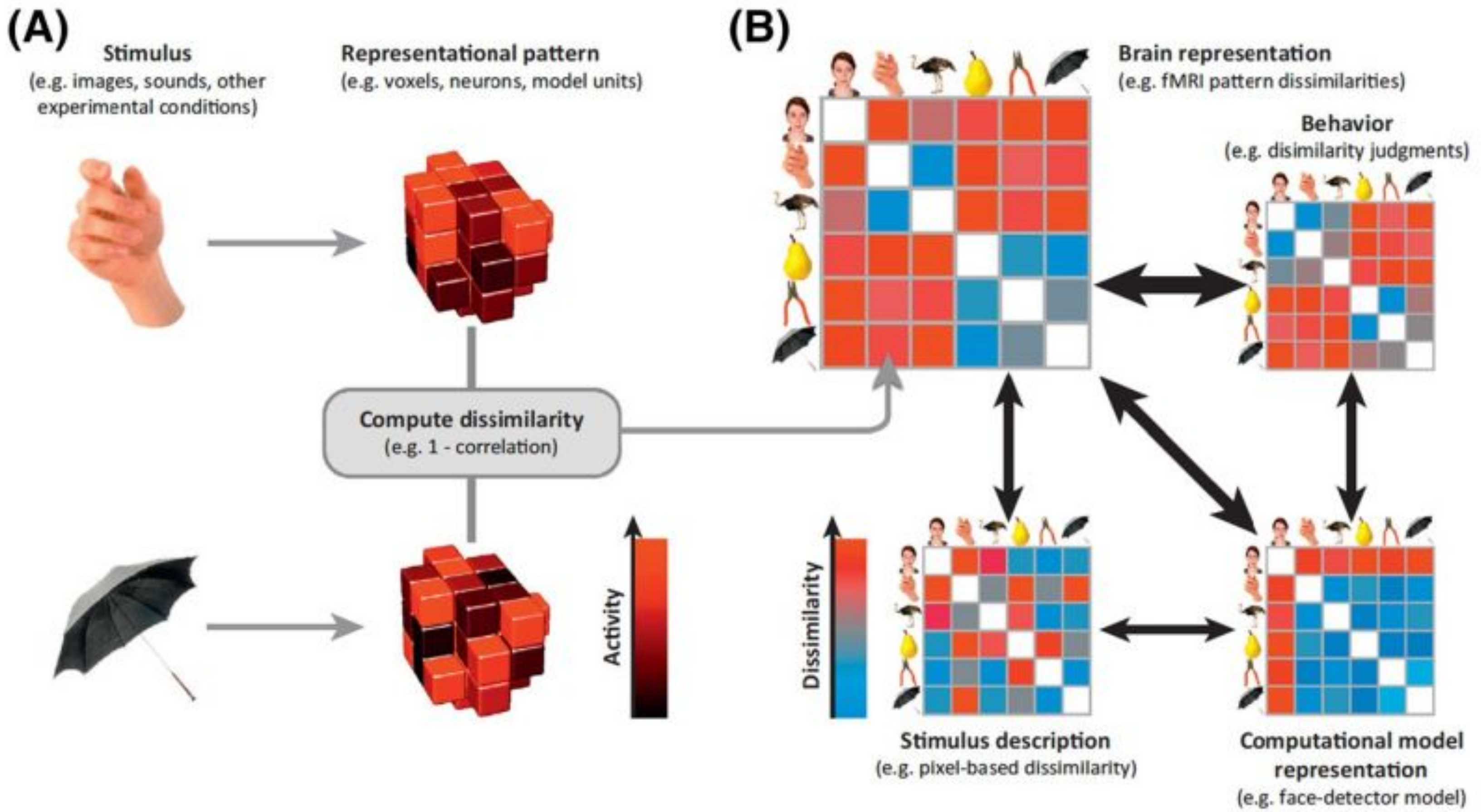
Representational Similarity Analysis (Kriegeskorte 2008)

- A type of MVPA - multivariate pattern analysis [Haxby 1991]
- First proposed by N. Kriegeskorte in 2008 to compare different representational geometries
- Aimed explicitly at identifying informational processing
- Philosophically analyzed recently by Roskies (2021)
- A two step procedure

RSA - Step One: Constructing the Representational Dissimilarity Matrix (RDM)

- Firstly, patterns of brain activity during different experimental manipulations are encoded as vectors
- Distance metric is used to calculate the distance (Euclidean, angular, etc.) between pairs of vectors.
- These distance measurements comprise a **representational dissimilarity matrix, or RDM**, of the activity in that task.
- The individual entries in an RDM are thus numbers characterizing the dissimilarity of the brain activity patterns across the two conditions, yielding a diagonalized matrix.
- As RDM entries are scalar quantities, they are **format-neutral** and abstract from particularities of their provenance, such as spatial layout

RSA - First step



RSA - Step Two: Meta-Analysis

- In this phase, **second-order comparisons** between different RDMs are performed
- Different measurements over the same experimental conditions
- Allows comparisons between sets of measurements that differ in substantial ways.
- Provides a means of comparing structural relationships embedded in multivariate data to the structural relationships in other datasets
- **It is this second-order comparison between RDMs that makes RSA so versatile**
- Turns out a **lot of structure is retained** throughout representational geometries coming not only from different areas or modes of measurement, but also different subjects or even species
- **<u>Strong evidence that coding in representational schema is not arbitrary, but organised according to similarity-based information</u>**

RSA - Second step

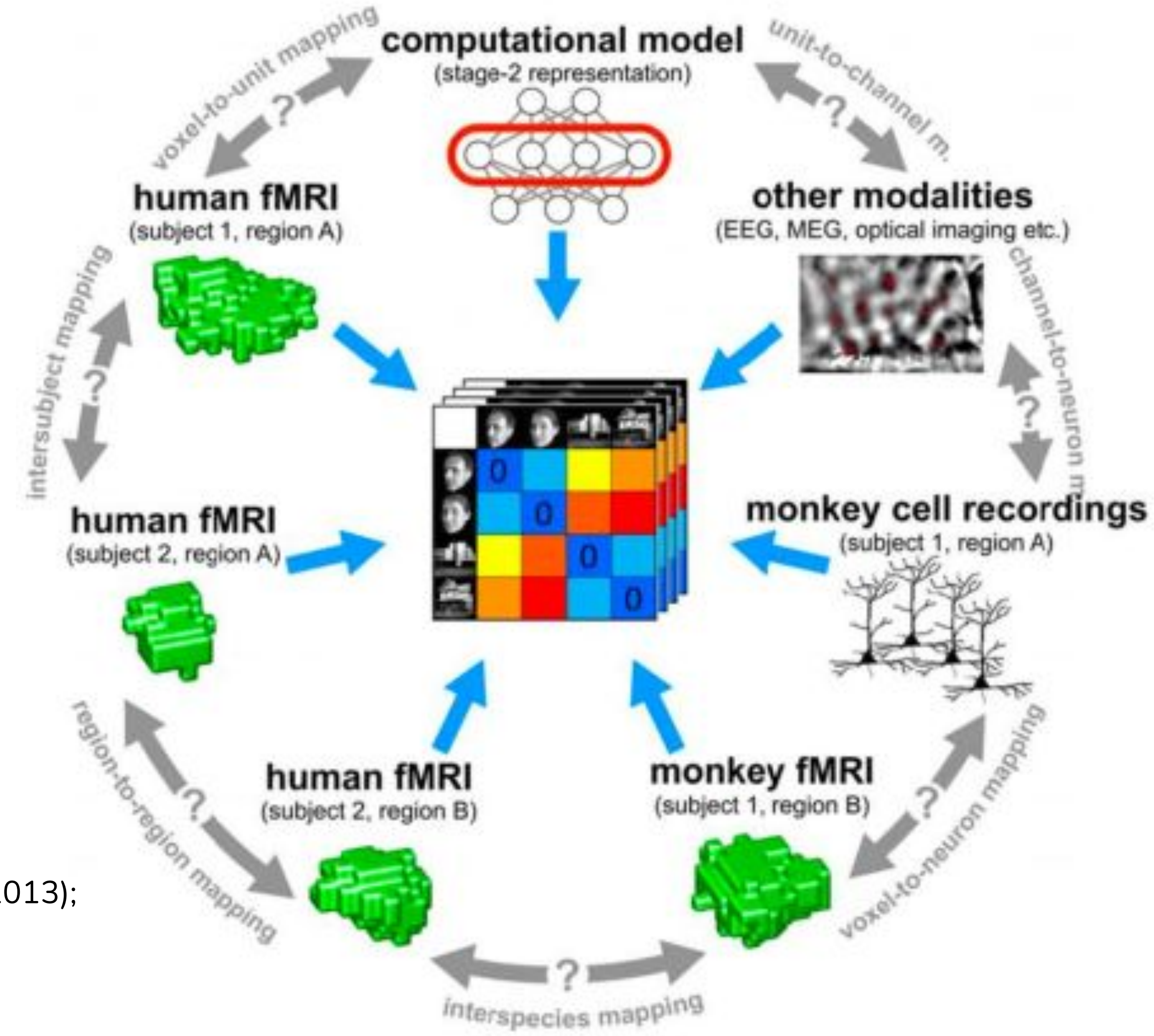
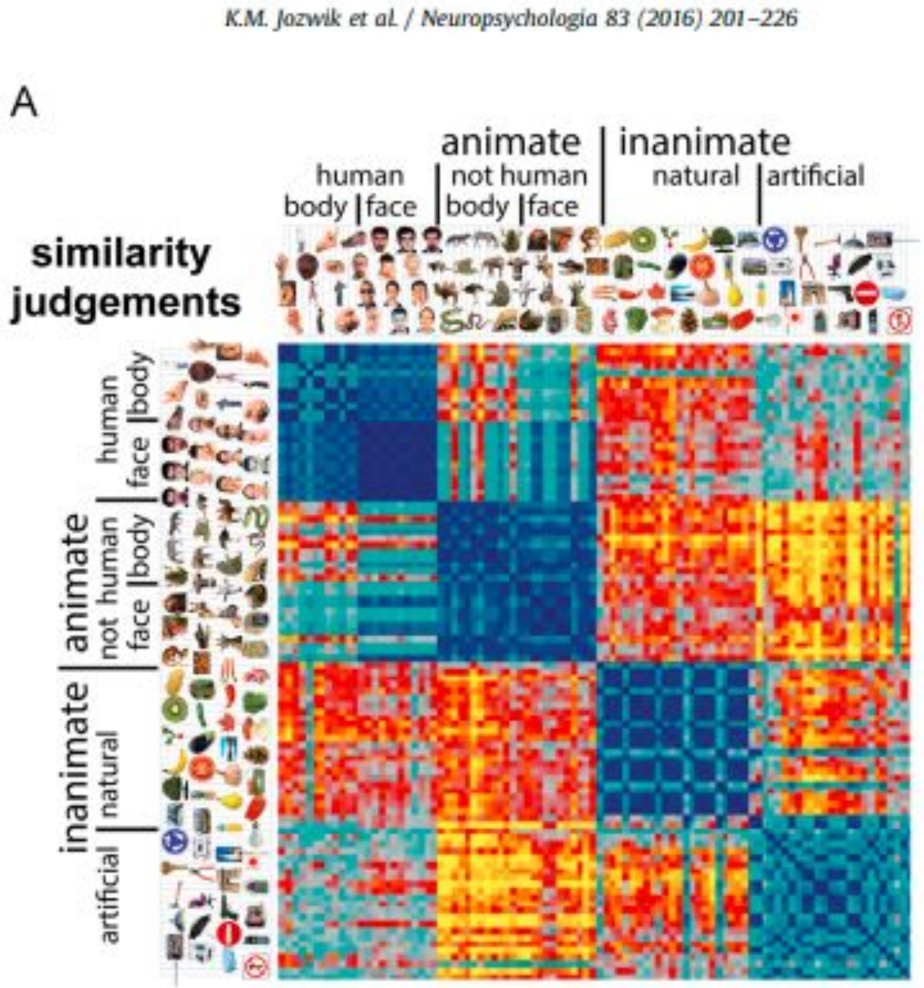


Diagram comes from (Kriegeskorte 2008);
previous slide's from (Kriegeskorte & Kievit 2013);
both adapted from (Roskies 2021)

Retainment of categorical structure in clusters



To recapitulate the main points

- **T1**: There is an infocorrespondence relation both between the representational geometries themselves [G–G] and geometries and the external world [G–W] (pairwise).

- **T2:** This relation is precisely what grants semantic information flow.

- **T3:** The existence of semantic information better explains the retainment of relation of properties between the representational geometries that the theory of arbitrary coding (i.e. Shannonian) struggles with.

- **T4:** Structural similarity licences surrogative inferences on the researchers (external observer) part.

Structural information in RSA - a lot to like?

(or at least a bit?)

- Correspondence component: representations in the matrices have **satisfaction conditions**, insofar they accurately inform about their targets
- It is **through the second step** that structure is preserved not due to correlational information, nor causal processes (each first-order RDM can, and usually does, follow different causal path)
- The meta-level component that makes RSA so versatile and powerful comes from **format-friendliness** i.e. the internal structural features are the ones that count
- A **structure-preserving relation** holds (retainment of core categorical features), but there is some counter-preservation, too (similarity score is almost never = 1)
- It allows researchers to draw inferences about *semantic* properties from one representational geometry to another (surrogative inference)
- RDMs code gradable properties (from 0-1 range) ↠ according to the infocorrespondence formalization of information flow
    - one RDM is informative of another, actively and successfully used by neuroscientists for prediction of neural data)
- Tentatively, measurement in conceptual geometry can provide **quantificational** aspect to CTSI

Some potential objections and open problems

- **An abstract generalized account only? Yes, but it's a feature. not a bug.**
- **Unidirectionality of bidirectionality? Symmetric or asymmetric similarity?**
- What about transitivity?
- How arbitrary is the correspondence? Issue of too many correspondences (Kotarbińska, Goodman)
    - Pragmatically-decided grounding has to be assumed to limit which types of similarities one is interested in when inferring surrogatively
- Aren't the similarities spurious?
- **Is such neural information explicitly represented *for* the system? Can we compare similarity geometries to cognitive maps (such as place cell grid)?**
- **What 'uses' information? Who is a consumer?** Next layer, different voxel, region, personal level operator, or maybe merely an external observer (researcher)?
- Information =/ representation. Additional constraints needed in practice.

Selected bibliography

- Barwise, J., & Seligman, J. (1997). *Information flow: The logic of distributed systems*. Cambridge University Press.
- Cao, R. (2012). A teleosemantic approach to information in the brain. *Biology & Philosophy*, *27*(1), 49–71
- Dretske, F. I. (1981). *Knowledge and the Flow of Information*. MIT Press
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4.
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in cognitive sciences*, *17*(8), 401–412
- Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British journal for the philosophy of science*, *70*(2), 581–607.
- Miłkowski, M. (2021). Correspondence Theory of Semantic Information. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/714804
- Roskies, A. L. (2021). Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations. *Synthese*, 1–19.
- Sprevak, M. (2020). Two kinds of information processing in cognition. *Review of Philosophy and Psychology*, *11*(3), 591–611.
- Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*(3), 449–508.

Thank you for your attention!

For follow-up questions, do not hesitate to contact me: wojciech.mamak@gmail.com