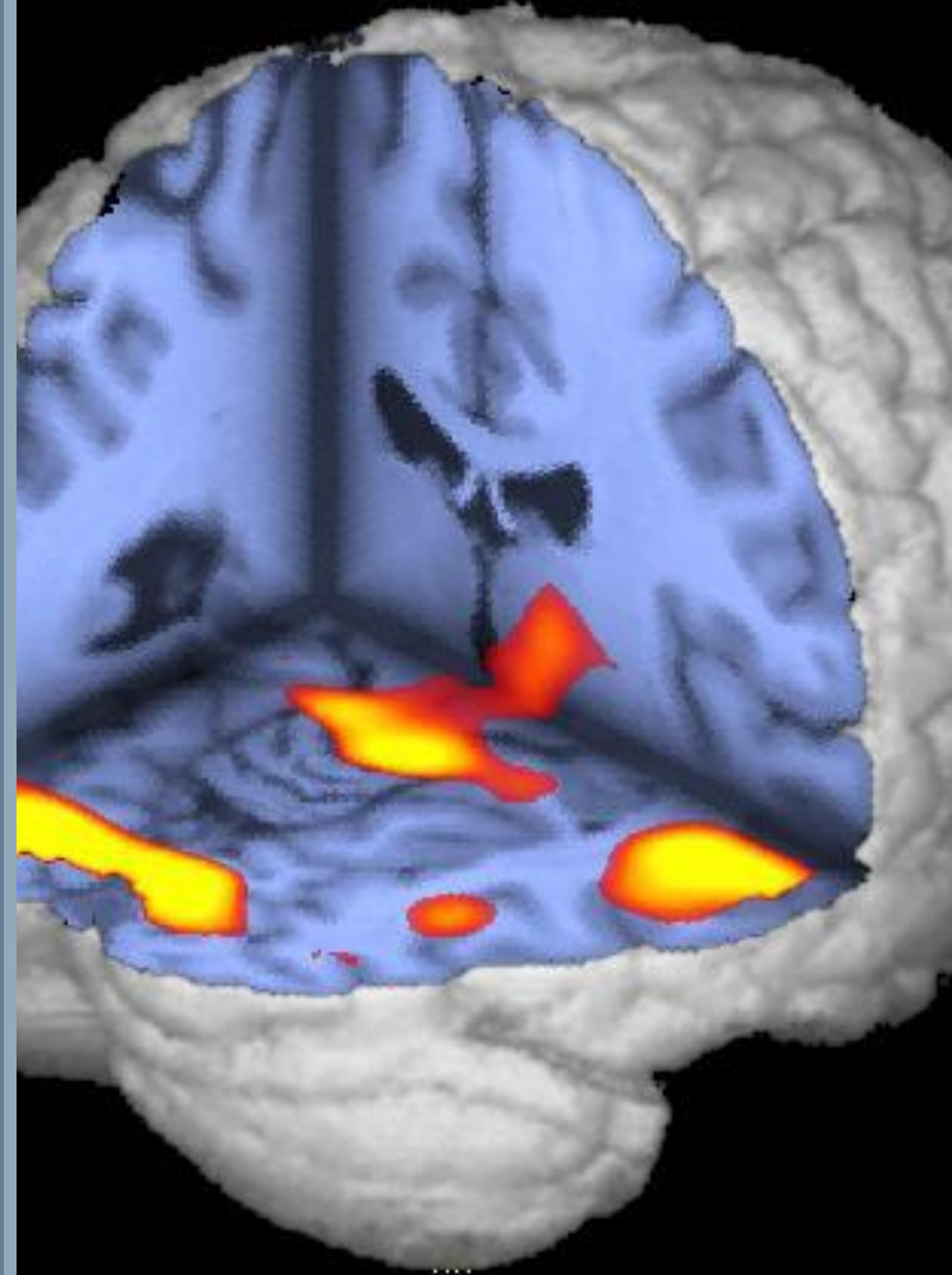# Putting The Best of Human Nature into Natural-Artificial Intelligence: A Brain-Based Approach

David Silbersweig MD

Department of Psychiatry
Center for the Neurosciences

Brigham and Women's Hospital
Harvard Medical School

Brain, Mind and Computers

"And men ought to know that from nothing else but thence [from the brain] come joys, delights, laughter and sports, and sorrows, griefs, despondency, and lamentations.

And by this… we acquire wisdom and knowledge, and see and hear and know what are foul and what are fair… And by the same organ we become mad and delirious, and fears and terrors assail us…"
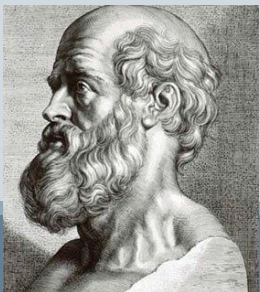
Hippocrates (460-377 BCE)

FUNCTIONALISM

– Computational Functionalism
  • Representational Theory of the Mind (Jerry Fodor - 1975, Stephen Stich)
    – Knowledge of the world is embedded in mental representations
    – Mental representations are symbols (the "language of thought" or "mentalese")
    – Symbols can be "computed"
    – The mind is endowed with a set of rules to operate on such representations, i.e. to "compute" the symbols
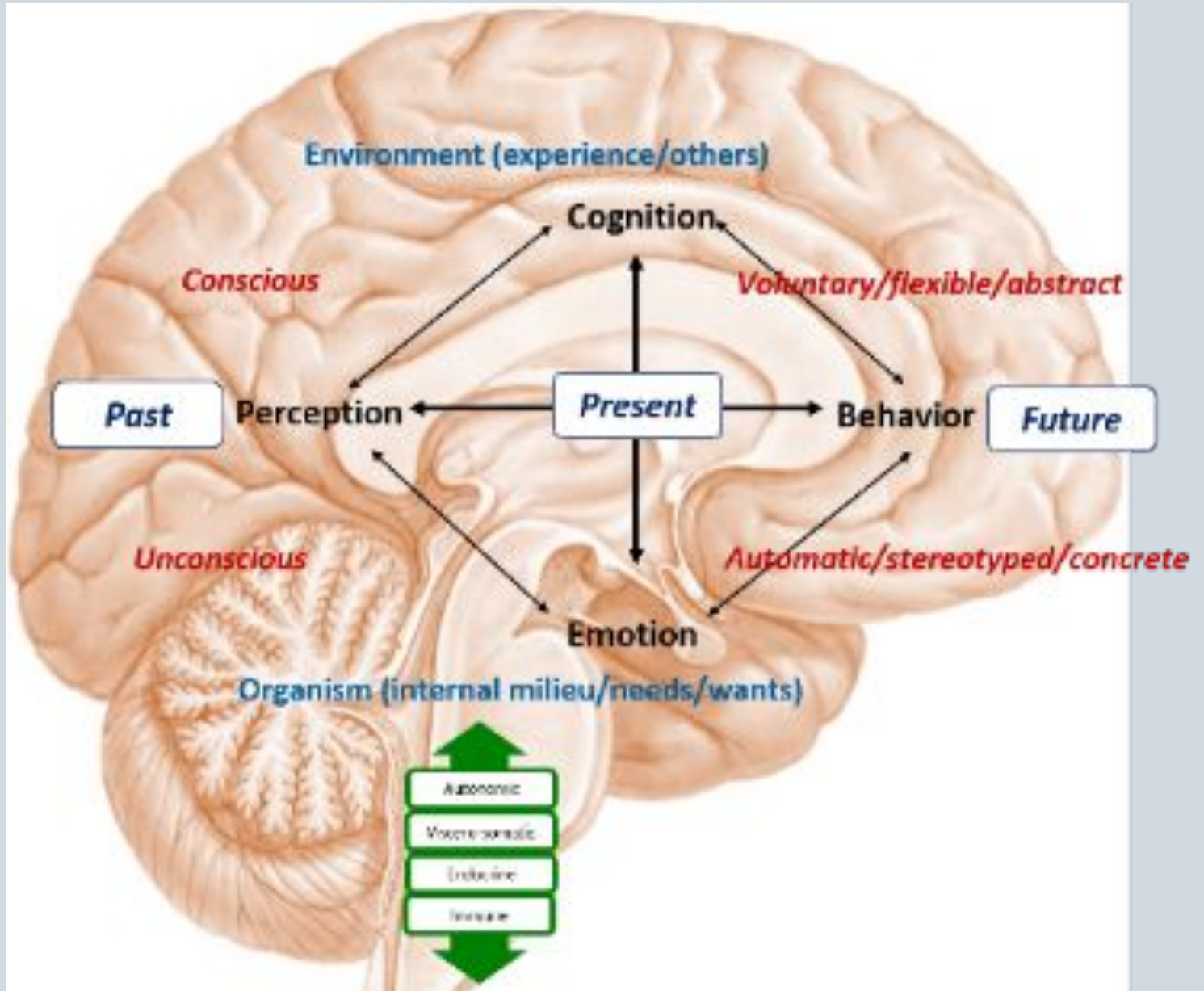   – Cognitive life is the output of those rules

Translational Methods

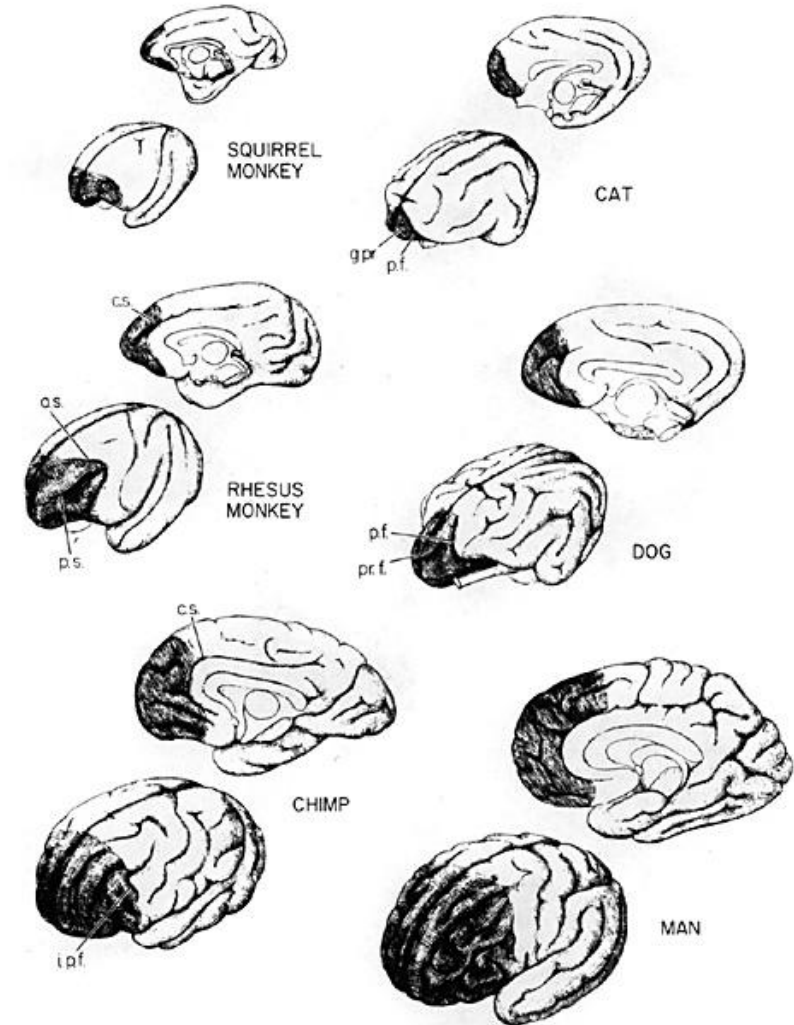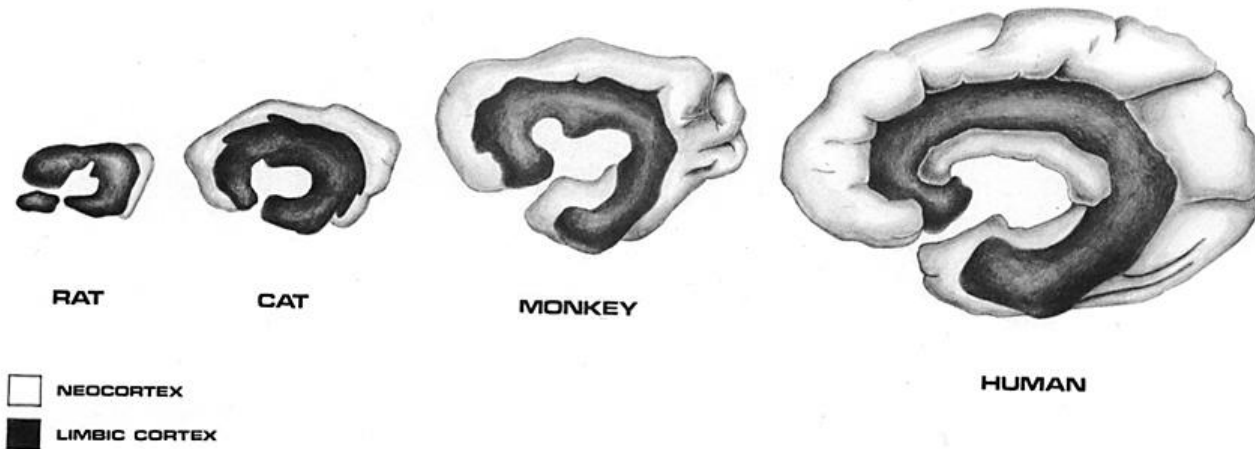Basic Science

Clinical Systems Level Neuroscience

+ cognitive, affective, social neuroscience

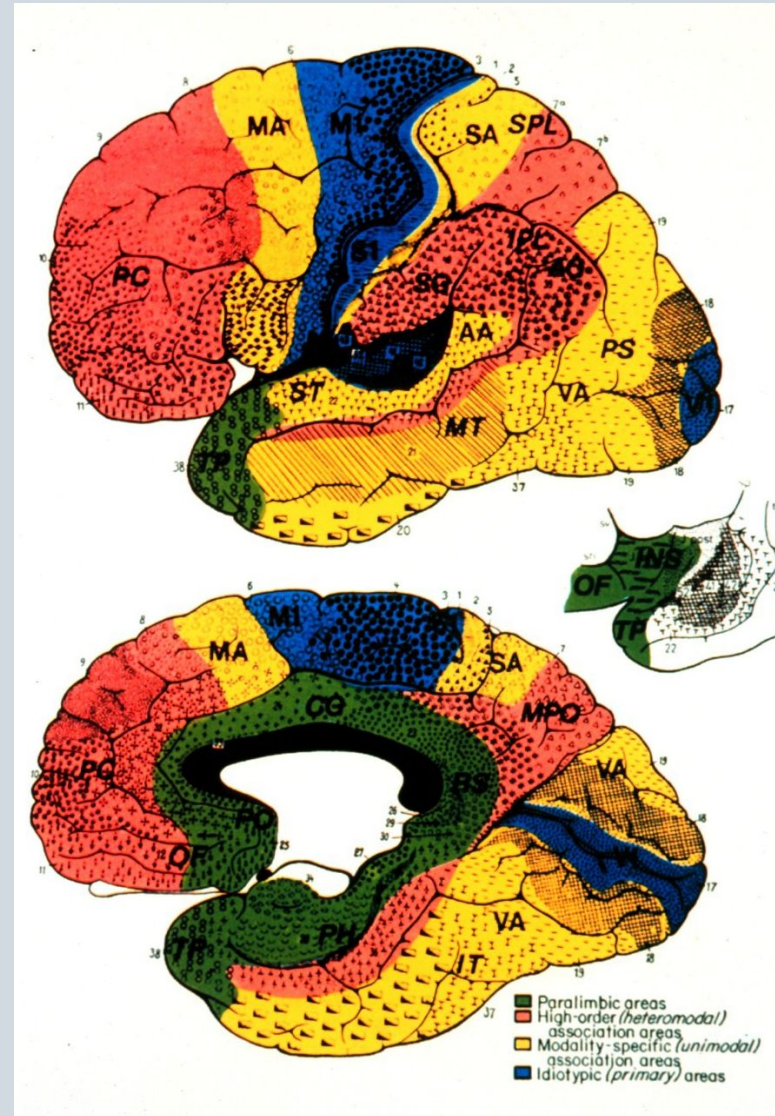**Integrated brain circuit model of perception, cognition, emotion & behavior**

Core processes and functions: mapping gradients onto brain circuits

Evolutionary Context:
Limbic and Prefrontal Cortex Across Species

Mesulam

Hierarchical, Modular, Integrated, Feedforward/Feedback Complex Spatio-Temporal Systems with Emergent Functions

Information
Mapping
Representation
Classification, Context
Pattern Detection
Salience Labeling
Memory, Change
Symbolic, Language
Prediction Modeling
Decision Making, Planning
Action, Interaction
Monitoring
Control

The Human Connectome

Anatomy
Klingler's method for fiber tract dissection uses freezing of brain matter to spread nerve fibers apart. Afterwards, tissue is carefully scratched away to reveal a relief-like surface in which the desired nerve tracts are naturally surrounded by their anatomical brain areas.
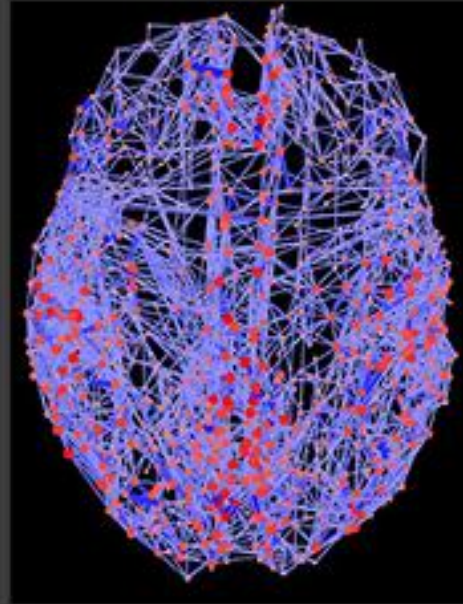
Connectome
Shown are the connections of brain regions together with "hubs" that connect signals among different brain areas and a central "core" or backbone of connections, which relays commands for our thoughts and behaviors.
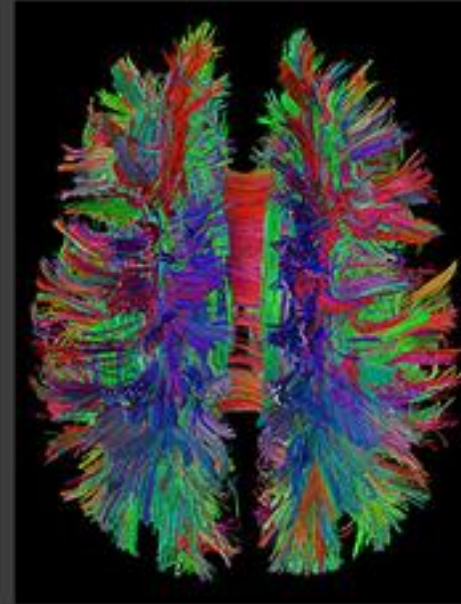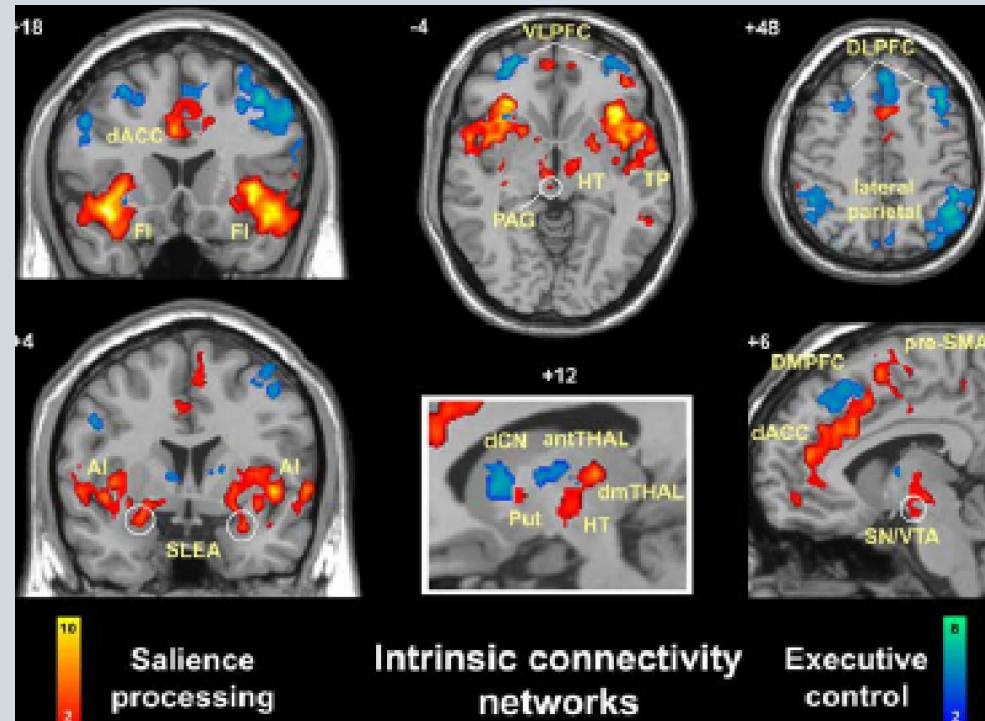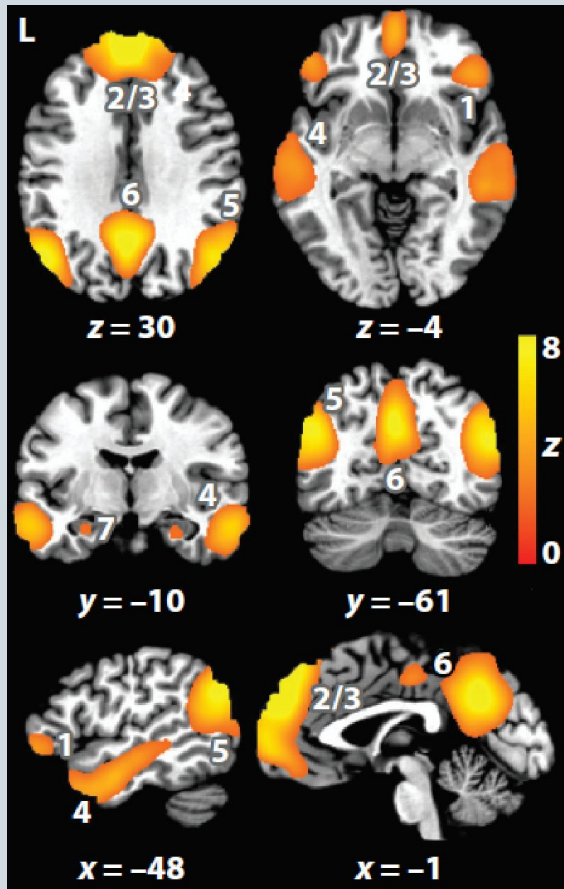
Neuronal Pathways
A new MRI technique called diffusion spectrum imaging (DSI) analyzes how water molecules move along nerve fibers. DSI can show a brain's major neuron pathways and will help neurologists relate structure to function.
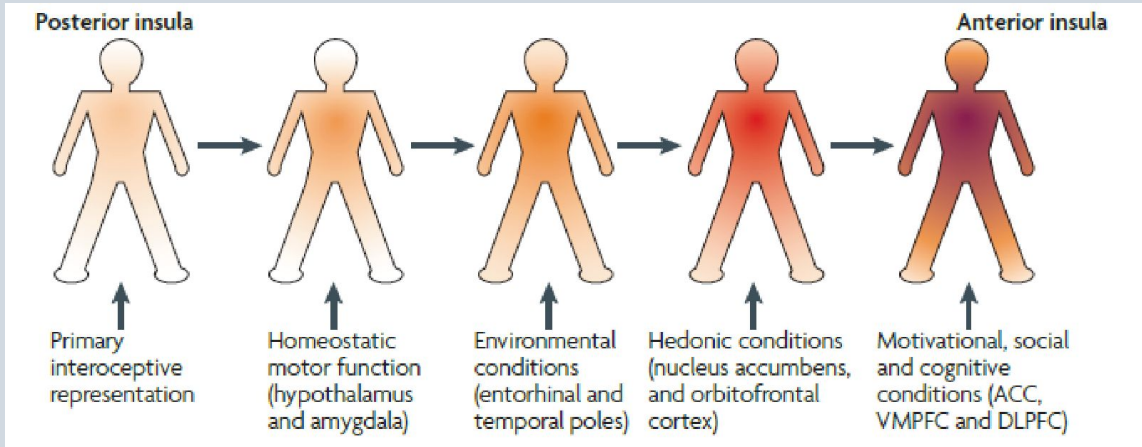
Mele

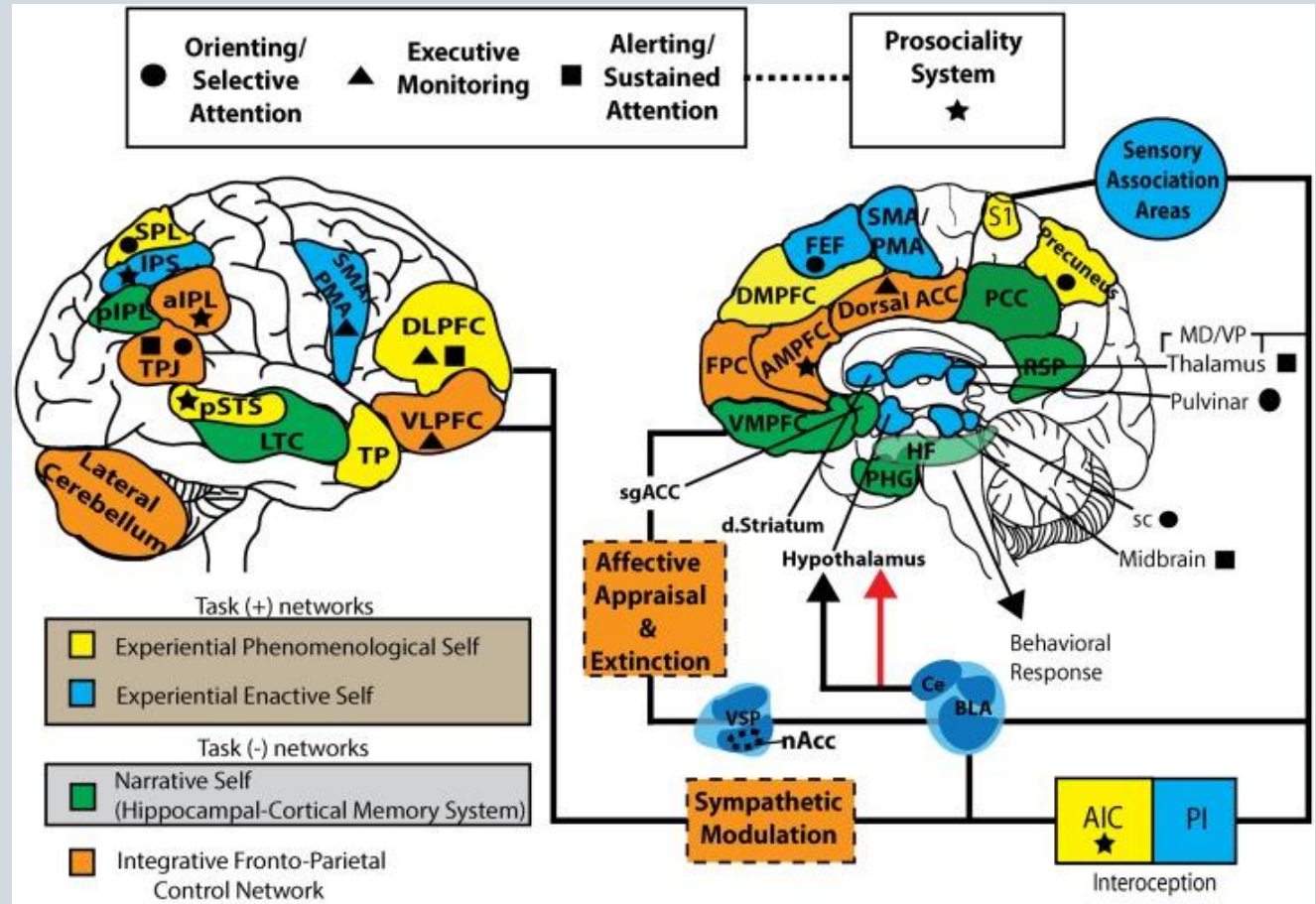# Resting State Networks:
Default mode, Executive, Salience
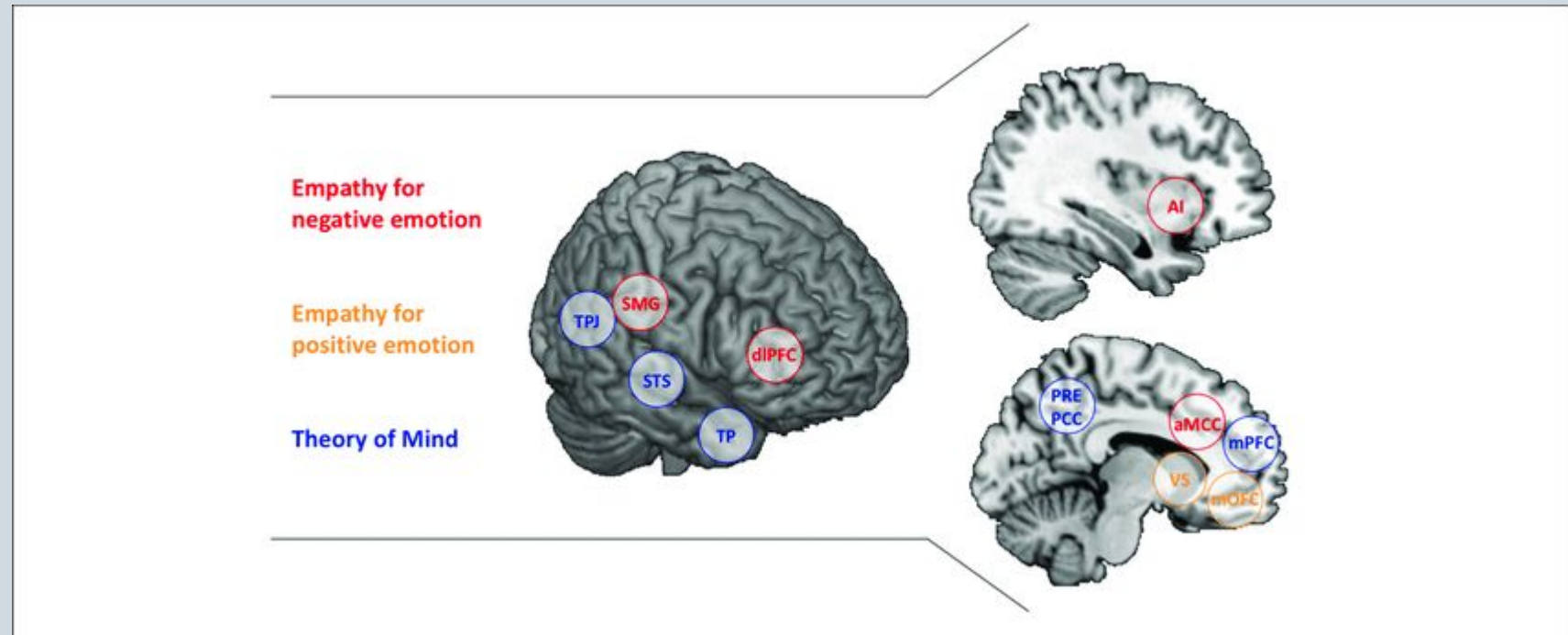

Raichle, Seely

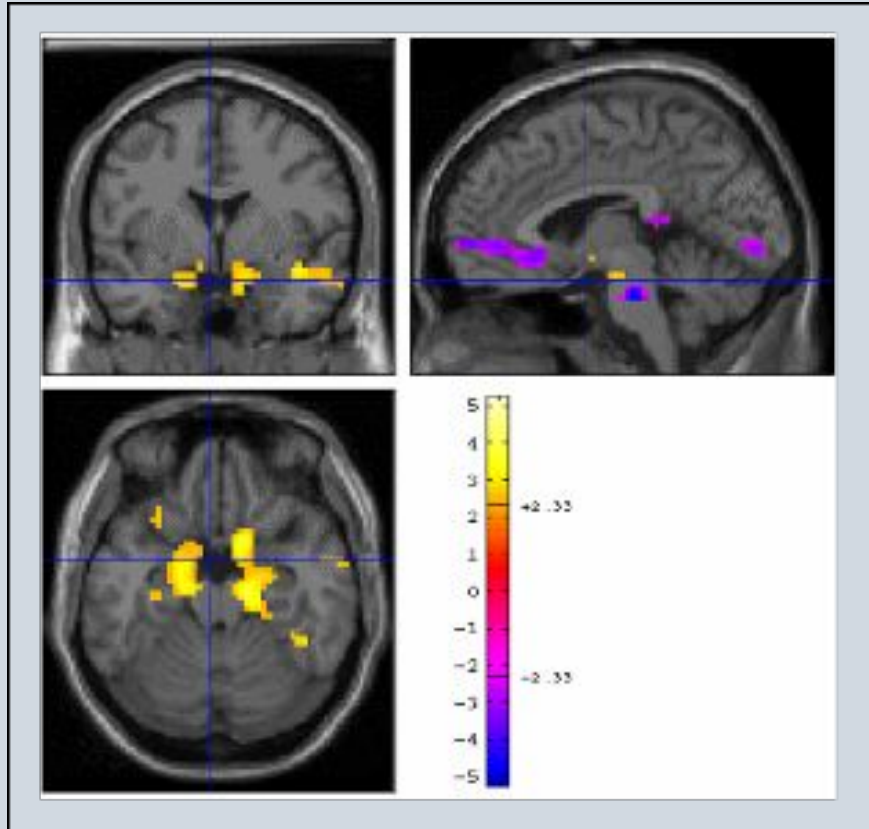Awareness of Self
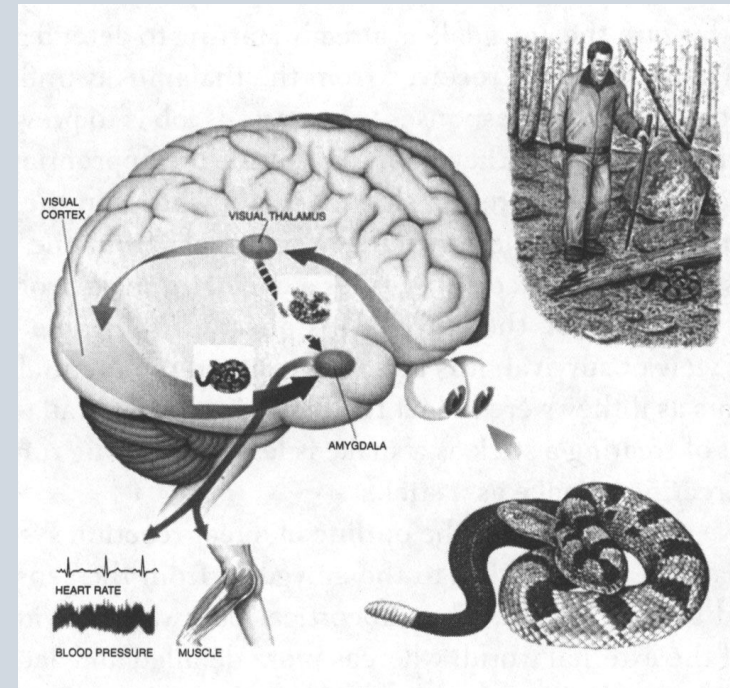
Craig

Vago and Silbersweig

Awareness of Others

Paranoid Delusions in Schizophrenia: Hyperactivation of Threat Circuitry
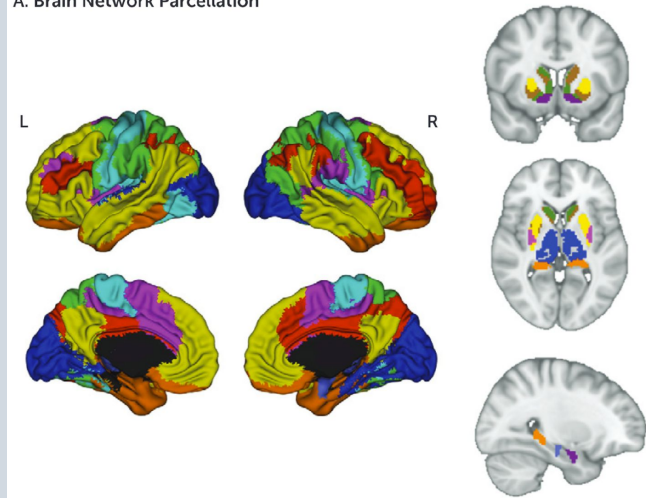

Silbersweig/Stern

LeDoux

Brain Connectivity Predictors of Antidepressant Response, Targeting of Therapeutic Brain Stimulation

FIGURE 1. Brain network parcellation and moderators of treatment response in a study of functional brain connectivity and antidepressant treatment response[a]

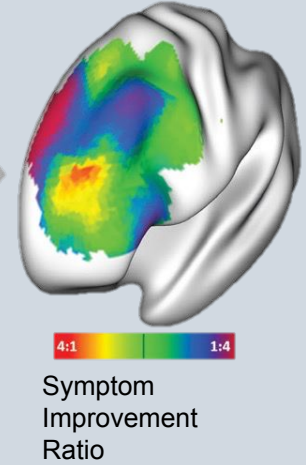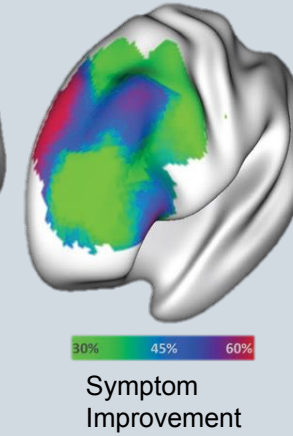A. Brain Network Parcellation

B. Moderators of Treatment Response
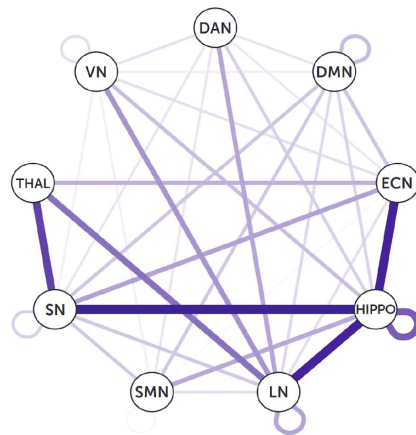
Dysphoric

Anxiosomatic

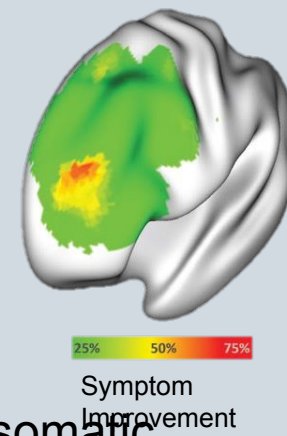Symptom Improvement

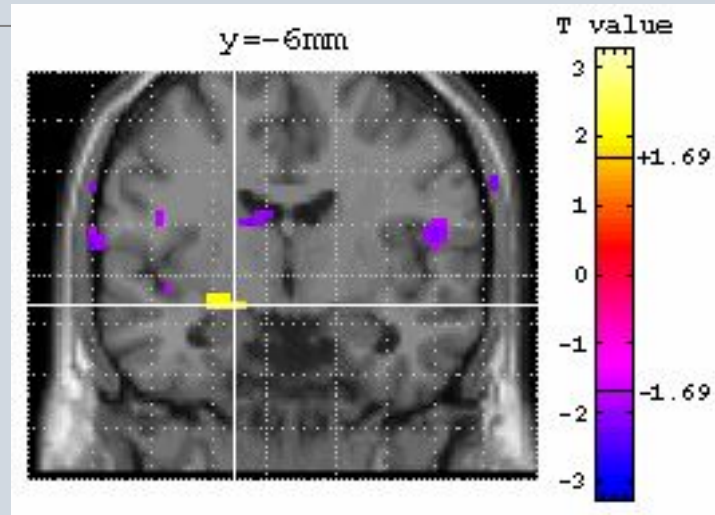Symptom Improvement

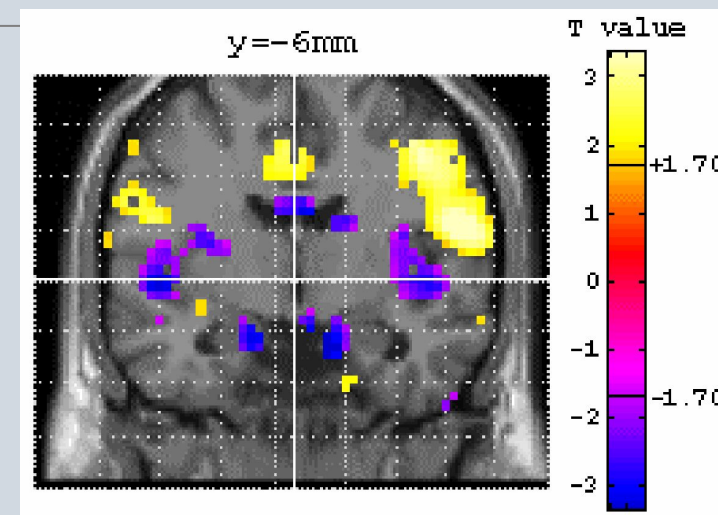Symptom Improvement Ratio

Fatt et al

Siddiqi et al

Neuromodulation by Psychotherapy: Schizophrenia and Paranoid Delusions Before and After Group CBT
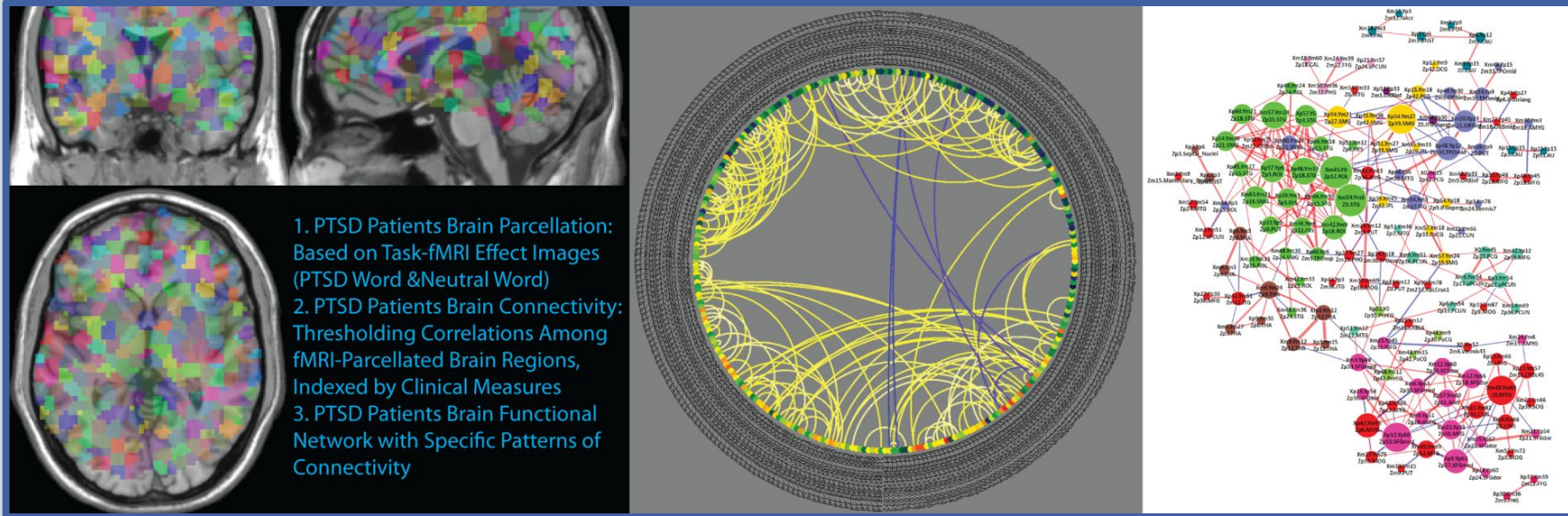
BEFORE
(P6 score =4)

 Extended amygdala

AFTER
(P6 score =2)

 Amygdala
 Dorsal AC
 DLPFC

PTSD *Connectivity/Networks*

1. PTSD Patients Brain Parcellation: Based on Task-fMRI Effect Images (PTSD Word &Neutral Word)
2. PTSD Patients Brain Connectivity: Thresholding Correlations Among fMRI-Parcellated Brain Regions, Indexed by Clinical Measures
3. PTSD Patients Brain Functional Network with Specific Patterns of Connectivity

|	Assortativity	Global efficiency	Characteristic path length	Network radius	Network diameter	Density	Mean Clustering Coefficient	Transitivity	Mean global routing efficiency	No. Vertices	No. edges
PT_EMPT	0.215	0.2111	5.334	12.3045	11.3193	0.0471	0.178	0.178	0.2065	184	793
PT_EMNU	0.3794	0.2547	4.514	6.0757	10.5011	0.0898	0.2467	0.2849	0.2547	183	1496
PT_EMPTvNU	0.3238	0.2158	5.2429	7.0876	11.2641	0.0511	0.1751	0.1722	0.2158	180	824
NL_EMPT	0.4367	0.2178	5.4471	7.2566	13.2059	0.0508	0.2029	0.2806	0.2178	183	846
NL_EMNU	0.3259	0.2142	5.3608	6.7804	12.8861	0.0413	0.1797	0.1776	0.2142	183	688
NL_EMPTvNU	0.2996	0.2545	4.6517	1.0563	13.119	0.074	0.2768	0.3124	0.2379	181	1206

Multivariate Analytics ML/AI
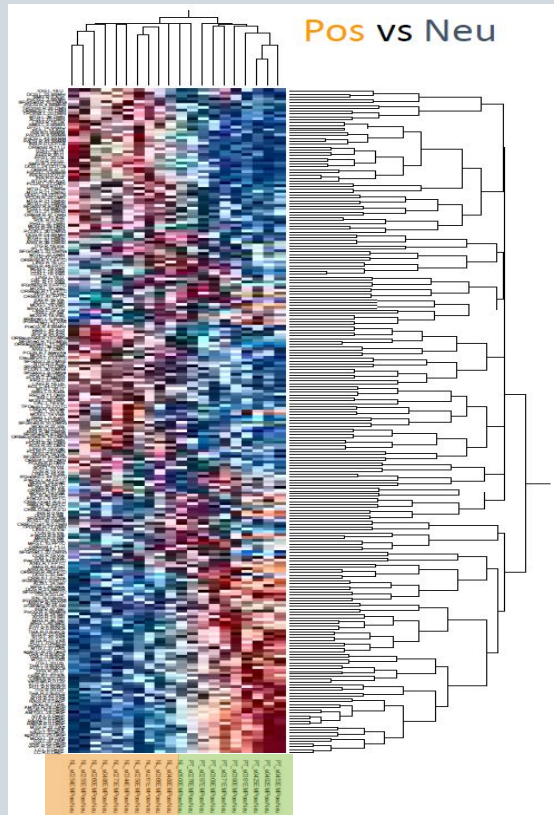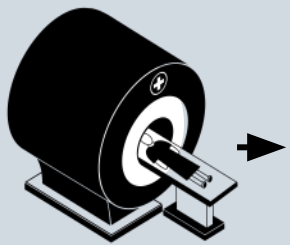
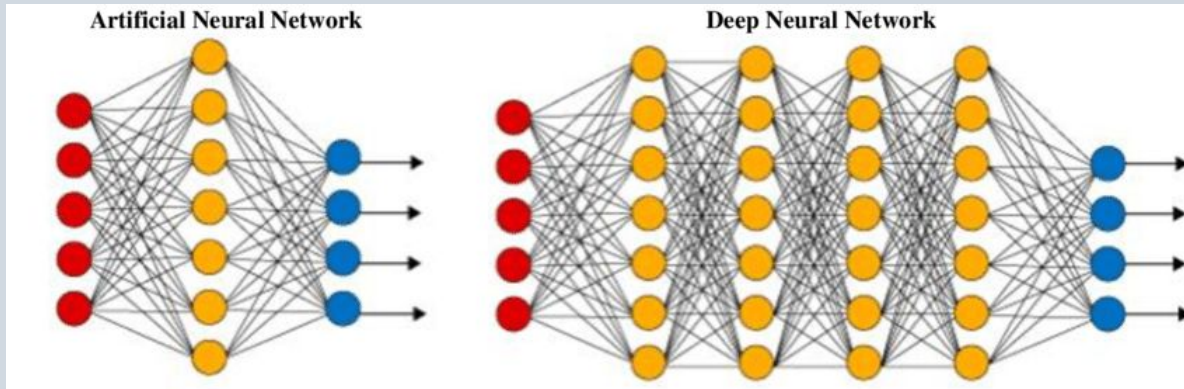Patient/HC and circuit stratification based on brain network activity
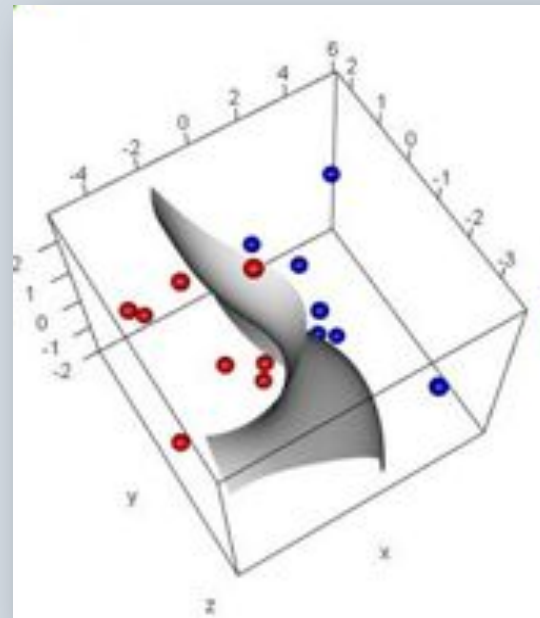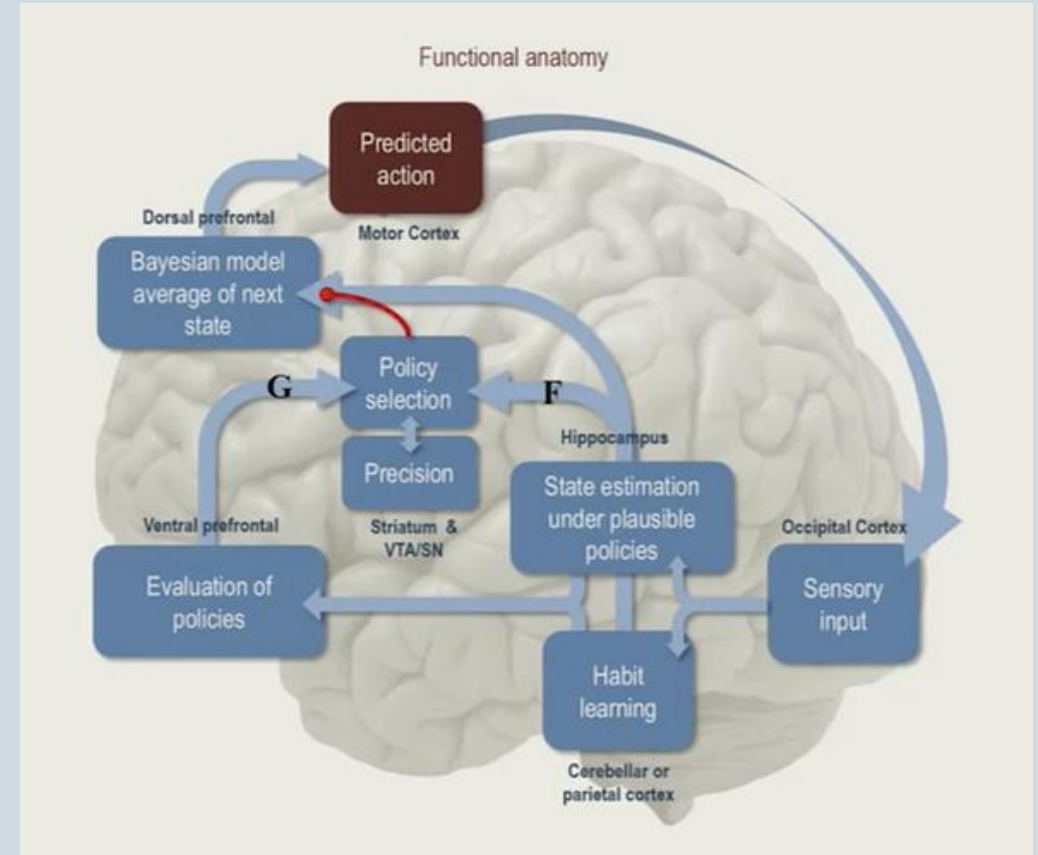
AI-based classification
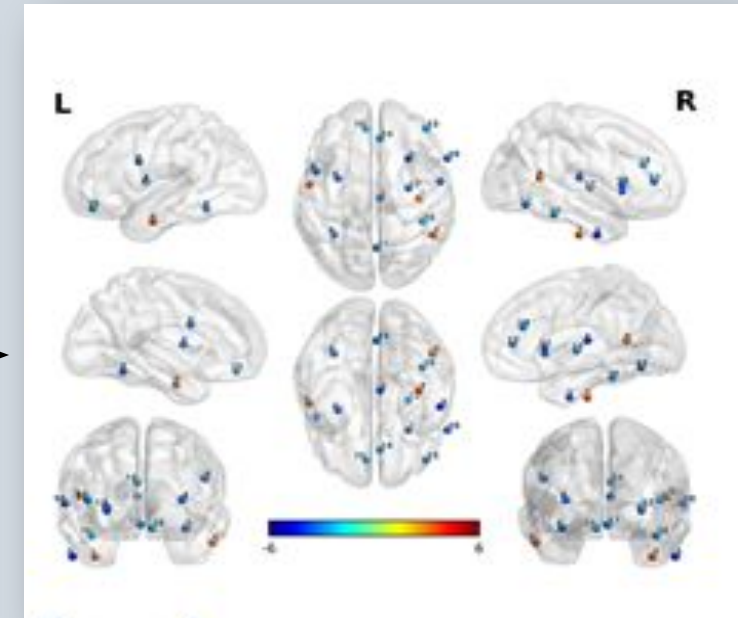
Define Disease Subtypes: brain nodes that drive classification

Common Models:
Neural Networks, Free Energy Principle
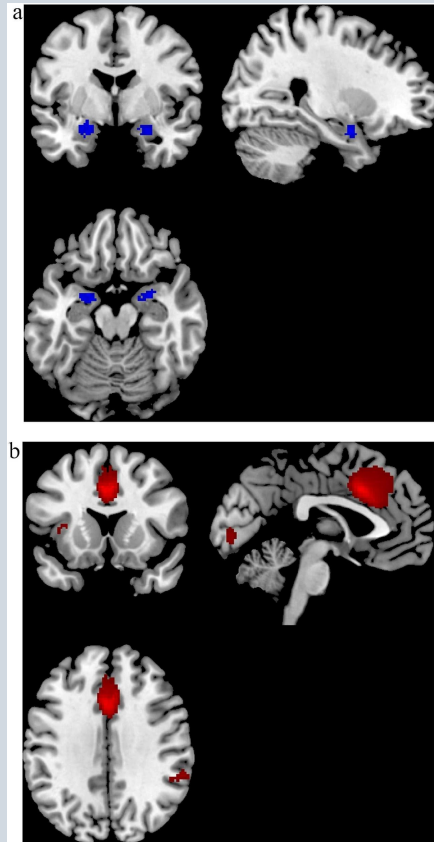

Bharath

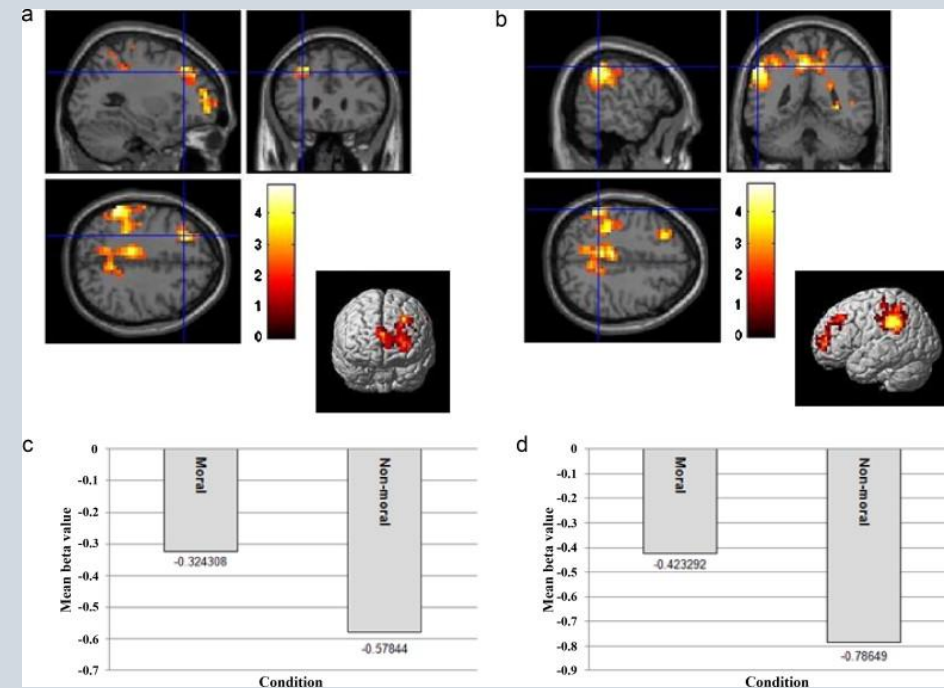Friston

Neurology of Political Ideology

Ideology

Brain and Behavior Affiliations	High Conservatism	High Liberalism
Personality	Stability; opposition to change	Novelty
	Conformity	Unconventional; self-expression
	Tradition	New experiences and sensations
	Order, structure, and closure	Flexibility and variability
	Favor less complexity; harder categorization	Tolerance for uncertainty and ambiguity
	Purity	Minimization of harm
	Authority	Equality
	Conscientiousness	Empathy
	Distinctions with out-groups	Universal community
	Expressions of power	Expressions of warmth
Cognitive	Negativity bias	No clear bias
	Greater sensitivity to threat or loss	Greater sensitivity to cues for altering habitual response patterns
	Sensitivity to disgust	
Physiological	Greater activation of right amygdala	Greater conflict-related anterior cingulate cortex activity
Neuroimaging	Increased gray matter volume in right amygdala and other right anterior structures	Increased gray matter volume in anterior cingulate cortex

[a] References and studies for cognitive, physiological, and neuroimaging discussed in the article text.

Mendez et al

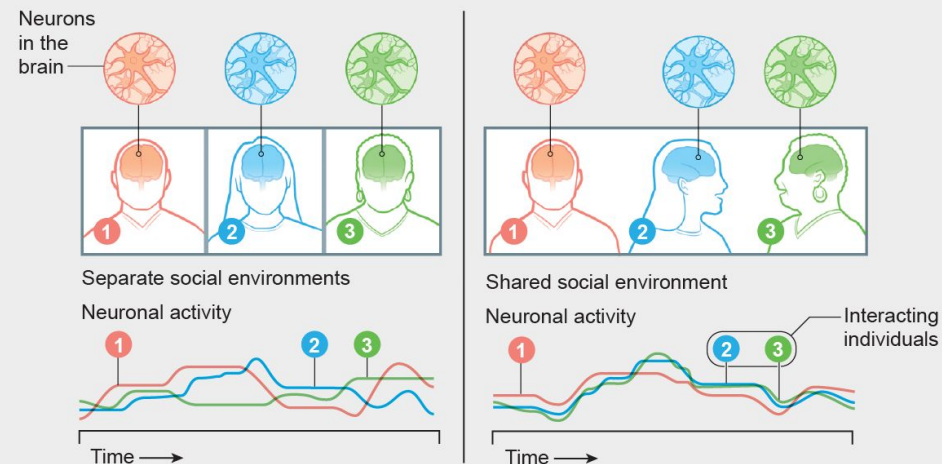# Processing of In-Group vs Out-Group Bias and Prejudice and Moral Decision Making

Saarinen et al

Renier et al

Interacting Brain-Minds

Wheatley

Advancing Human-Centered AI through Integration with Natural Systems and Neuroscience: Toward Policy and Societal Well-Being

- Combining insights from neuropsychiatry and behavioral neuroscience with advances in generative AI

- Common models and approaches such as Free Energy Principle, Complex Systems, Neural Networks

- Enhancing pro-social and diminishing anti-social thinking, feeling, behavior

- Decreasing polarization, hatred, violence

- Improving collaborative/collective decision making and problem solving