



# **The cybernetic Bayesian brain**

**Anil Seth**

Sackler Centre for Consciousness Science,  
University of Sussex, Brighton, UK

[www.anilseth.com](http://www.anilseth.com)

[www.neurobanter.com](http://www.neurobanter.com)

[www.sussex.ac.uk/sackler](http://www.sussex.ac.uk/sackler)

@anilkseth

Zurich, December 2015

## Why am I here?

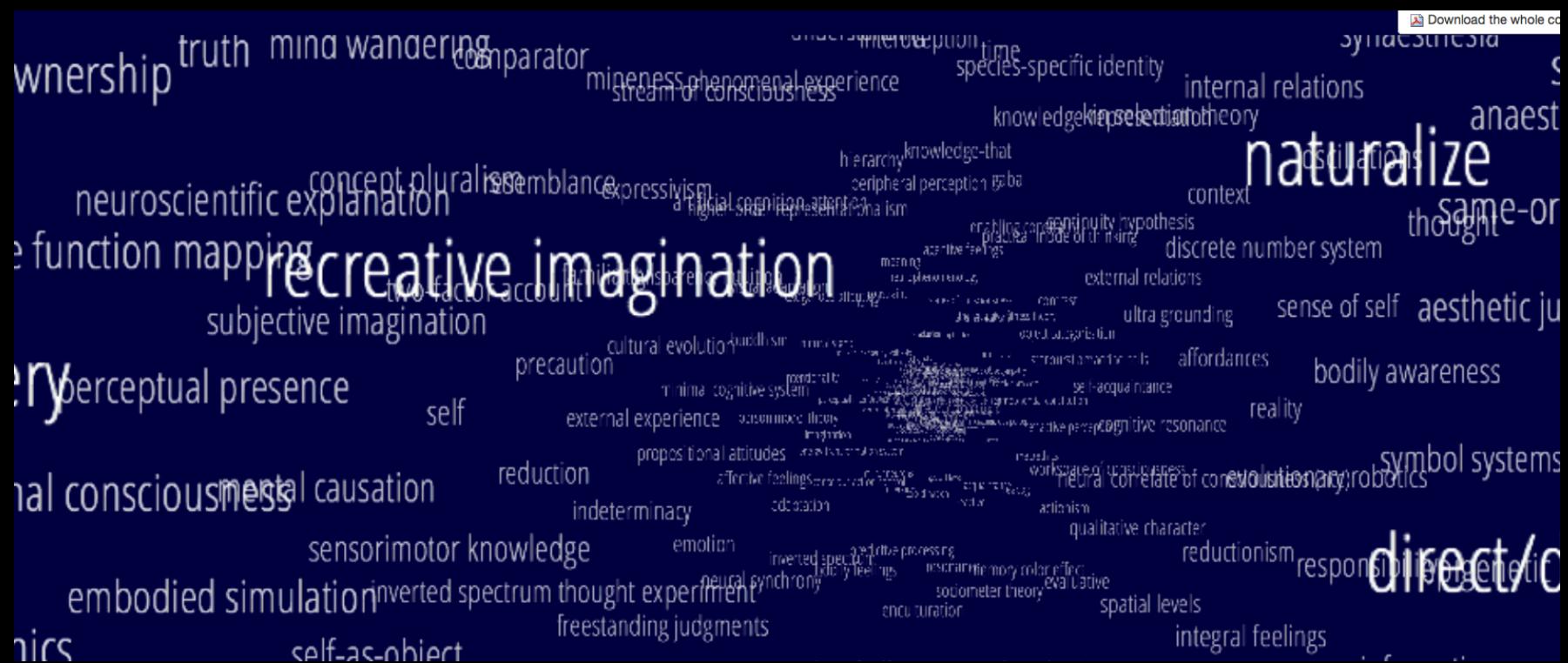
- Research on consciousness: level, content, and self
- Highlights phenomenology, especially the experience of being an embodied self.
- Need to understand the variety of experience, in health and in disease, in order to target models of psychiatric conditions.

## Outline

- The self from within and without
- Predictive processing
- Interoceptive inference, emotion, and embodiment
- Cybernetics and ‘inference to the best prediction’
- Interoception in autism and in psychosis
- Run out of time
- Counterfactual predictions and perceptual presence
- Summary: The cybernetic Bayesian brain



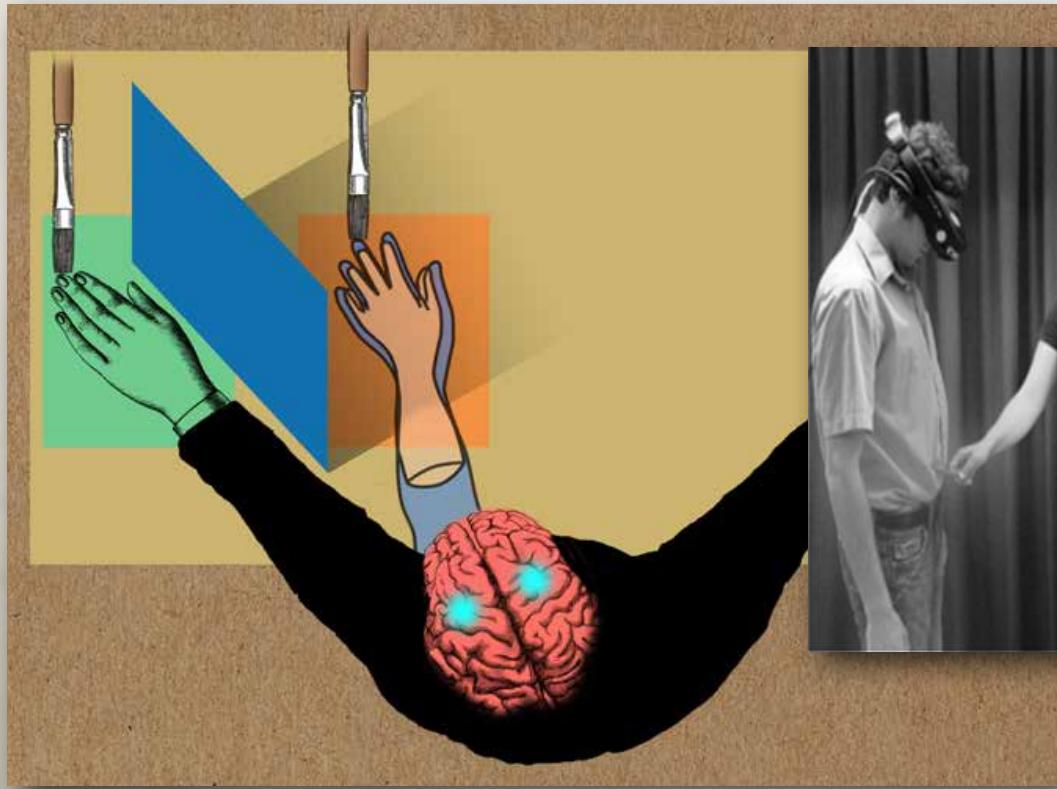
<http://open-mind.net/>



**the self, from within and from without**

# The self from within and from without

- A key property of conscious selfhood: owning and identifying with a particular body:
- One tradition - perception of the body **from the outside**.



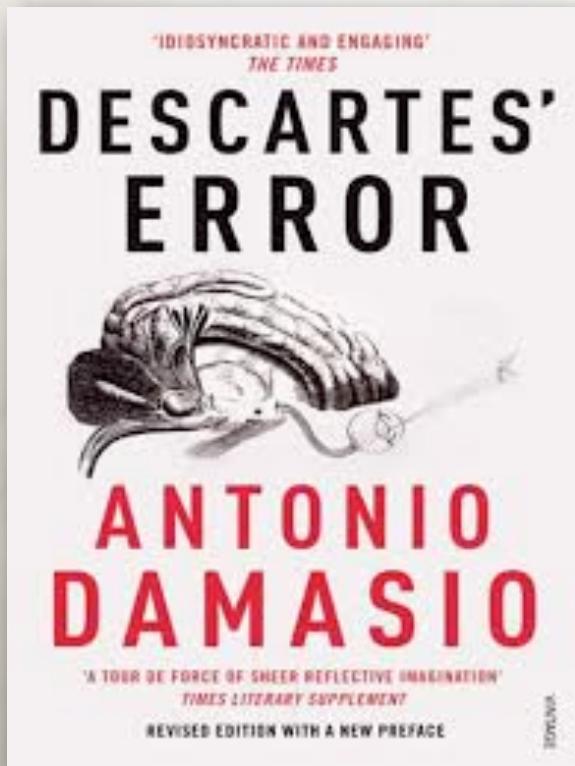
Rubber hand illusion



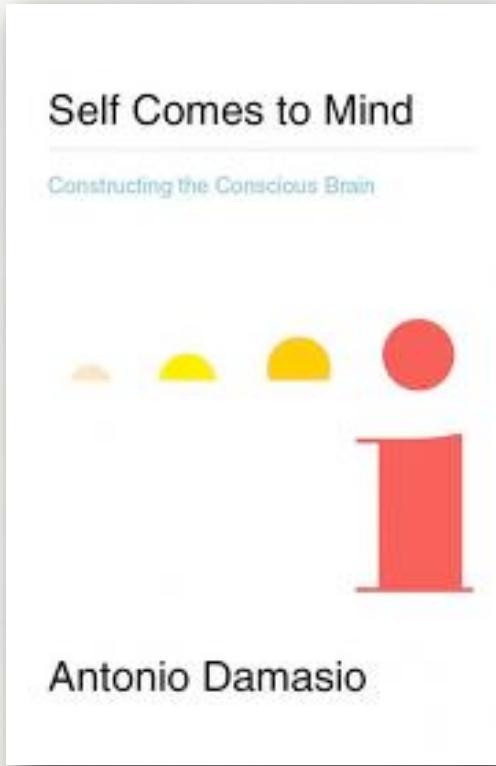
Body swap illusion

# The self from within and from without

- A key property of conscious selfhood: owning and identifying with a particular body:
- One tradition - perception of the body **from the outside**.
- A second tradition - perception of the body **from within**.



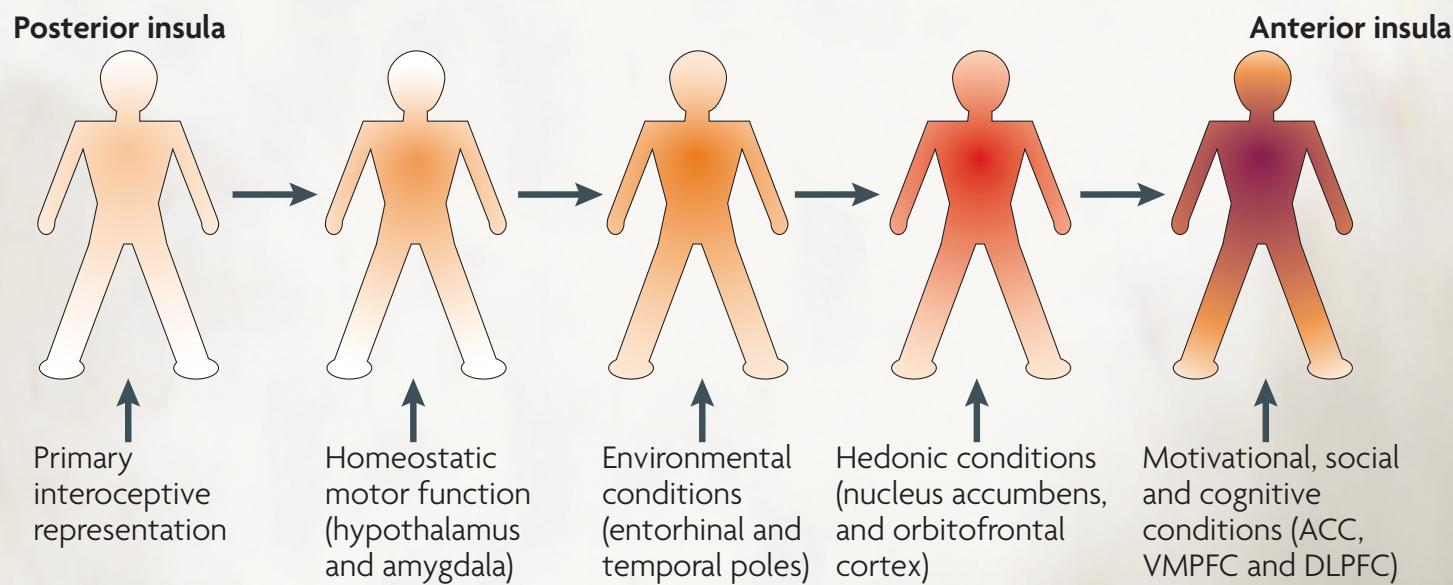
Damasio's "somatic marker" hypothesis



Damasio (1994/2010)  
Tsakiris et al (2011) *Proc Biol Soc*

# The self from within and from without

- A key property of conscious selfhood: owning and identifying with a particular body:
- One tradition - perception of the body **from the outside**.
- A second tradition - perception of the body **from within**.



Bud Craig's “material me”

Craig (2009) *Nat Neuro Rev*  
Tsakiris et al (2011) *Proc Biol Soc*



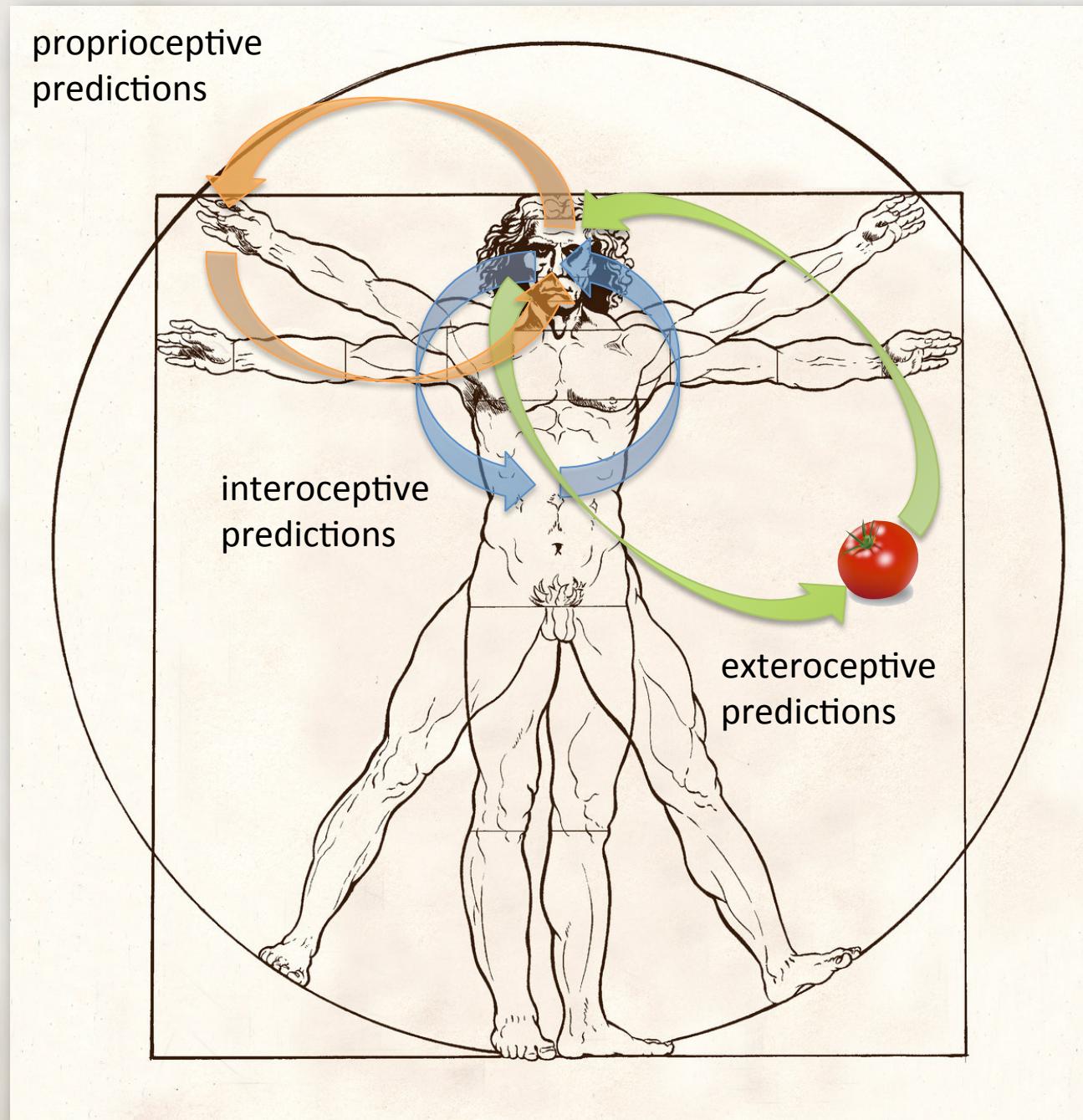
“Our mental life is knit up in  
our corporeal frames”

William James

# Different traditions, common mechanisms?

- Perception of the body **from the outside**:
  - multisensory integration.
- Perception of the body **from the inside**:
  - representations of internal physiological state.
- The basic idea:
  - Selfhood depends on predictive models of the causes of self-related signals, across exteroceptive, proprioceptive, and interoceptive domains.

# Different traditions, common mechanisms?



**predictive processing and perception**

# Perception as inference

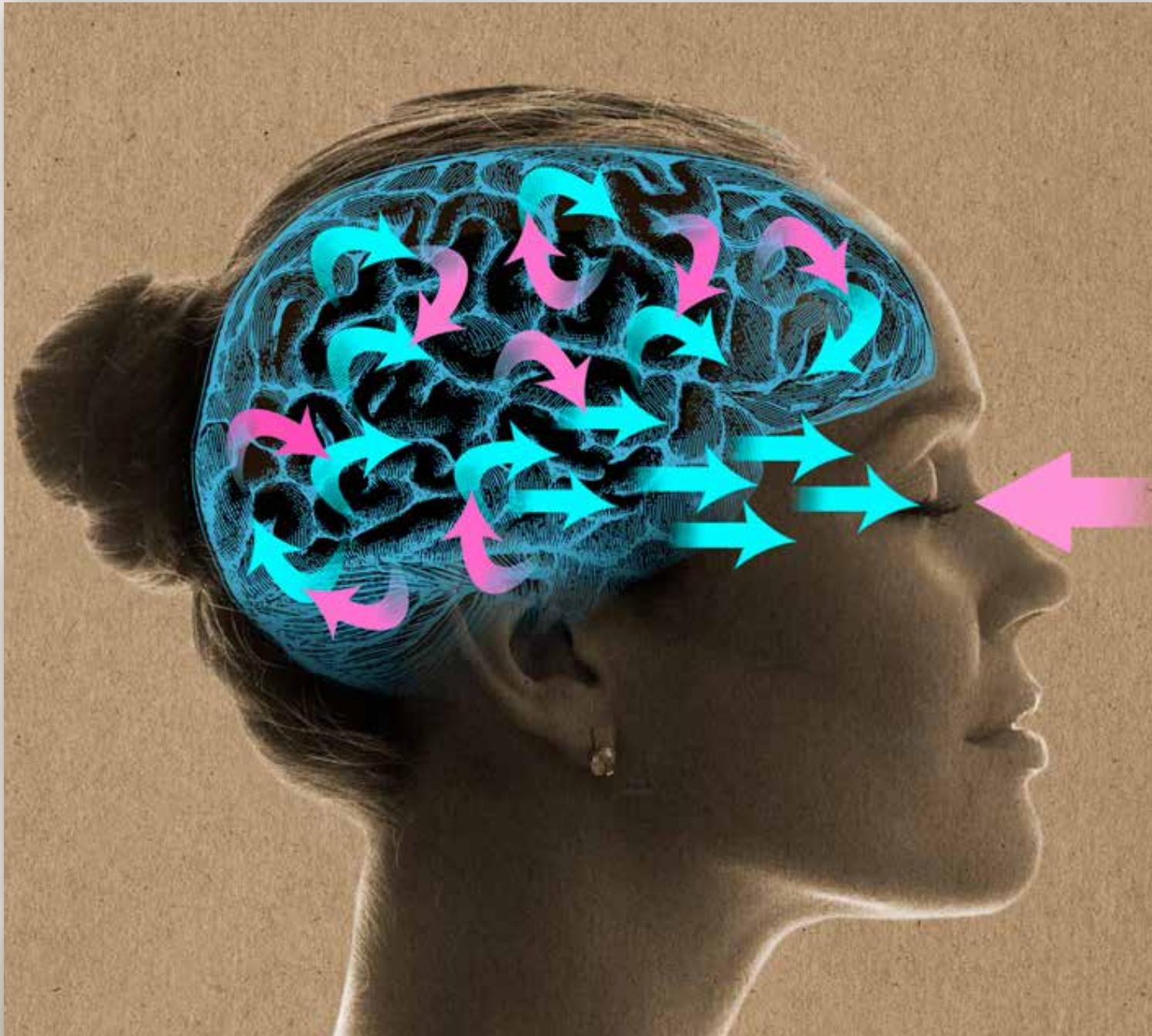


h/t Acer Chang

# Perception as inference

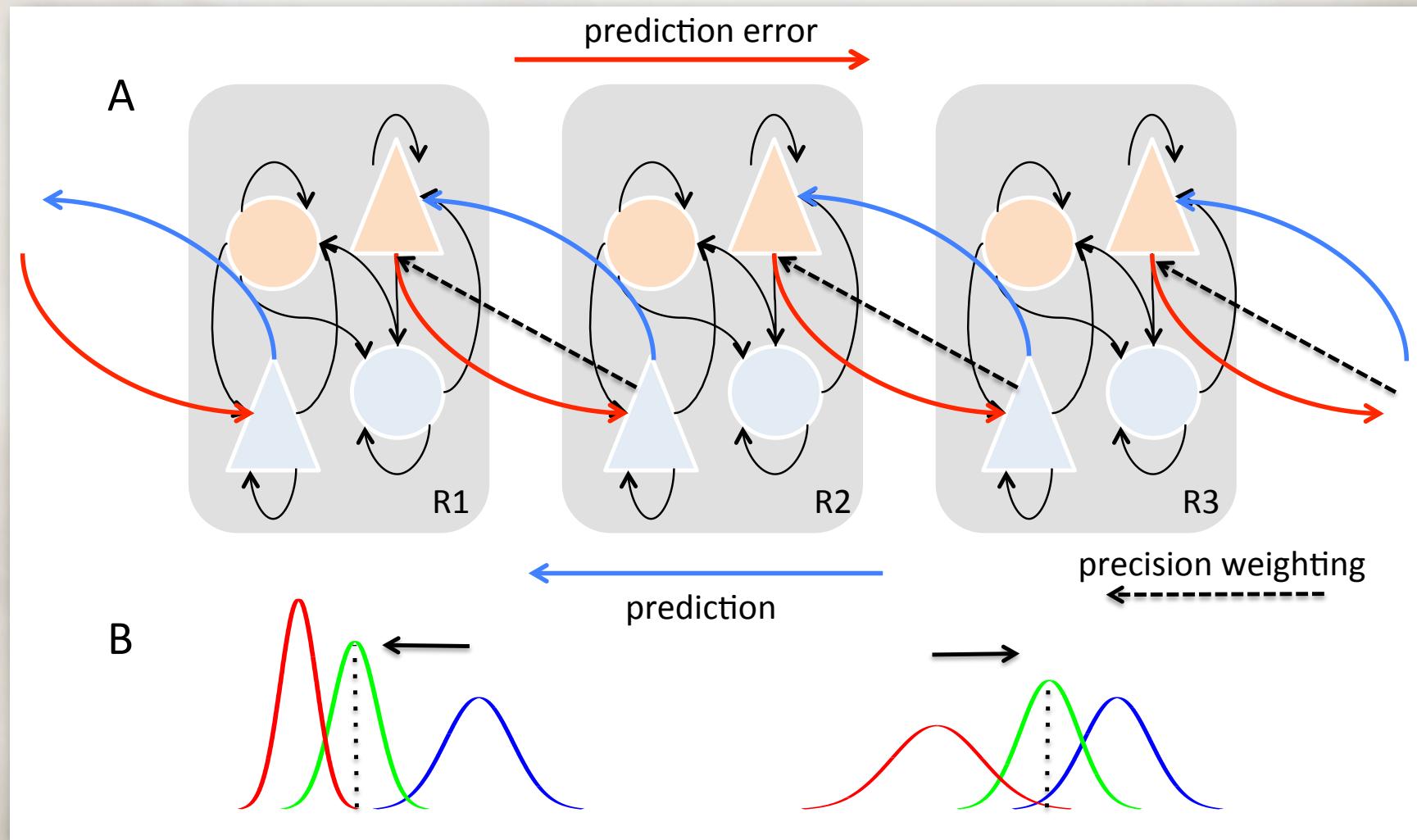


h/t Acer Chang



*30 Second Brain*, Seth (Ed), Ivy Press, 2014

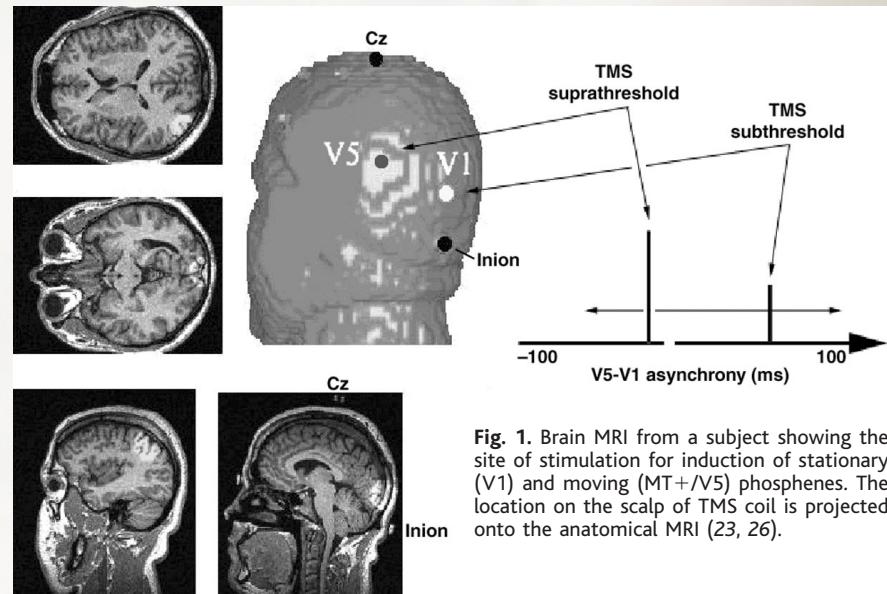
# Predictive processing in cortical networks



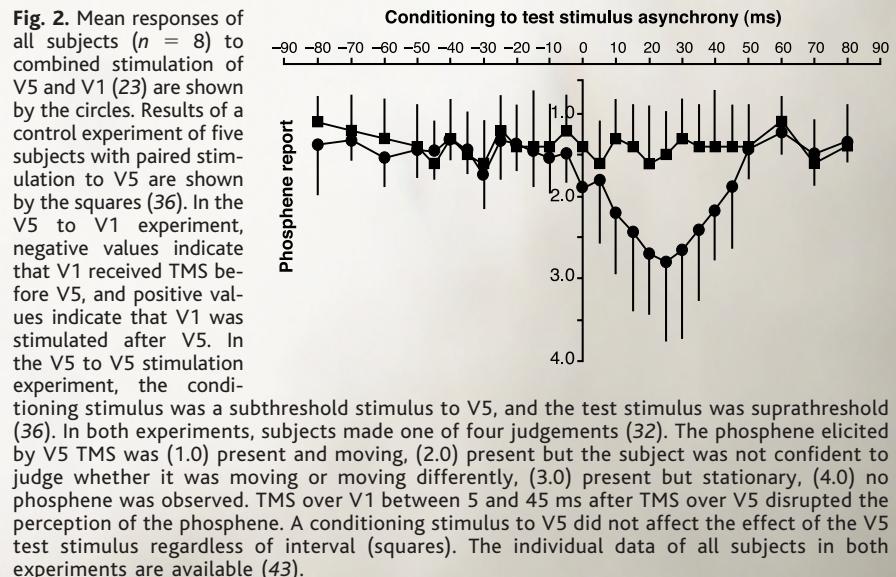
# Fast Backprojections from the Motion to the Primary Visual Area Necessary for Visual Awareness

Alvaro Pascual-Leone<sup>1\*</sup> and Vincent Walsh<sup>2</sup>

Much is known about the pathways from photoreceptors to higher visual areas in the brain. However, how we become aware of what we see or of having seen at all is a problem that has eluded neuroscience. Recordings from macaque V1 during deactivation of MT+/V5 and psychophysical studies of perceptual integration suggest that feedback from secondary visual areas to V1 is necessary for visual awareness. We used transcranial magnetic stimulation to probe the timing and function of feedback from human area MT+/V5 to V1 and found its action to be early and critical for awareness of visual motion.

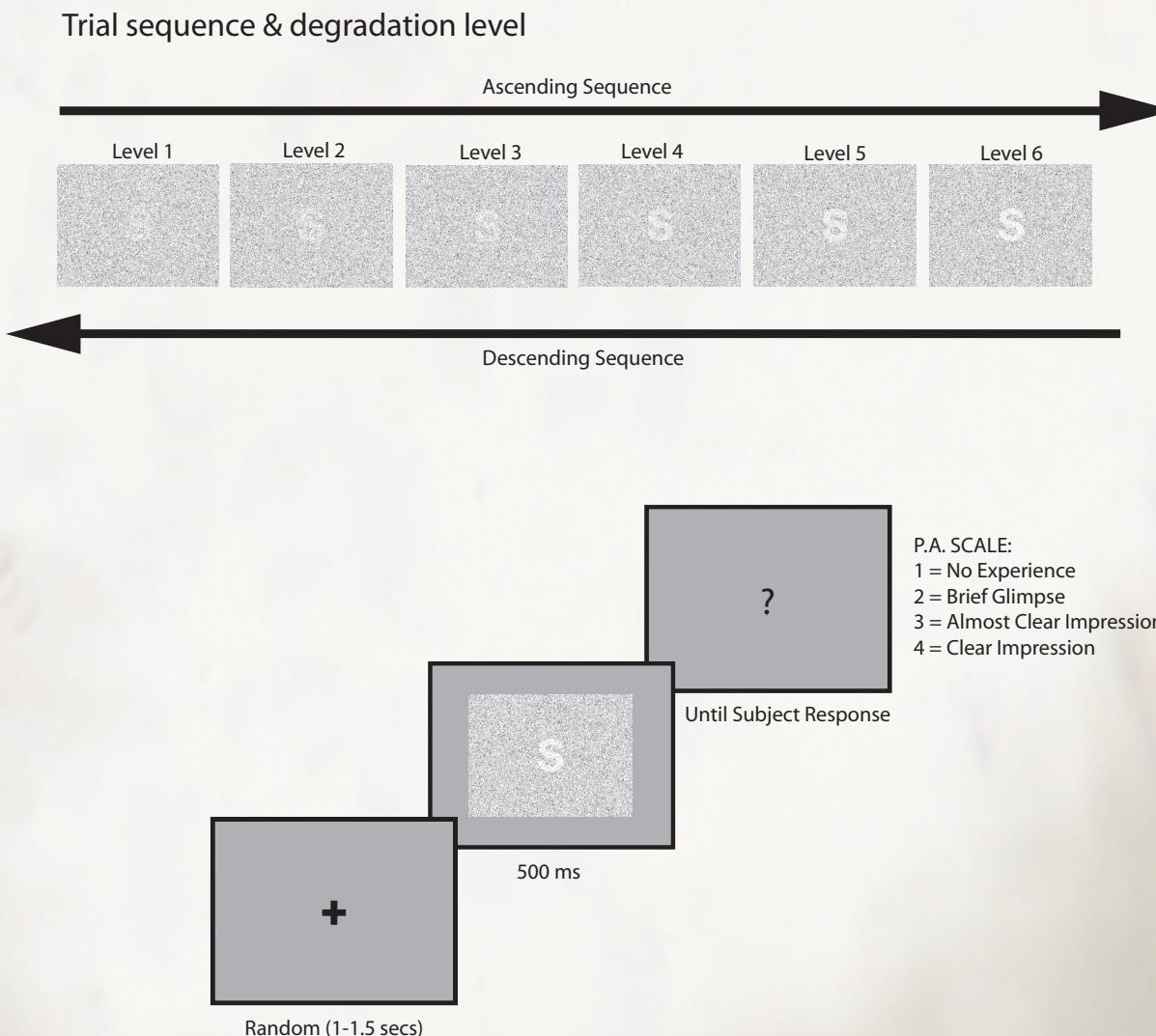


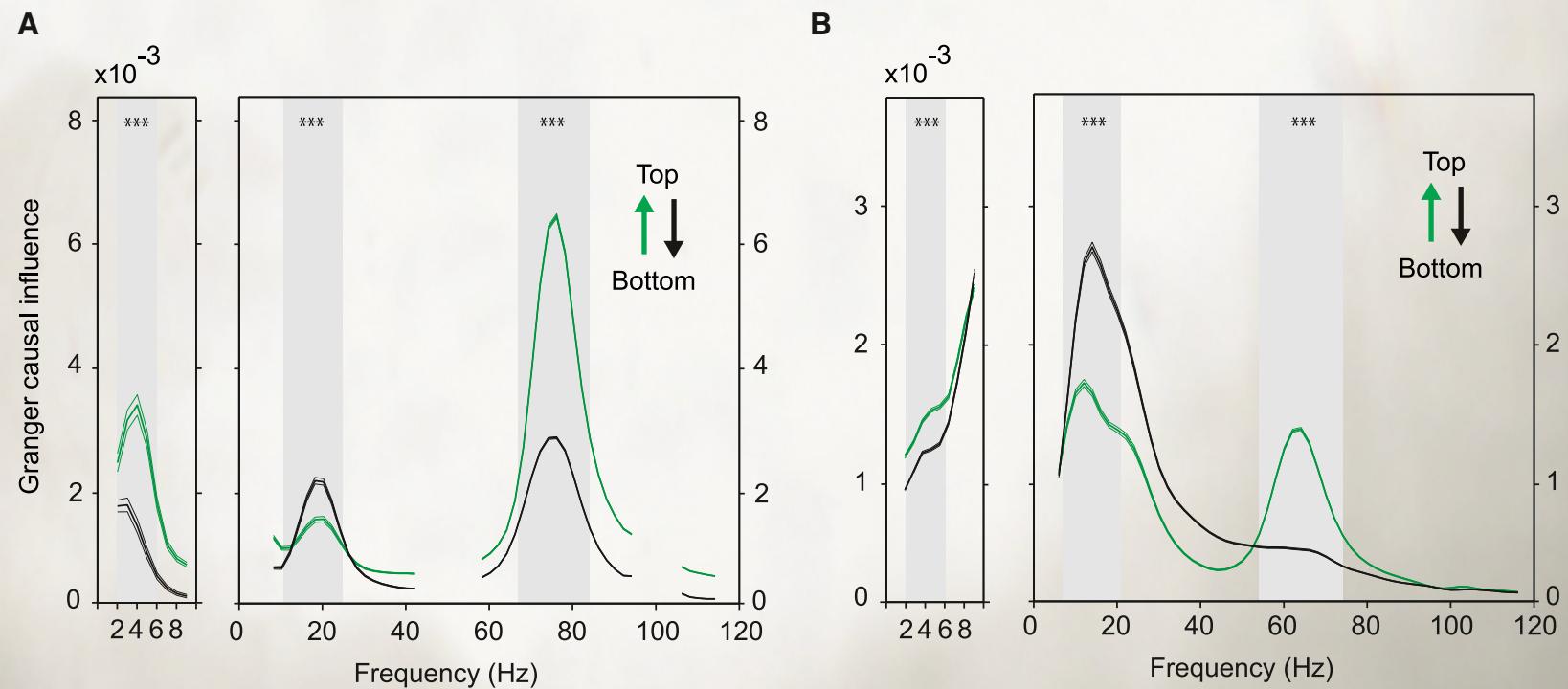
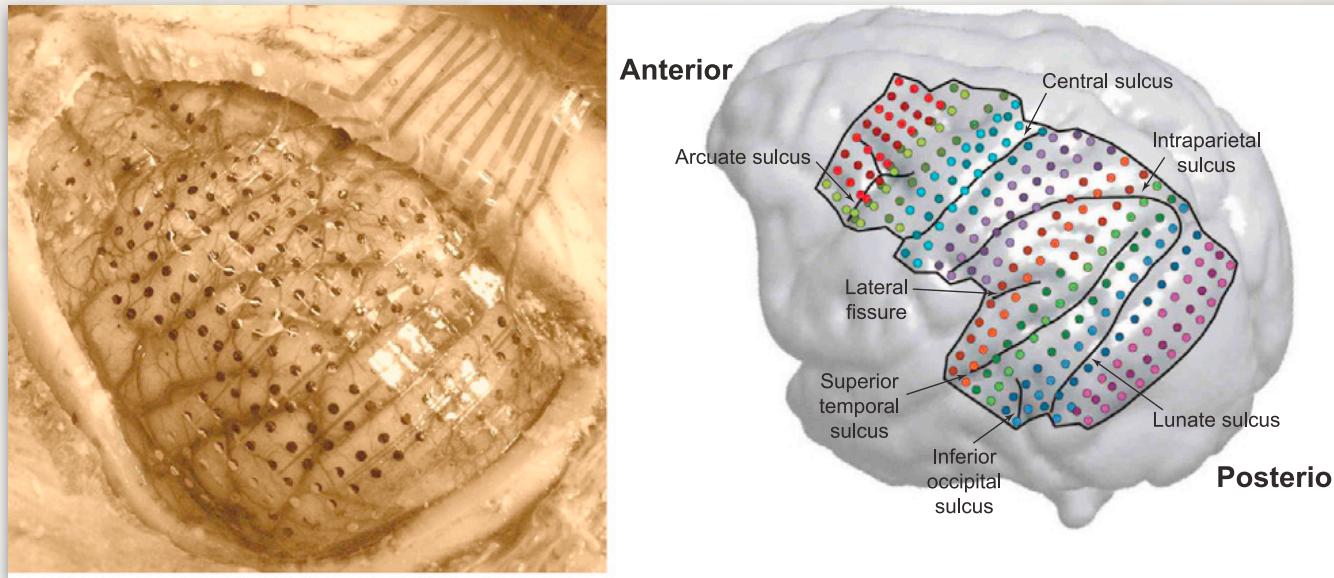
**Fig. 1.** Brain MRI from a subject showing the site of stimulation for induction of stationary (V1) and moving (MT+/V5) phosphene. The location on the scalp of TMS coil is projected onto anatomical MRI (23, 26).



**Fig. 2.** Mean responses of all subjects ( $n = 8$ ) to combined stimulation of V5 and V1 (23) are shown by the circles. Results of a control experiment of five subjects with paired stimulation to V5 are shown by the squares (36). In the V5 to V1 experiment, negative values indicate that V1 received TMS before V5, and positive values indicate that V1 was stimulated after V5. In the V5 to V5 stimulation experiment, the conditioning stimulus was a subthreshold stimulus to V5, and the test stimulus was suprathreshold (36). In both experiments, subjects made one of four judgements (32). The phosphene elicited by V5 TMS was (1.0) present and moving, (2.0) present but the subject was not confident to judge whether it was moving or moving differently, (3.0) present but stationary, (4.0) no phosphene was observed. TMS over V1 between 5 and 45 ms after TMS over V5 disrupted the perception of the phosphene. A conditioning stimulus to V5 did not affect the effect of the V5 test stimulus regardless of interval (squares). The individual data of all subjects in both experiments are available (43).

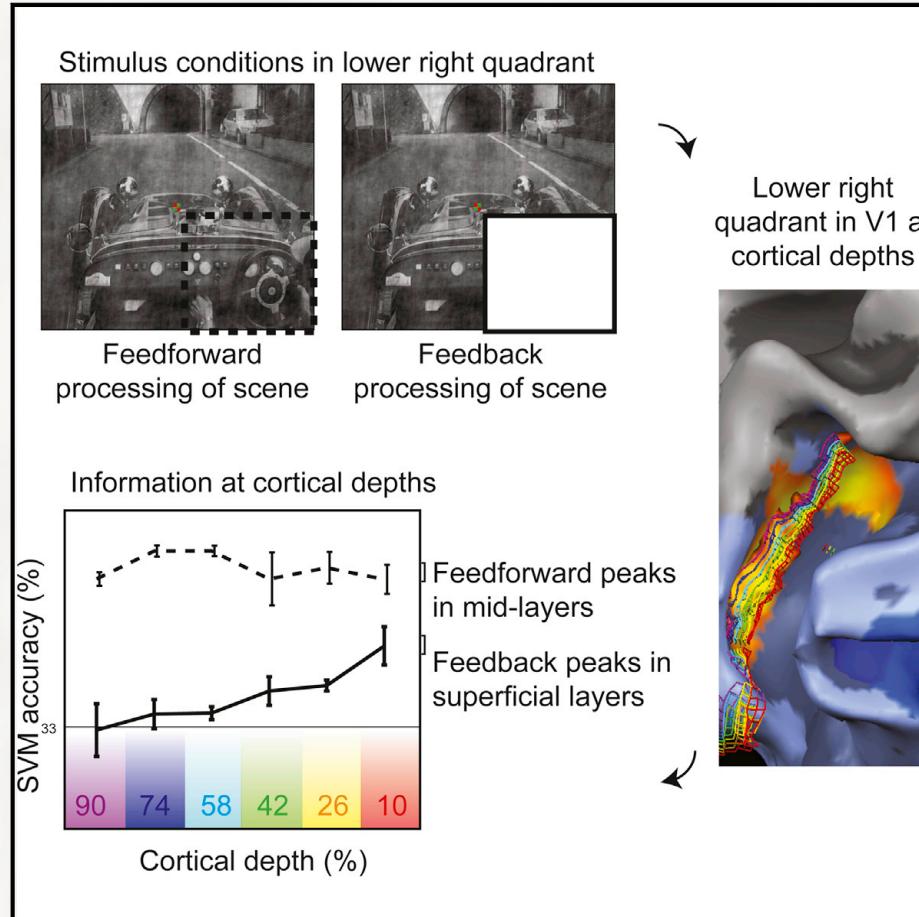
# Expectations can speed conscious access





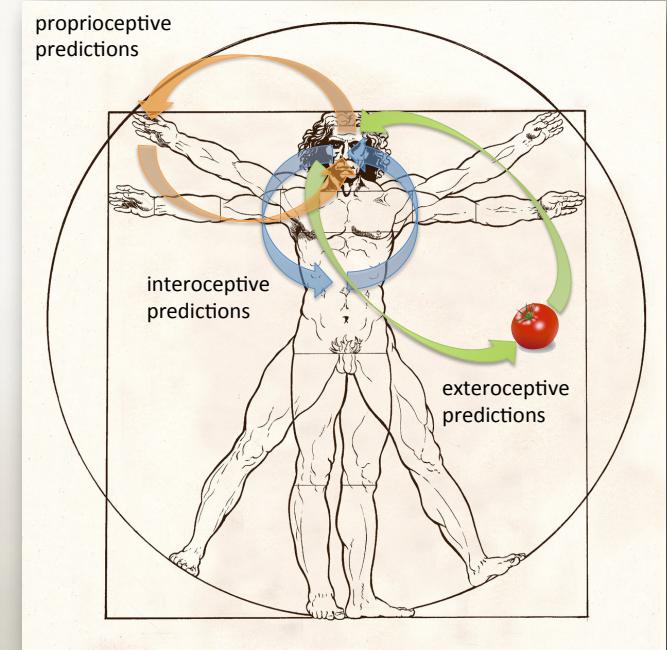
# Contextual Feedback to Superficial Layers of V1

Lars Muckli,<sup>1,\*</sup> Federico De Martino,<sup>2,4</sup> Luca Vizioli,<sup>1</sup> Lucy S. Petro,<sup>1</sup> Fraser W. Smith,<sup>3</sup> Kamil Ugurbil,<sup>4</sup> Rainer Goebel,<sup>2,5</sup> and Essa Yacoub<sup>4</sup>



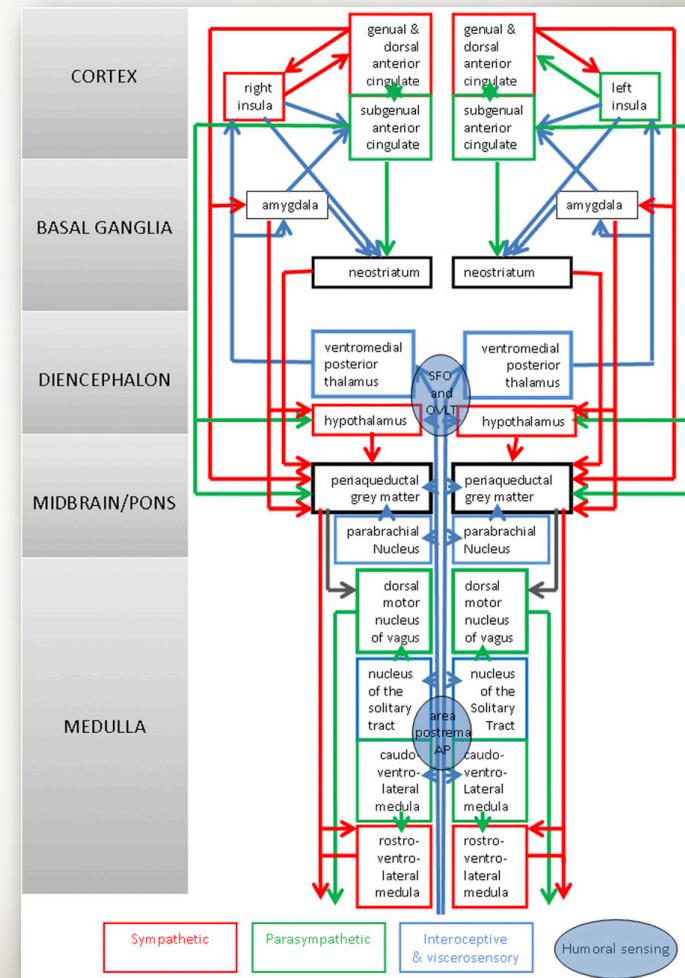
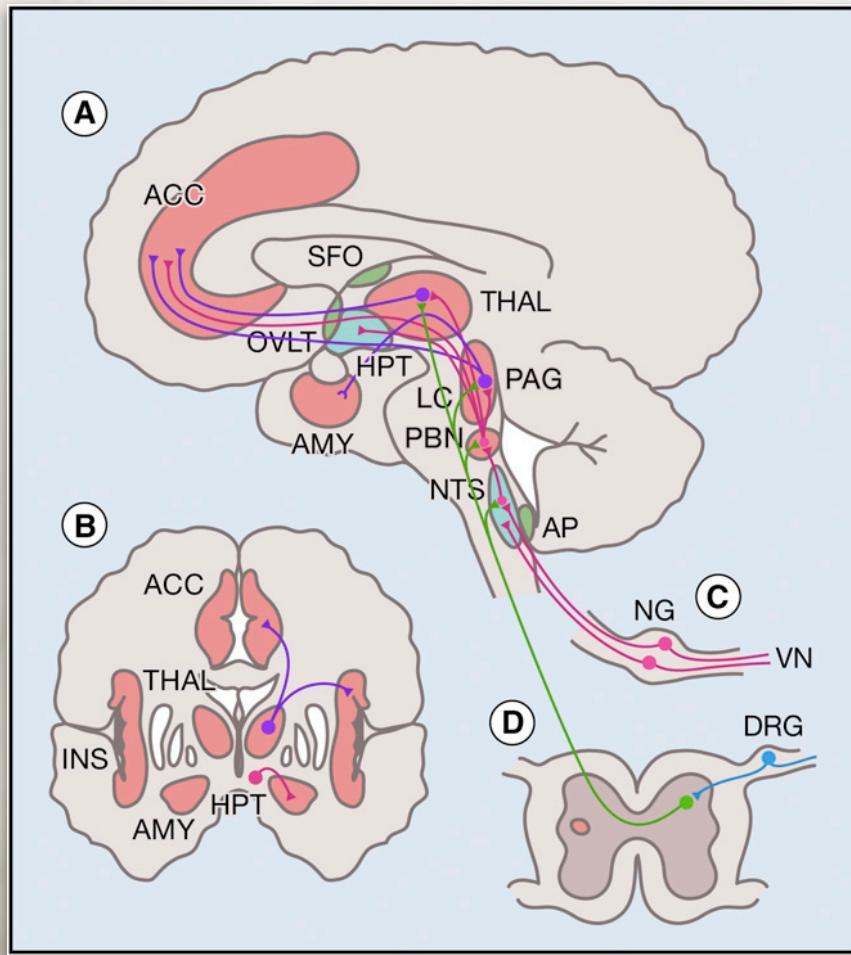
... also work by Maxine Sherman, Acer Chang, & Yair Pinto at Sussex

# interoceptive inference



# Interoception

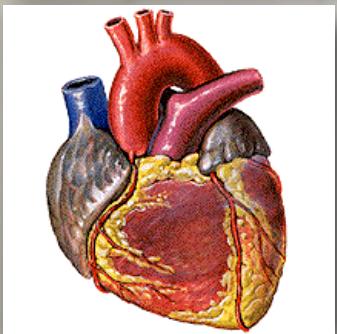
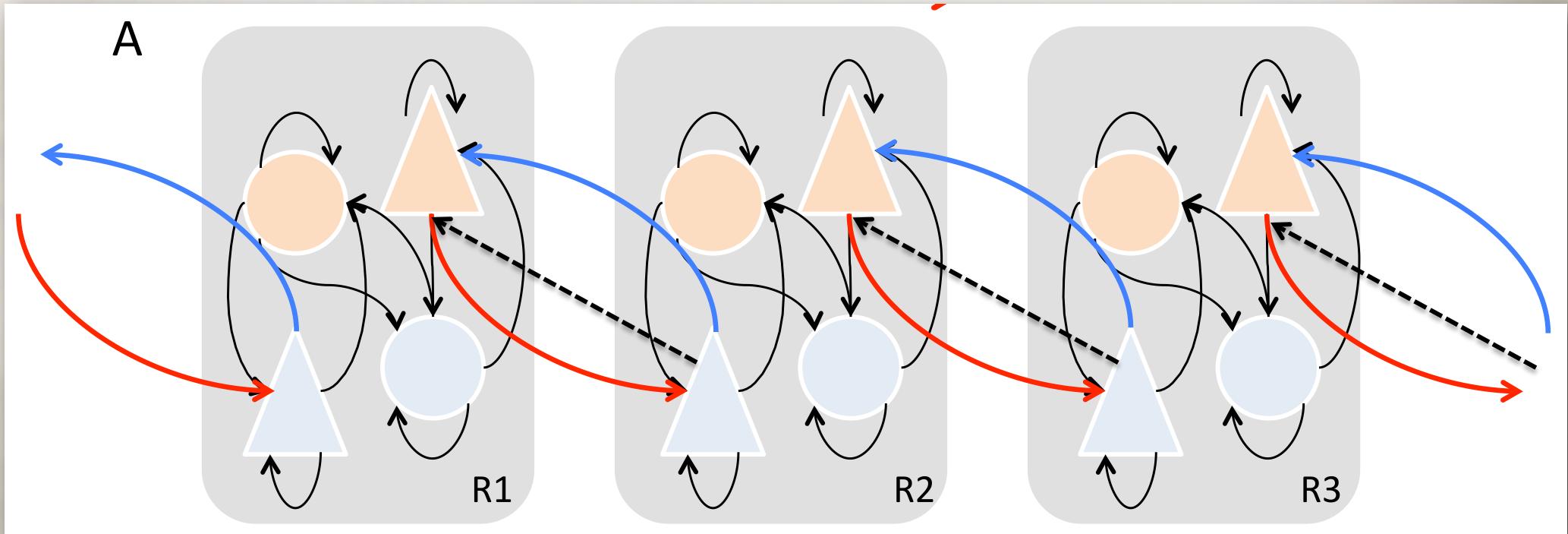
- The sense of the internal physiological condition of the body:



# Interoceptive inference

→

Interoceptive prediction error signals



Predicted interoceptive signals

←

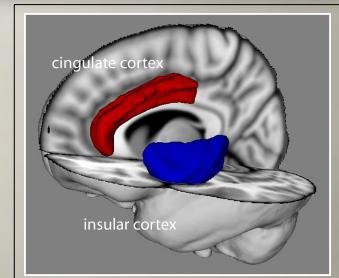


FIGURE 2 | The human cingulate (red) and insular (blue) cortices. Image generated using Mango (<http://hc.uthscsa.edu/mango>).

Paulus & Stein (2006) *Biol Psych*

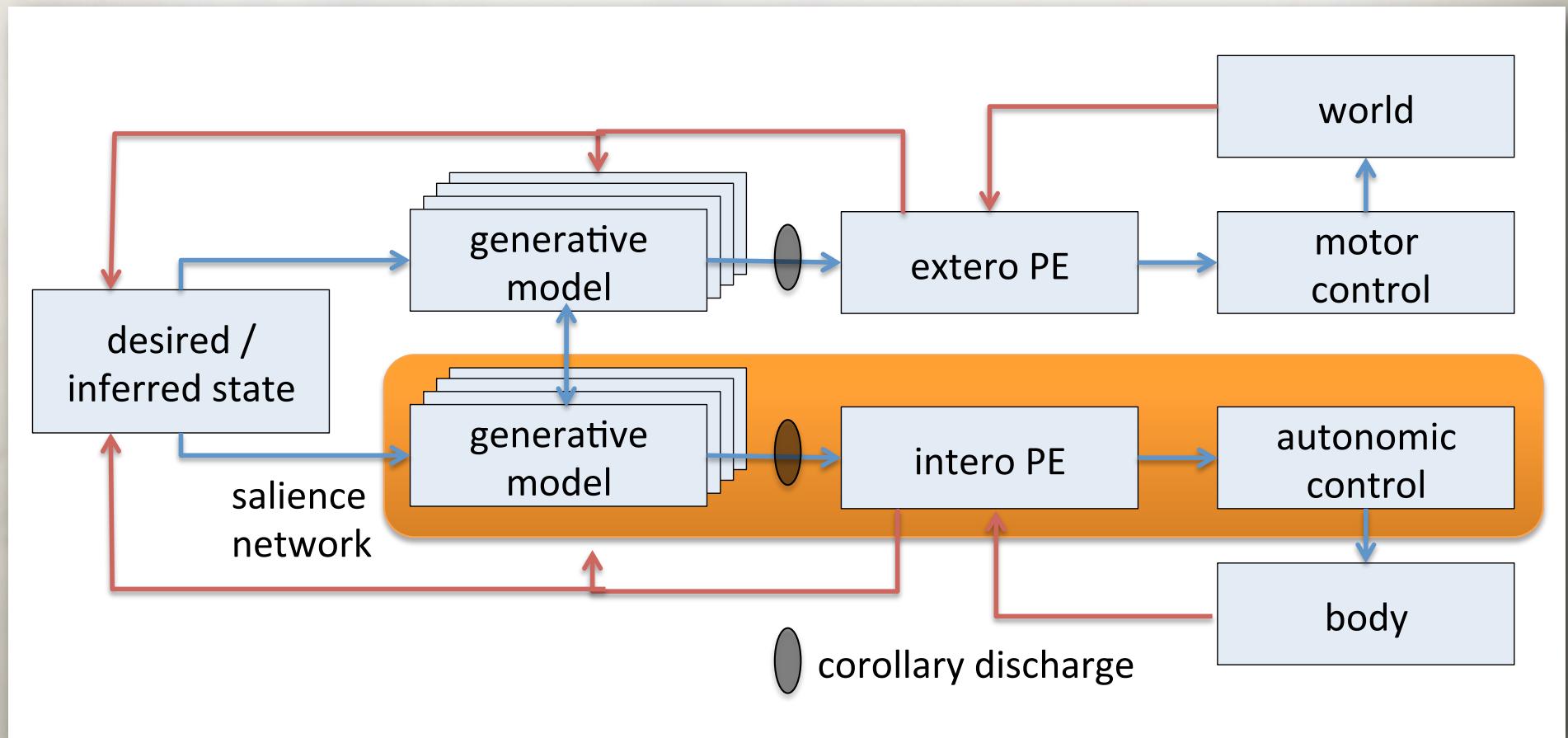
Pezzulo (2014) *CABN*

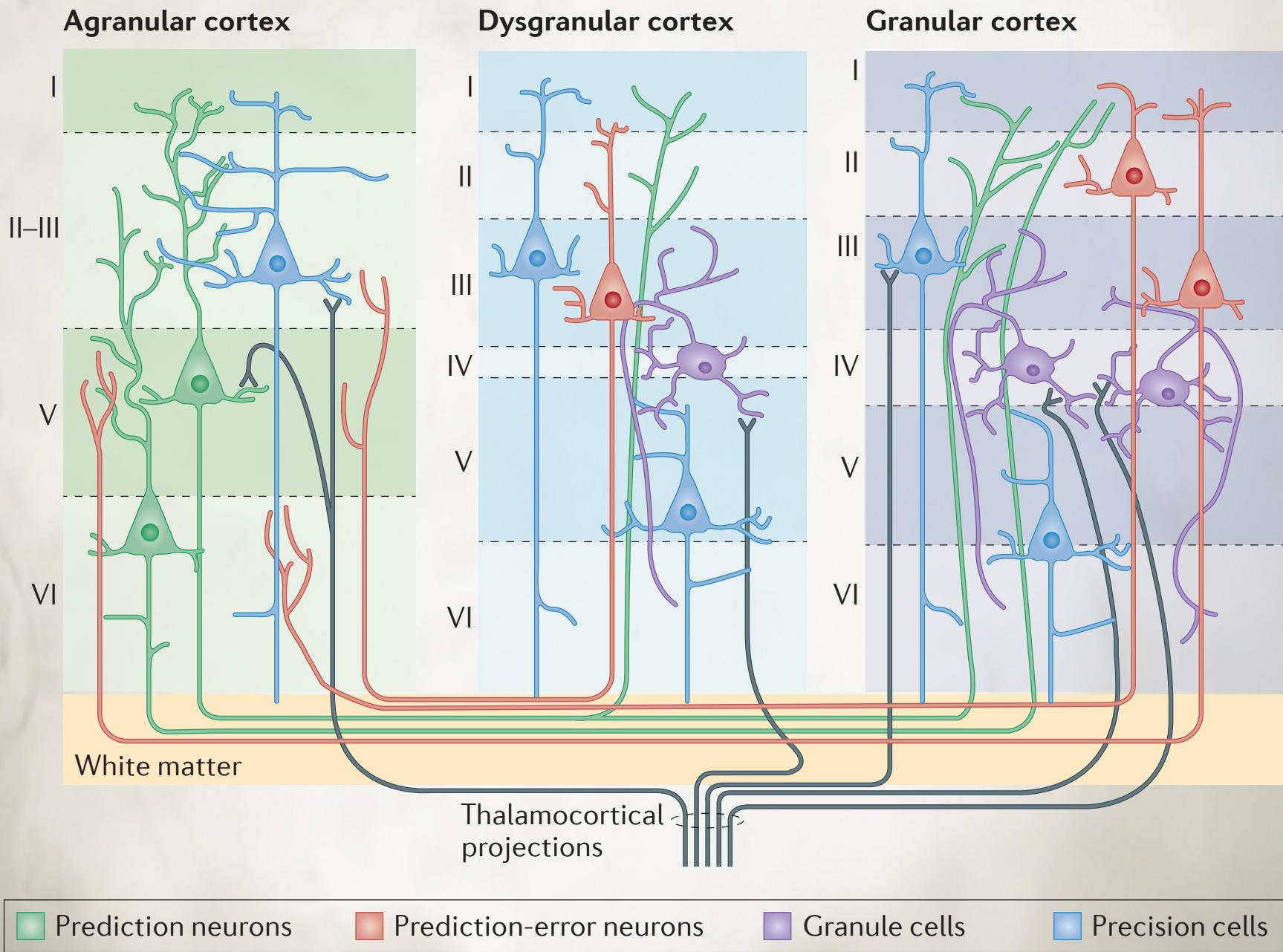
Seth (2013) *Trends Cog Sci*

Seth et al (2012) *Front. Psychol.*

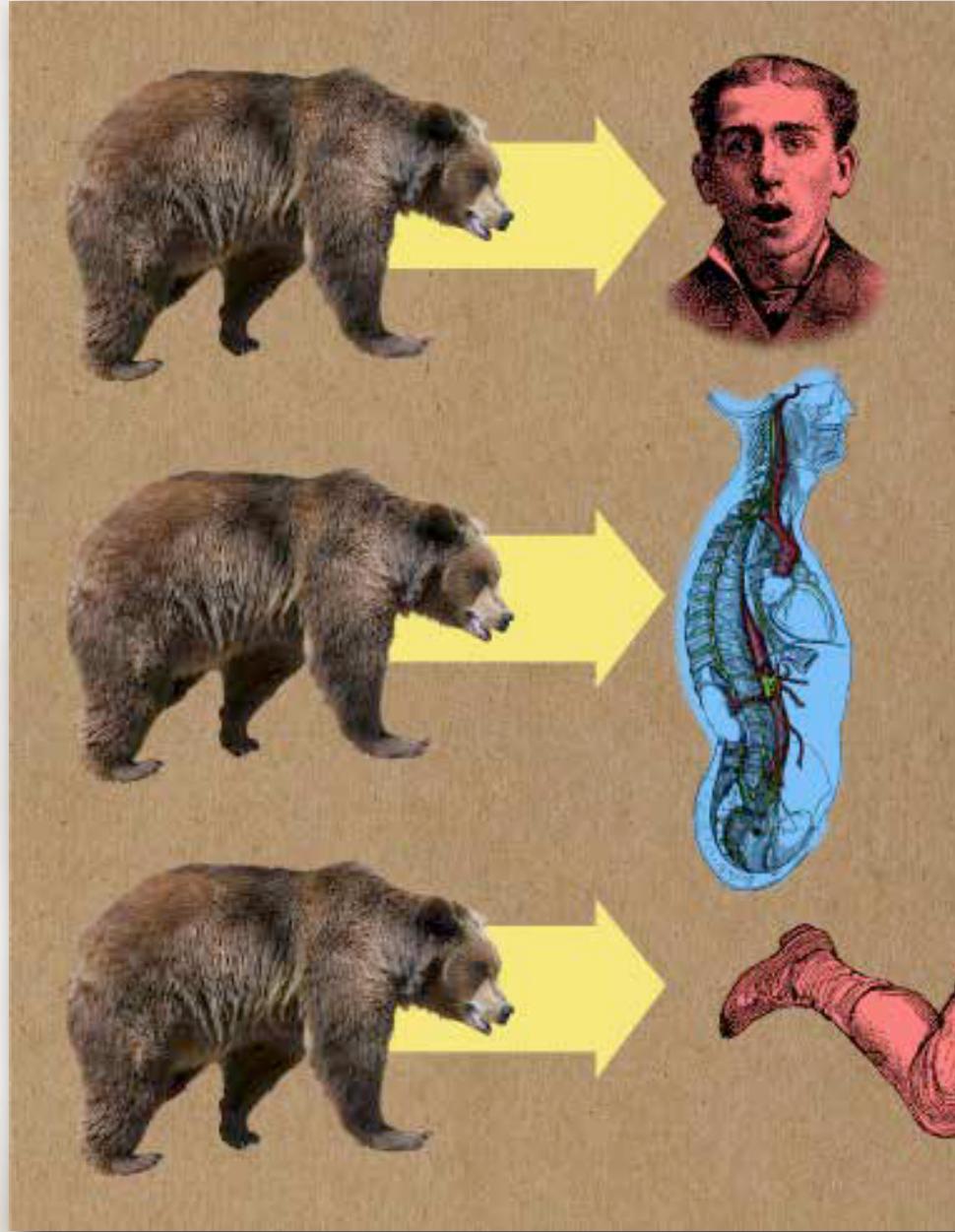
# Interoceptive inference

- Brain network centred on a ‘salience network’ including the anterior insula and cingulate cortices.





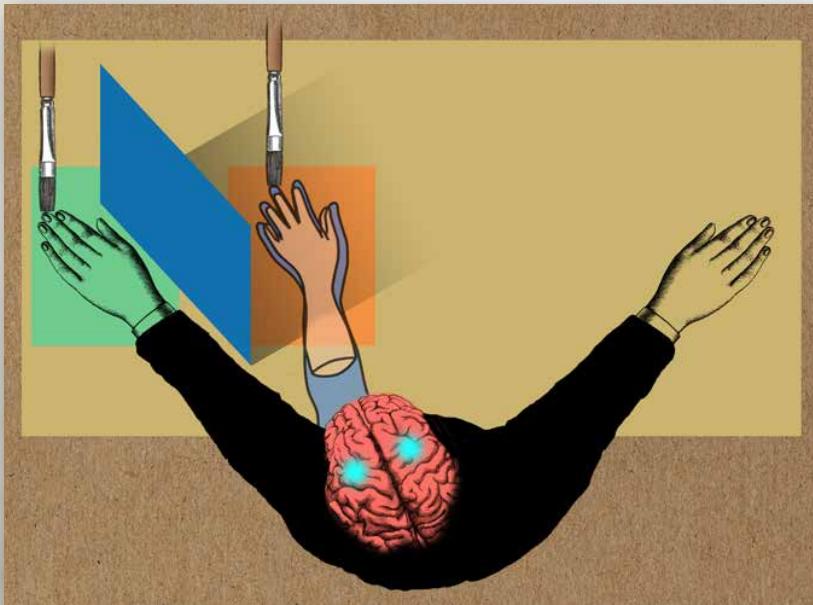
# Interoceptive inference and emotion



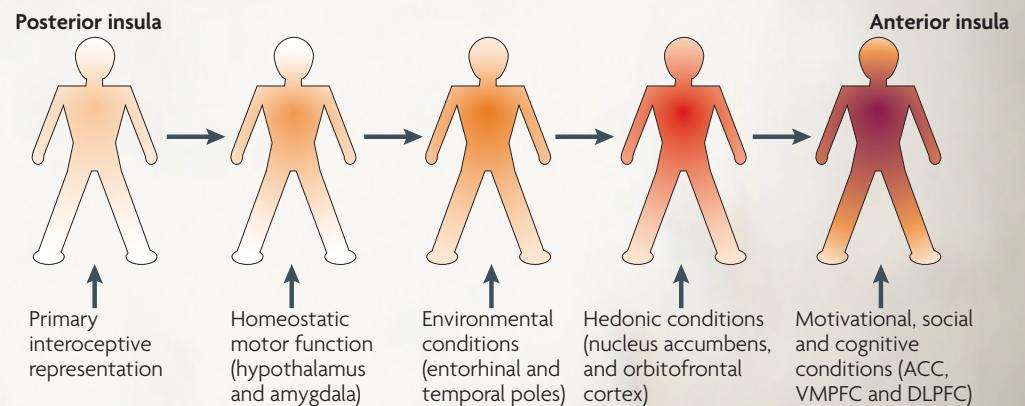
*30 Second Brain, Ivy Press*

# Embodied selfhood

- One tradition - perception of the body **from the outside**.
- A second tradition - perception of the body **from within**.



Rubber hand illusion



Bud Craig's “material me”

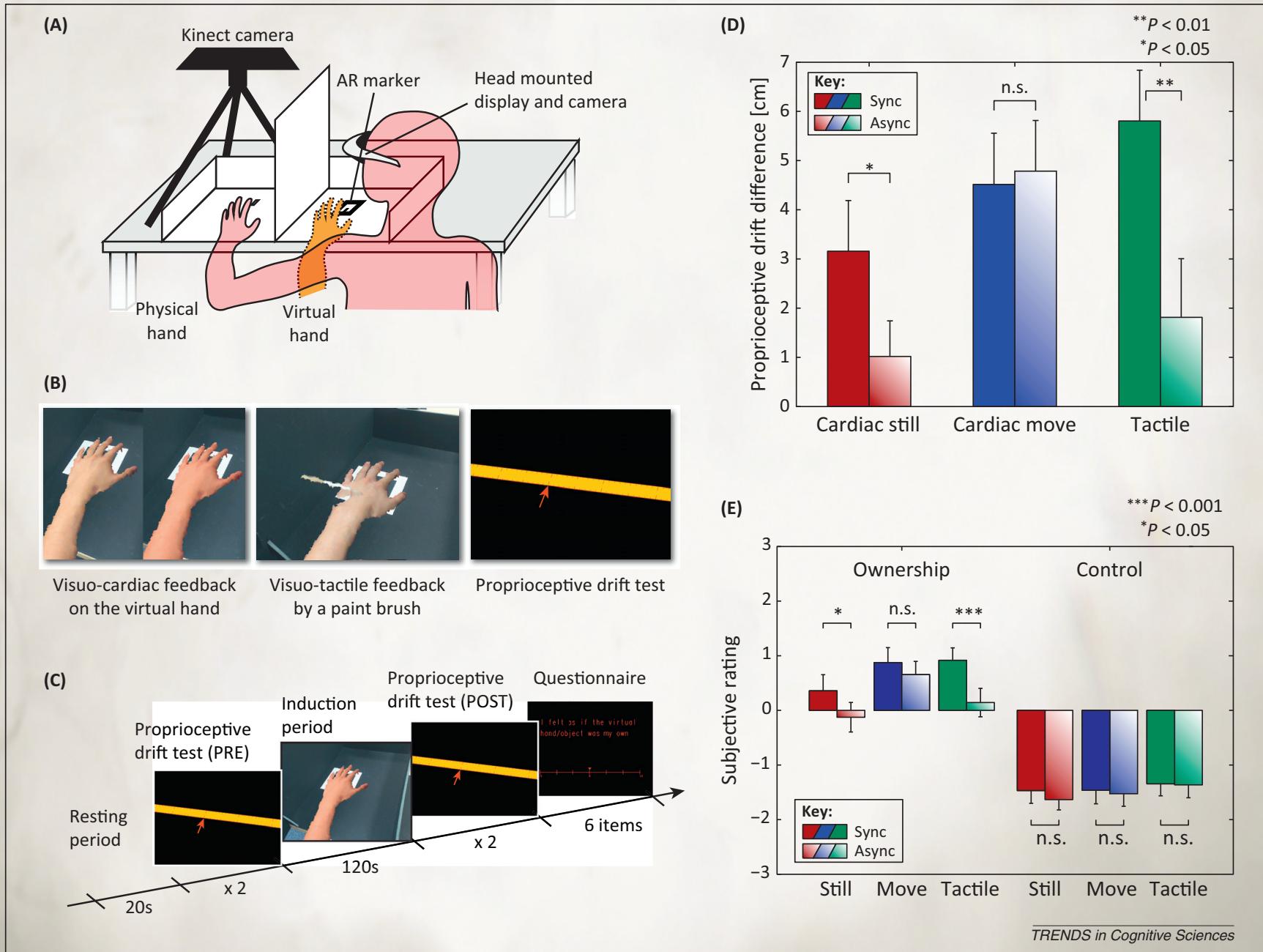
Craig (2009) *Nat Neuro Rev*

Botvinick & Cohen (1998) *Nature*



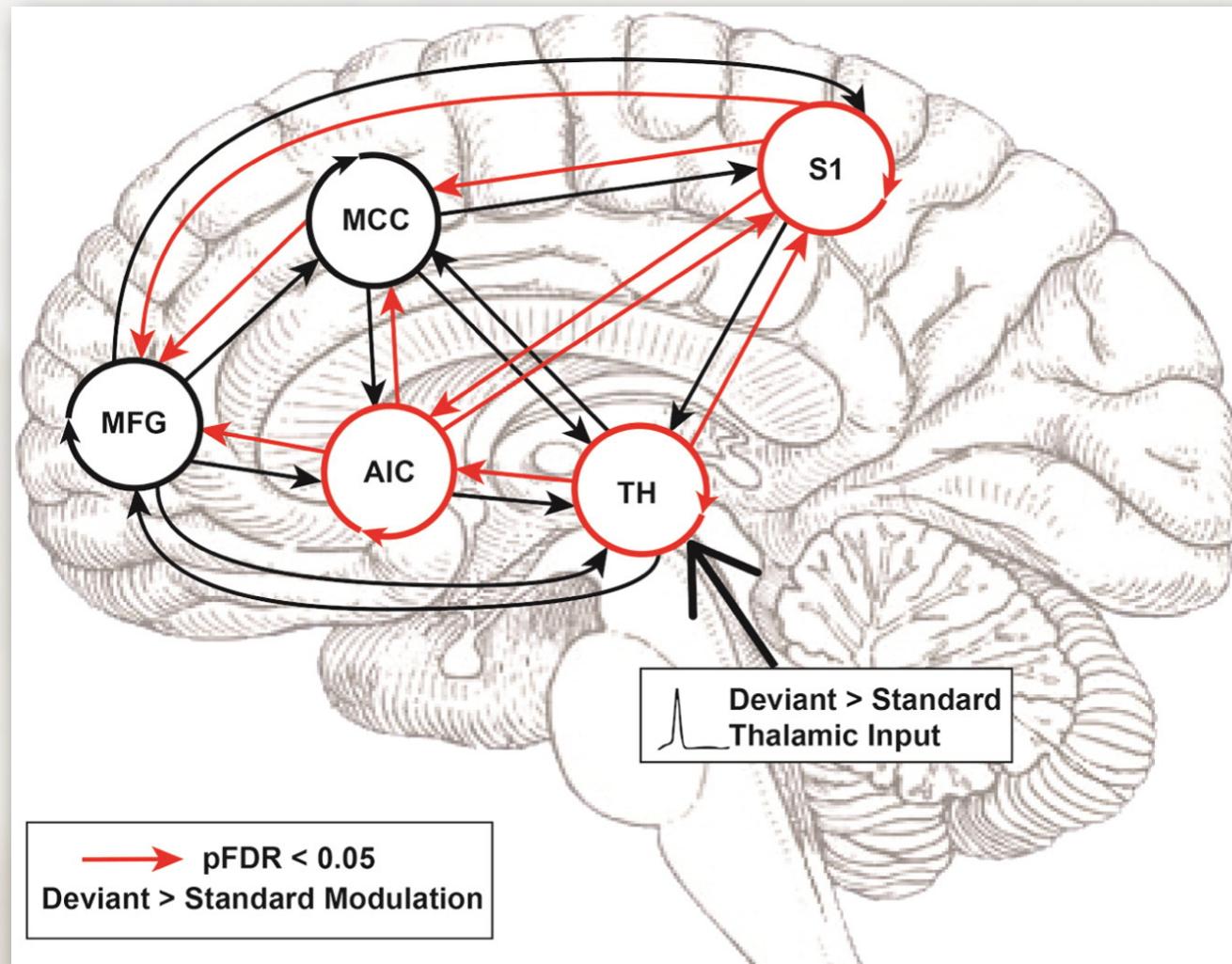
# Cardiac Feedback





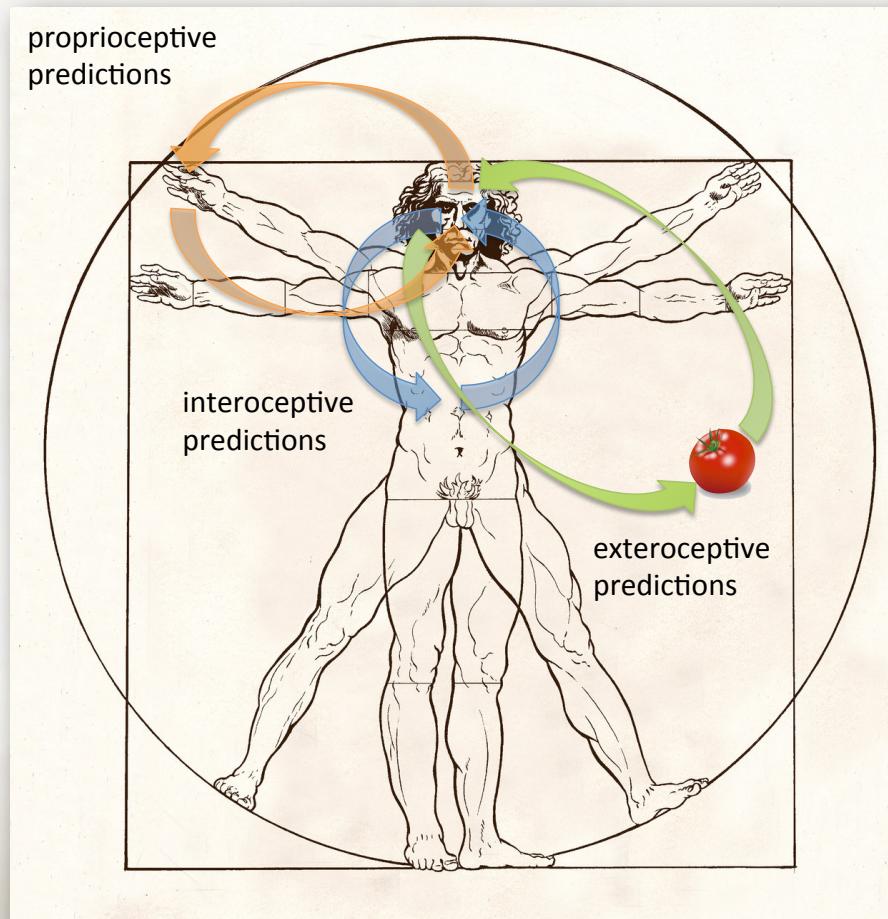
# Tactile mismatch responses

- “AIC coordinates hierarchical processing of tactile prediction error”.



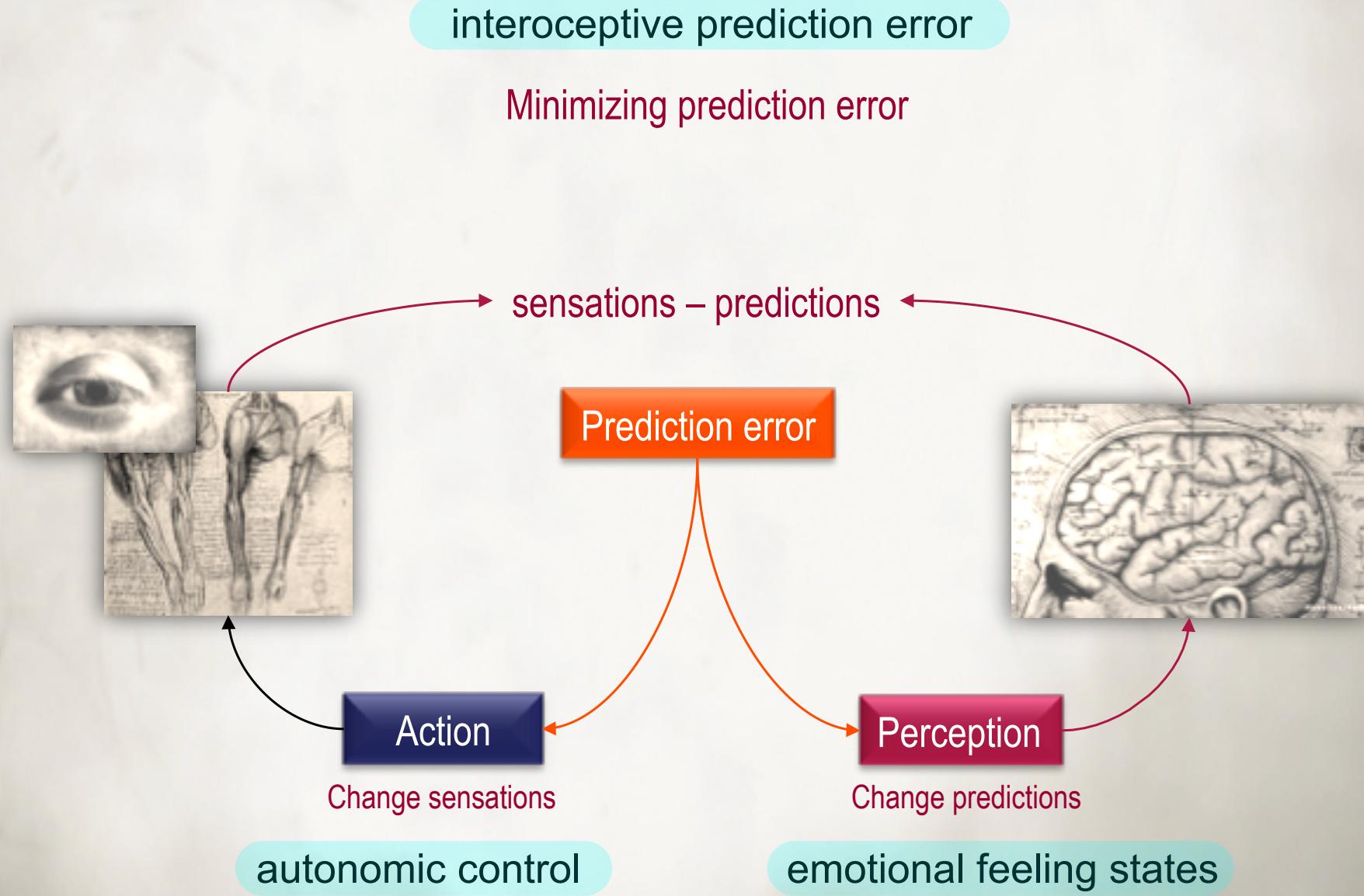
## Interim summary

- Embodied selfhood depends on control-oriented active inference in predictive models of the causes of self-related signals, across exteroceptive, proprioceptive, and interoceptive domains.

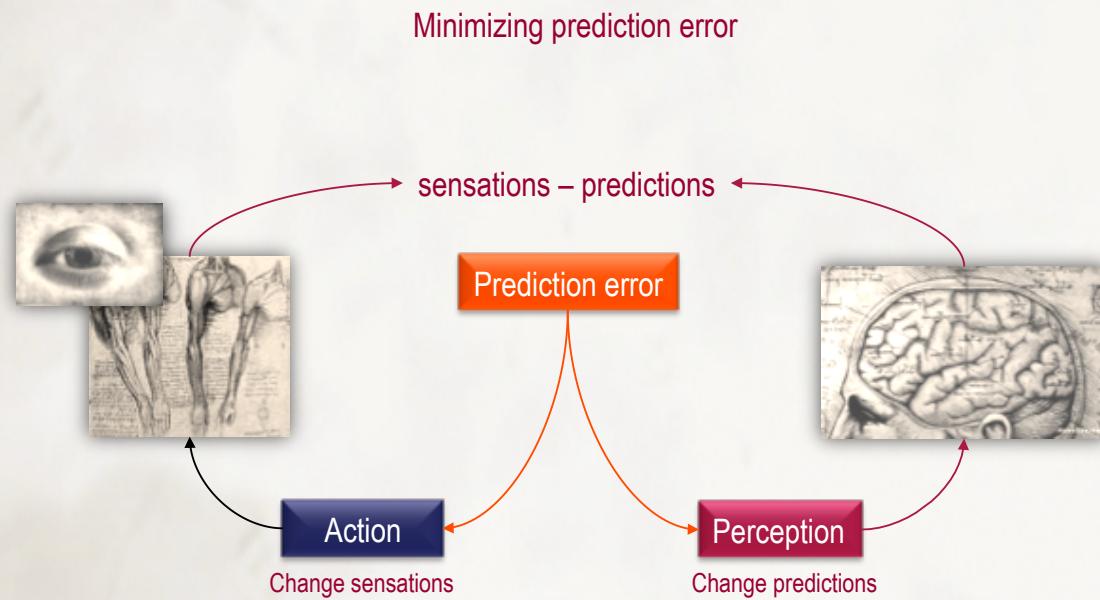


**inference to the best prediction**

# Active inference



# Active inference



... implies **counterfactual** predictive processing

# Epistemic and instrumental active inference

- Epistemic active inference deploys action to discover more about hidden causes of the sensorium.
- Instrumental active inference deploys action to maintain (essential) variables within viability bounds.



Helmholtz:  
view from inside the black box

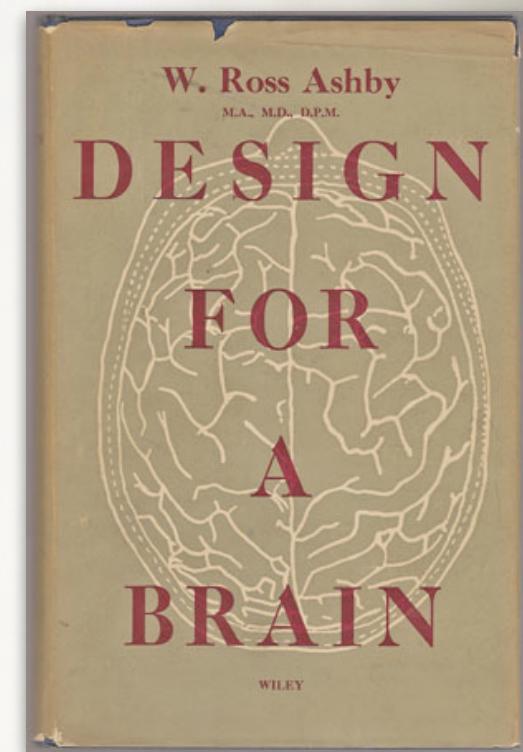
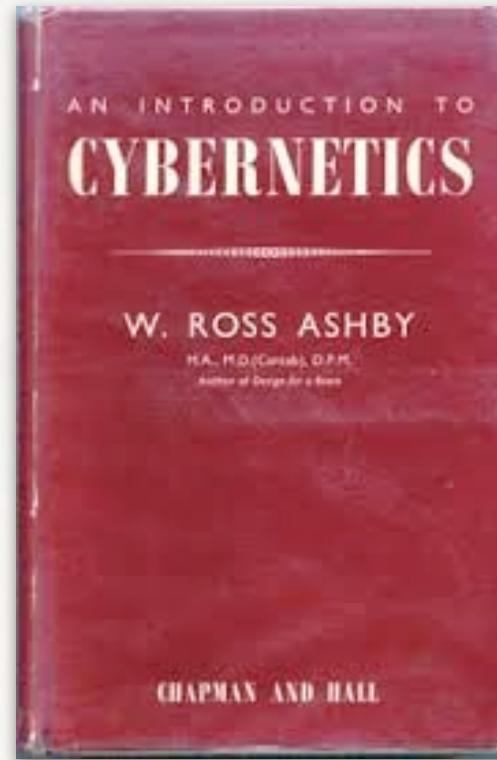


Ashby:  
good regulator theorem

# W. Ross Ashby

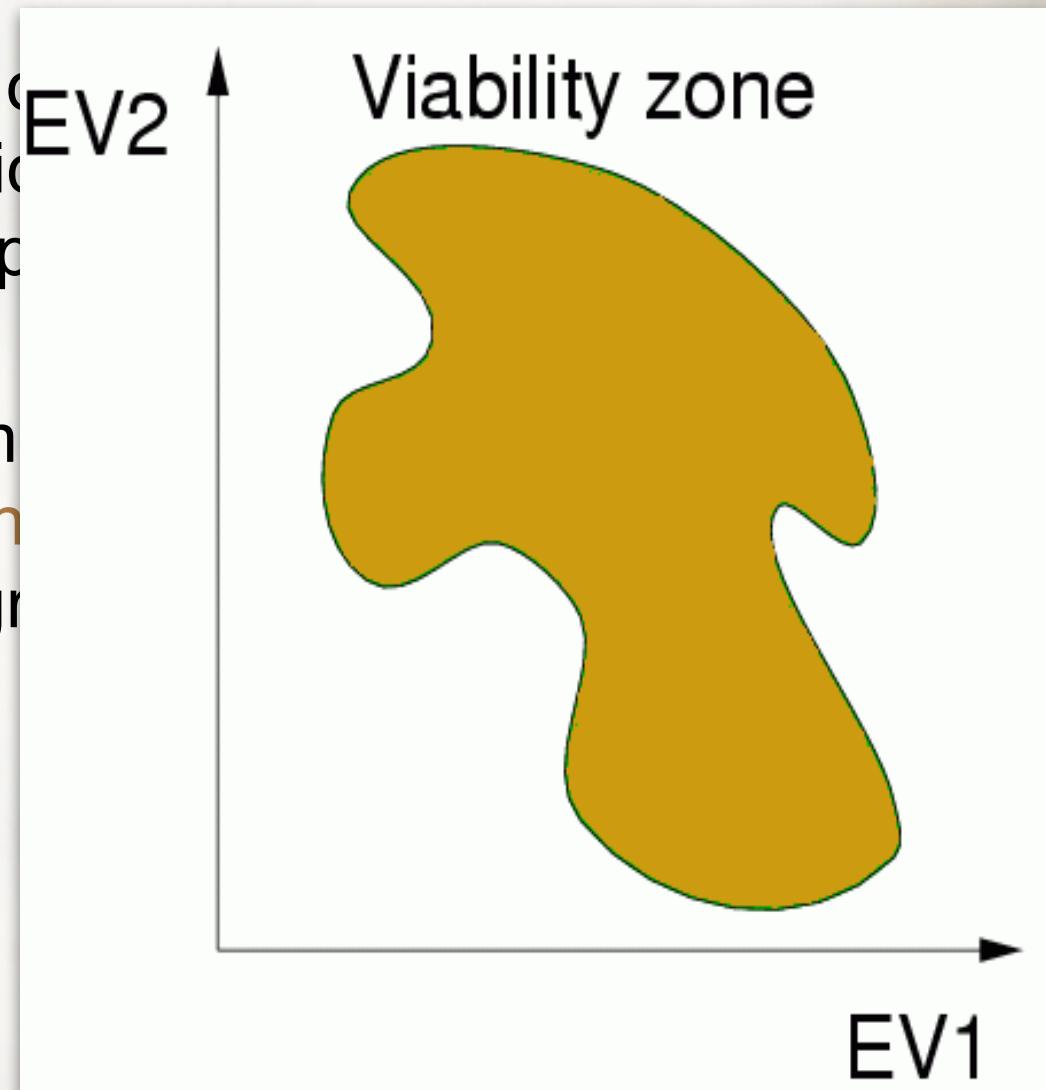


1903 - 1972



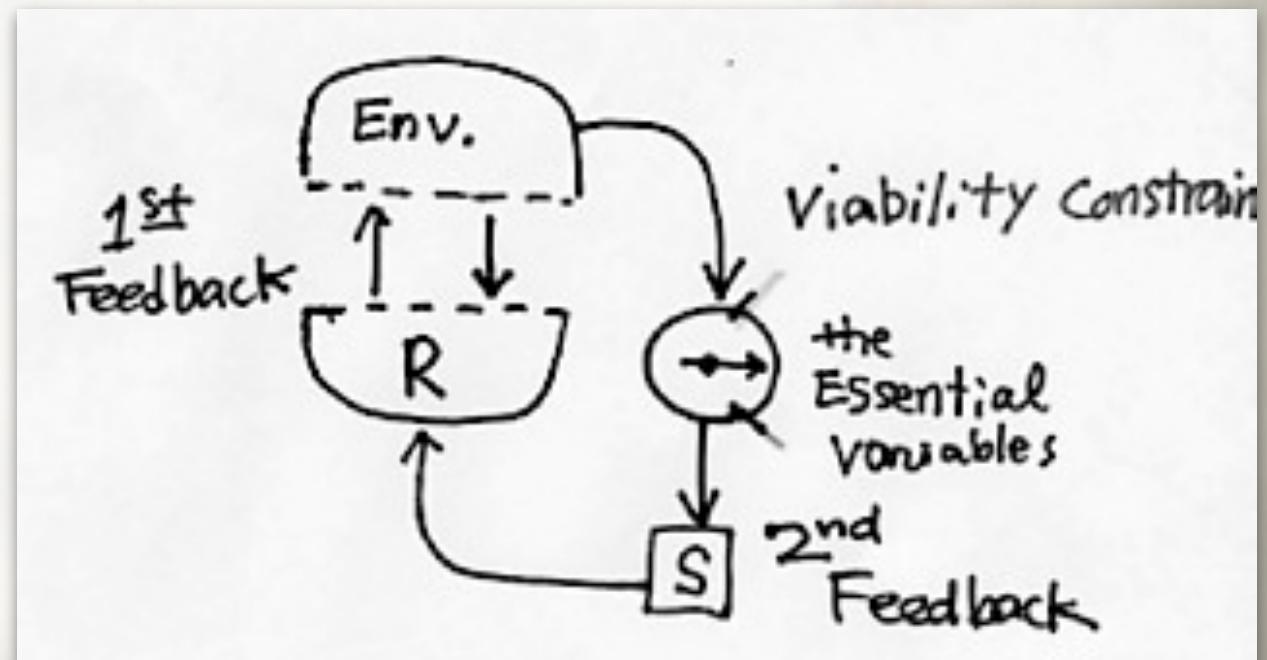
# Cybernetics

- The prediction and control of behaviour in teleological machines.
- The essential variables concern a related set of physiologically survival (e.g., body temperature, food intake, etc.)
- In order for an organism to survive, its internal environment must be maintained within certain limits. If this limit is exceeded, the organism faces disintegration.



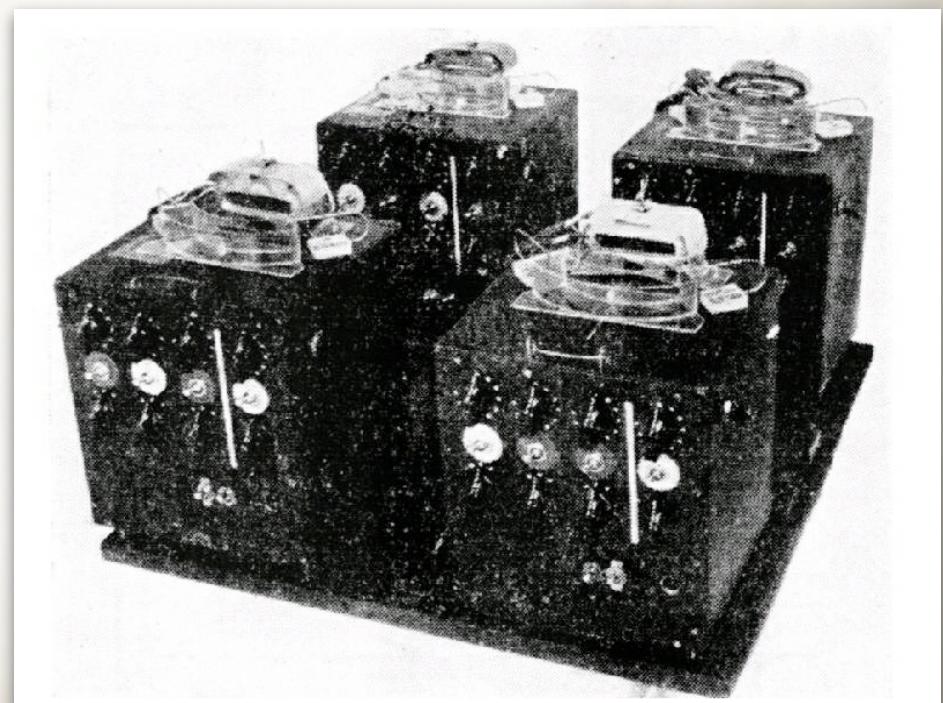
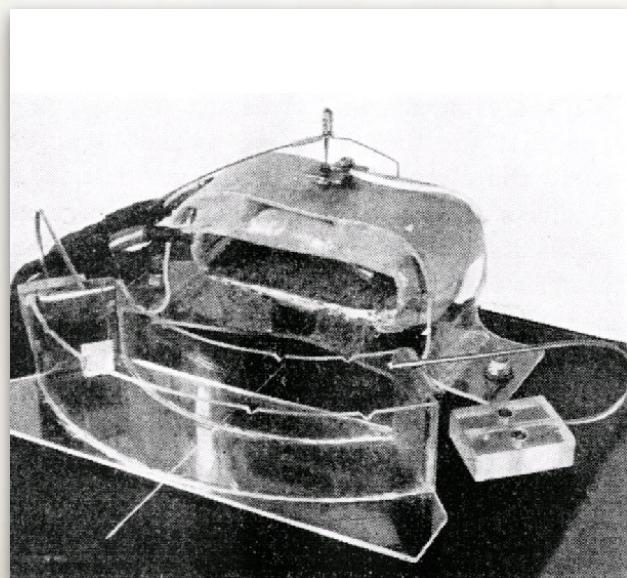
# Ultrastability

- How can a system remain within viability limits, given a limited ability to respond to perturbations?
- “Dynamic systems stop breaking when, and only when, they reach a state of equilibrium ...
- ... all dynamic systems change their internal organisation spontaneously until they arrive at some state of equilibrium”



## The homeostat

- An electromagnetic device consisting of 4 ultrastable systems that could be coupled in different ways.
- “Ashby’s brilliant idea of the unpurposeful random mechanism which seeks for its own purpose through a process of learning is ... one of the great philosophical contributions of the present day.”



Wiener (1964)

Ashby (1948)

# The good regulator theorem

Int. J. Systems Sci., 1970, vol. 1, No. 2, 89-97

## EVERY GOOD REGULATOR OF A SYSTEM MUST BE A MODEL OF THAT SYSTEM<sup>1</sup>

Roger C. Conant

Department of Information Engineering, University of Illinois, Box 4348, Chicago,  
Illinois, 60680, U.S.A.

and W. Ross Ashby

Biological Computers Laboratory, University of Illinois, Urbana, Illinois 61801,  
U.S.A.<sup>2</sup>

[Received 3 June 1970]

The design of a complex regulator often includes the making of a model of the system to be regulated. The making of such a model has hitherto been regarded as optional, as merely one of many possible ways.

In this paper a theorem is presented which shows, under very broad conditions, that any regulator that is maximally both successful and simple *must* be isomorphic with the system being regulated. (The exact assumptions are given.) Making a model is thus necessary.

The theorem has the interesting corollary that the living brain, so far as it is to be successful and efficient as a regulator for survival, *must* proceed, in learning, by the formation of a model (or models) of its environment.

## Interoceptive active inference

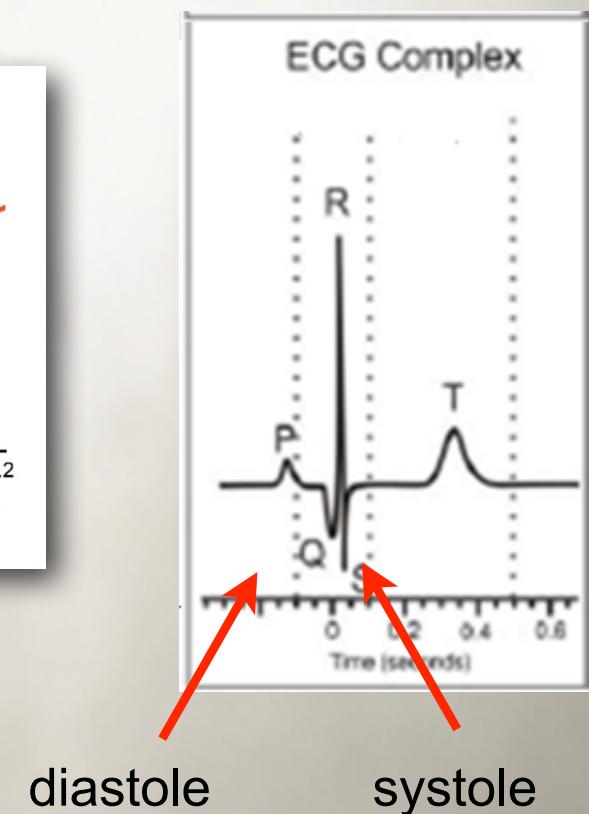
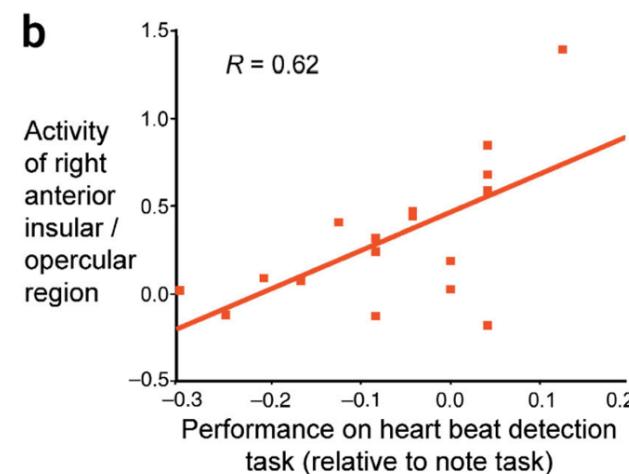
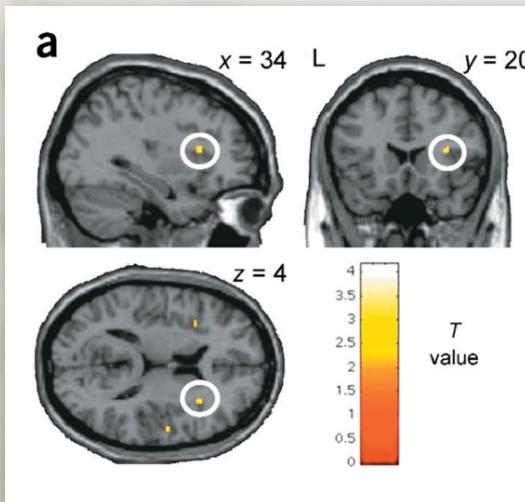
- Interoception and autonomic regulation emphasise **instrumental active inference**.
- One does not want to push one's essential variables to the limits, just to check that they can return.
- At the same time, a model of sensory causes is needed in order to ensure homeostasis and allostasis (predictive regulation) of EVs.
- cf. Karl Friston “I am therefore I think.”

# **interoceptive performance in autism and psychosis\***

**\*very much in progress**

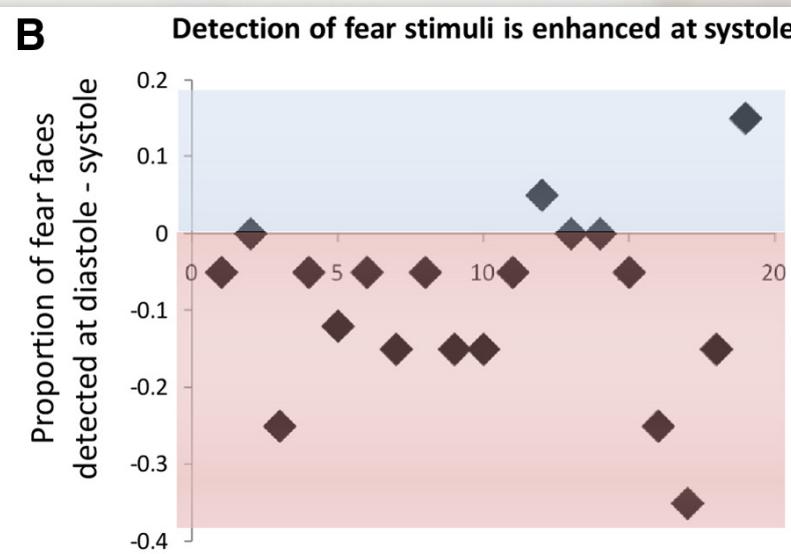
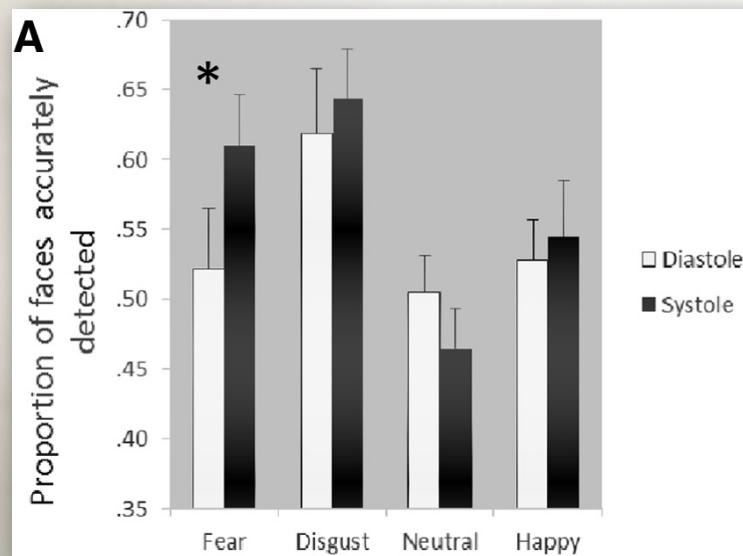
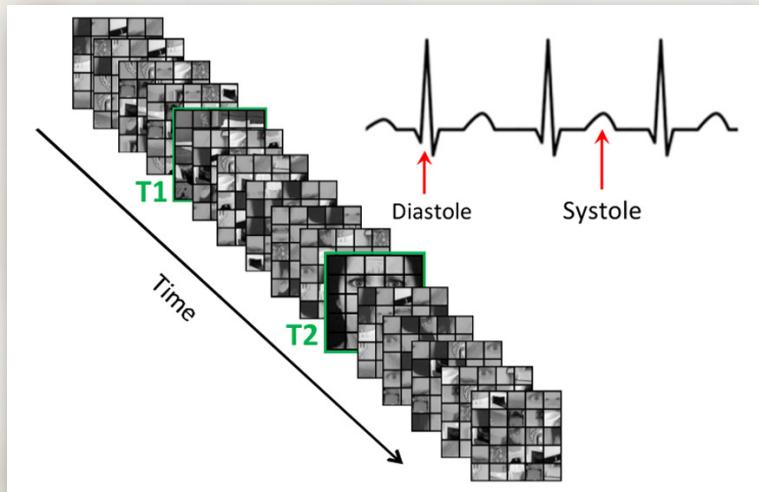
# Interoceptive (cardiac) feedback

- Interoceptive psychophysics is hard!
- Cardiac feedback provides an experimentally accessible channel for interoceptive signalling.
- Anterior insula activity is associated with interoceptive sensitivity.



# Interoceptive feedback and emotion perception

- Emotional ‘attentional blink’ shows breakthrough of emotional stimuli is modulated by cardiac cycle.



# Multiple levels of interoceptive report

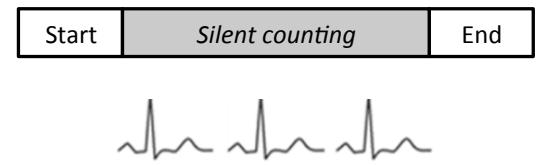
## Dimensions of Interoception

Interoceptive accuracy

*Objective performance*

Performance on  
behavioural tests

e.g. heartbeat detection tests



Schandry (1981)

Interoceptive sensibility

*Subjective belief*

Self report

e.g. questionnaire

BODY PERCEPTION QUESTIONNAIRE

Read the instructions for each sub-test and answer (a - e) next to each item

**I: AWARENESS (Imagine how aware you are of your body processes)**

Select the answer that most accurately describes you. Rate your awareness on each of the characteristics described below using the following 5-point scale:

a) never b) occasionally c) sometimes d) usually e) always

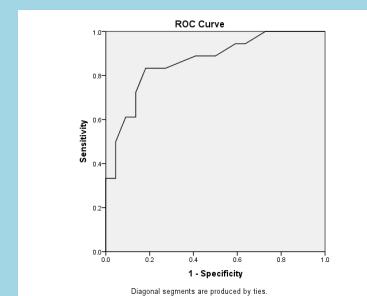
Porges (1993)

Interoceptive awareness

*Metacognitive accuracy*

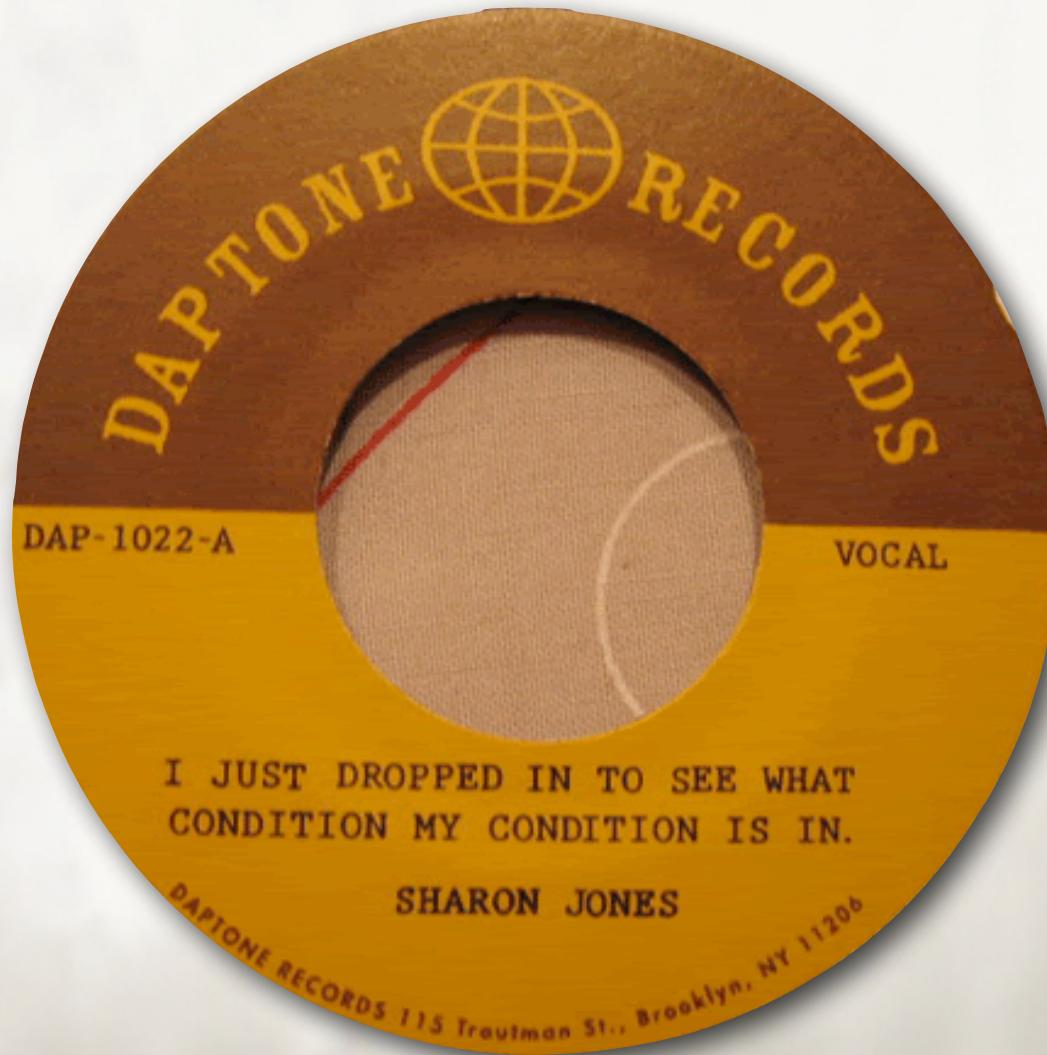
Insight into performance  
aptitude

e.g. confidence-accuracy correspondence,  
area under ROC curve

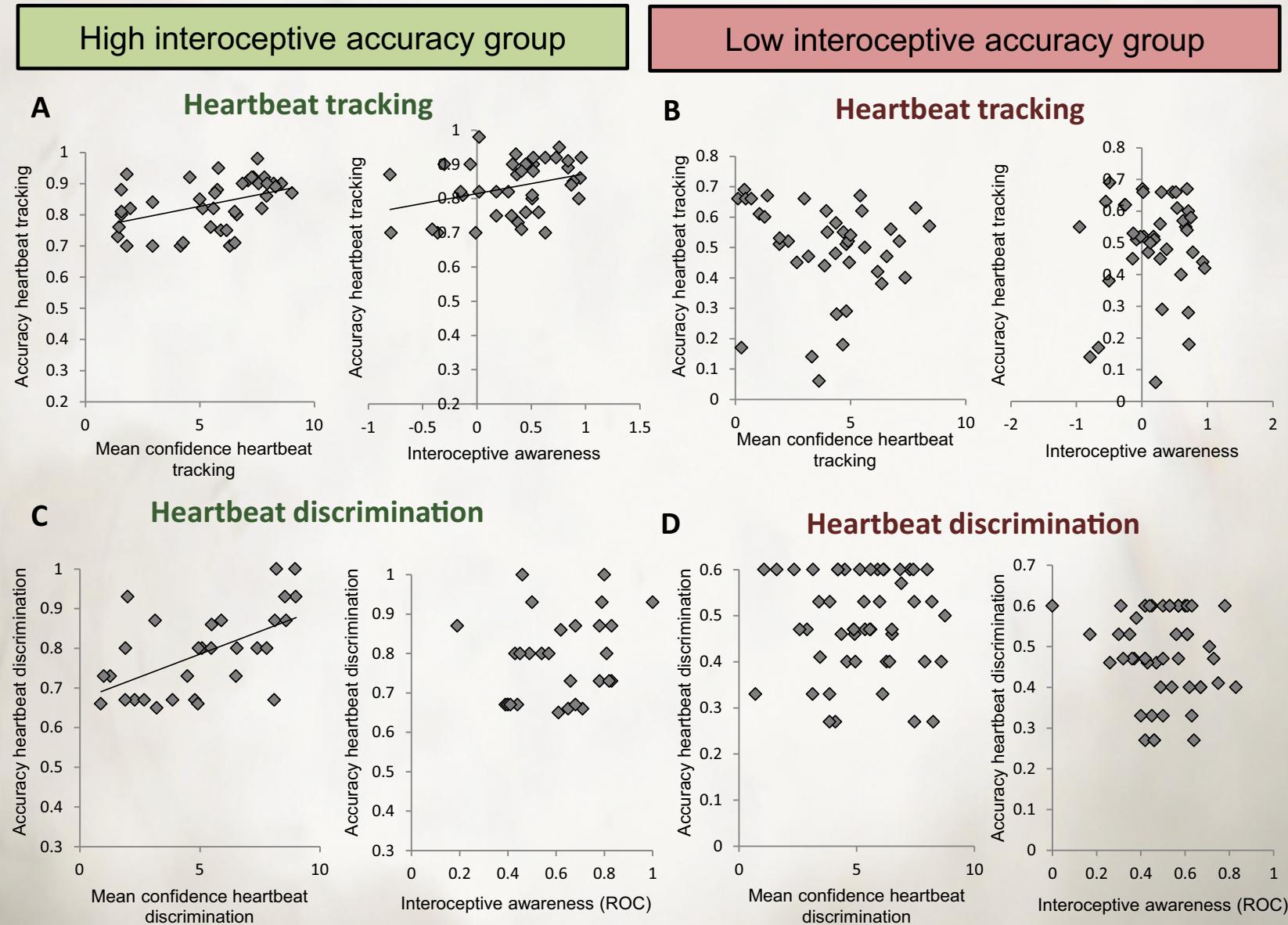


Garfinkel et al. (2015)

**“I just dropped in, to see  
what condition my condition was in”**



# Accuracy, awareness, and sensibility



## Accuracy, awareness, and sensibility

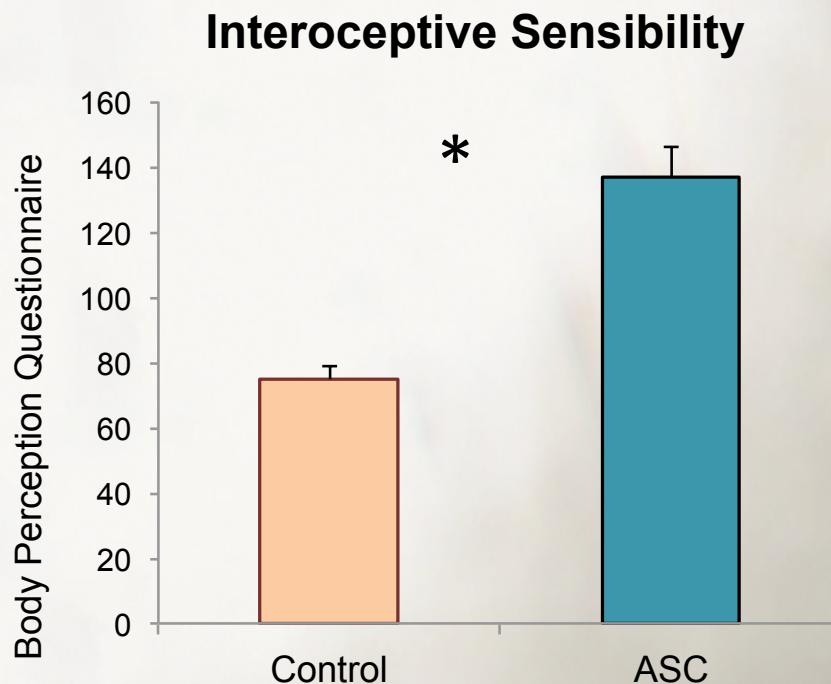
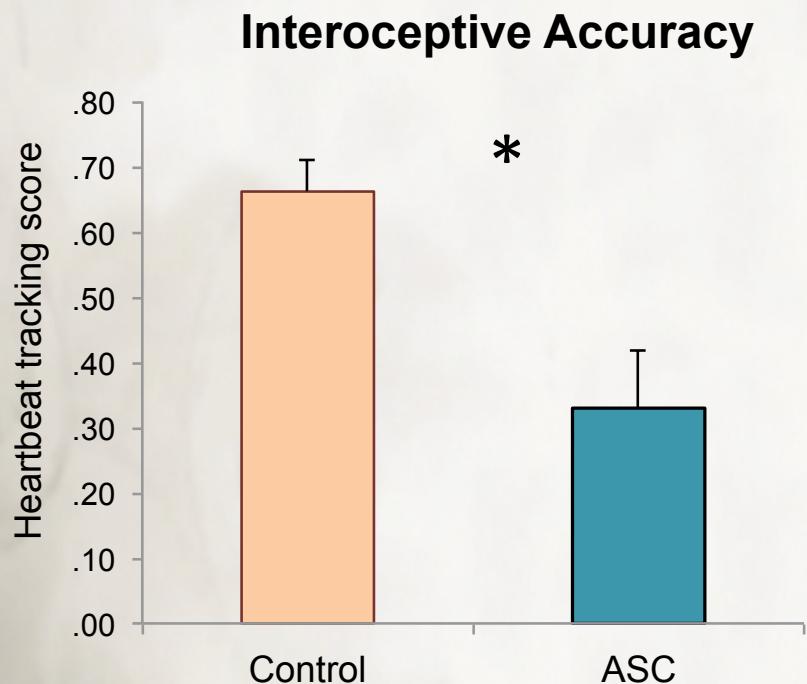
- Different levels of interoceptive representation are both related and dissociable.
- Strong correlations between different levels only emerge for individuals with high interoceptive accuracy.

## Autistic spectrum condition (ASC)

- ASC involves emotional processing and abnormal insular cortex responses in a variety of emotional tasks.
- Thus, ASC may be associated with **impaired** interoceptive ability.
- However, individuals with ASC often report **heightened sensitivity** to internal bodily sensations.
- We measured interoceptive accuracy, awareness, and sensibility in 20 ASC individuals (+ controls).
- We also measured **anxiety** (Speilberger state/trait) and **emotion sensitivity** (Cambridge empathy quotient)

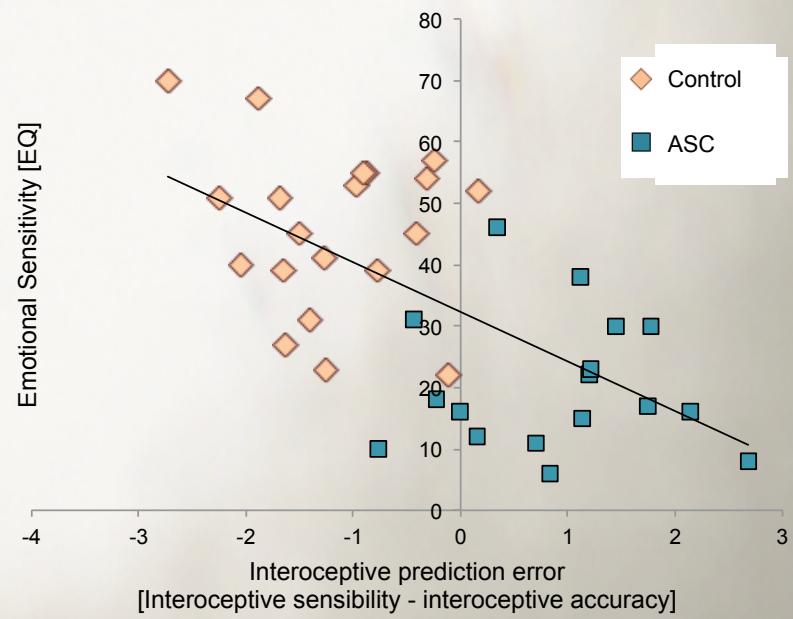
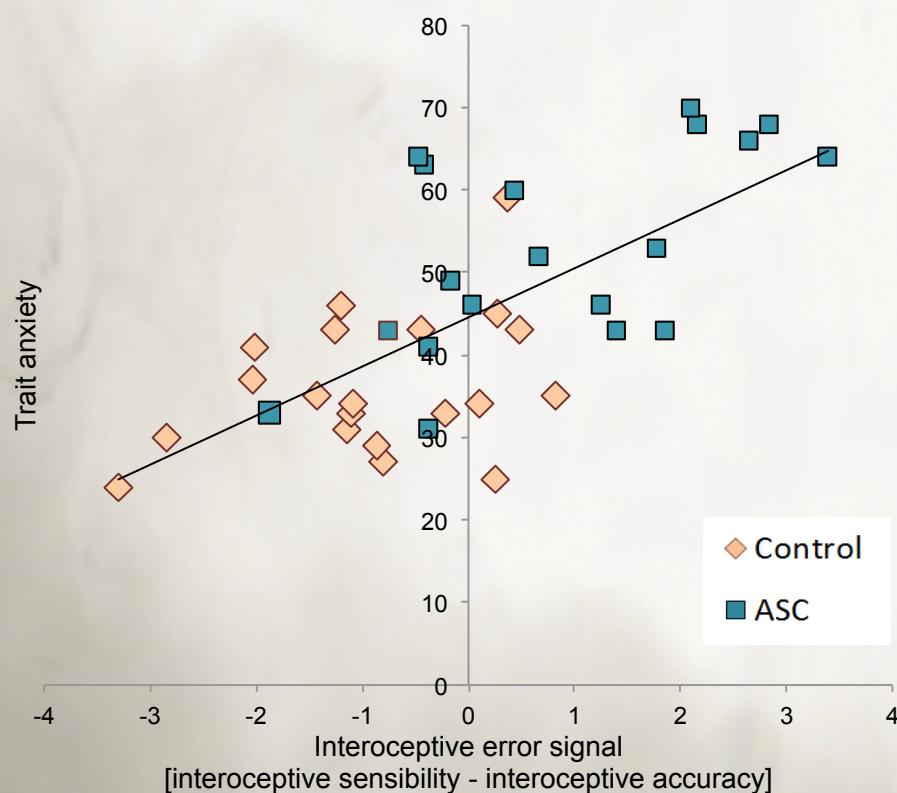
# Interoception and autism

- Prediction: ASC would show reduced interoceptive accuracy but heightened sensibility, compared to controls.



# Interoceptive ‘trait prediction error’

- An ‘interoceptive trait prediction error’ (ITPE) can be defined as the difference between **sensibility** and **accuracy**.
- Positive ITPE: overestimation of interoceptive ability.
- Negative ITPE: underestimation of interoceptive ability.
- Correlation between ITPE and anxiety (>ASC).



Paulus & Stein (2006) *Biological Psychiatry*  
Garfinkel et al (in press) *Biological Psychiatry*

# Psychosis

- Psychosis is sometimes considered “opposite” to ASC.
- ASC has been associated with weak priors (or failure in beliefs about prior precision)
- (Some) Bayesian models of psychosis emphasize overactive priors in shaping perceptual content.
- Not a great deal known about interoceptive (dys)function in psychosis ...
- ... though please correct me.

Wylie & Tregellas (2010) *Schiz Res*

Teufel et al (2015) *PNAS*

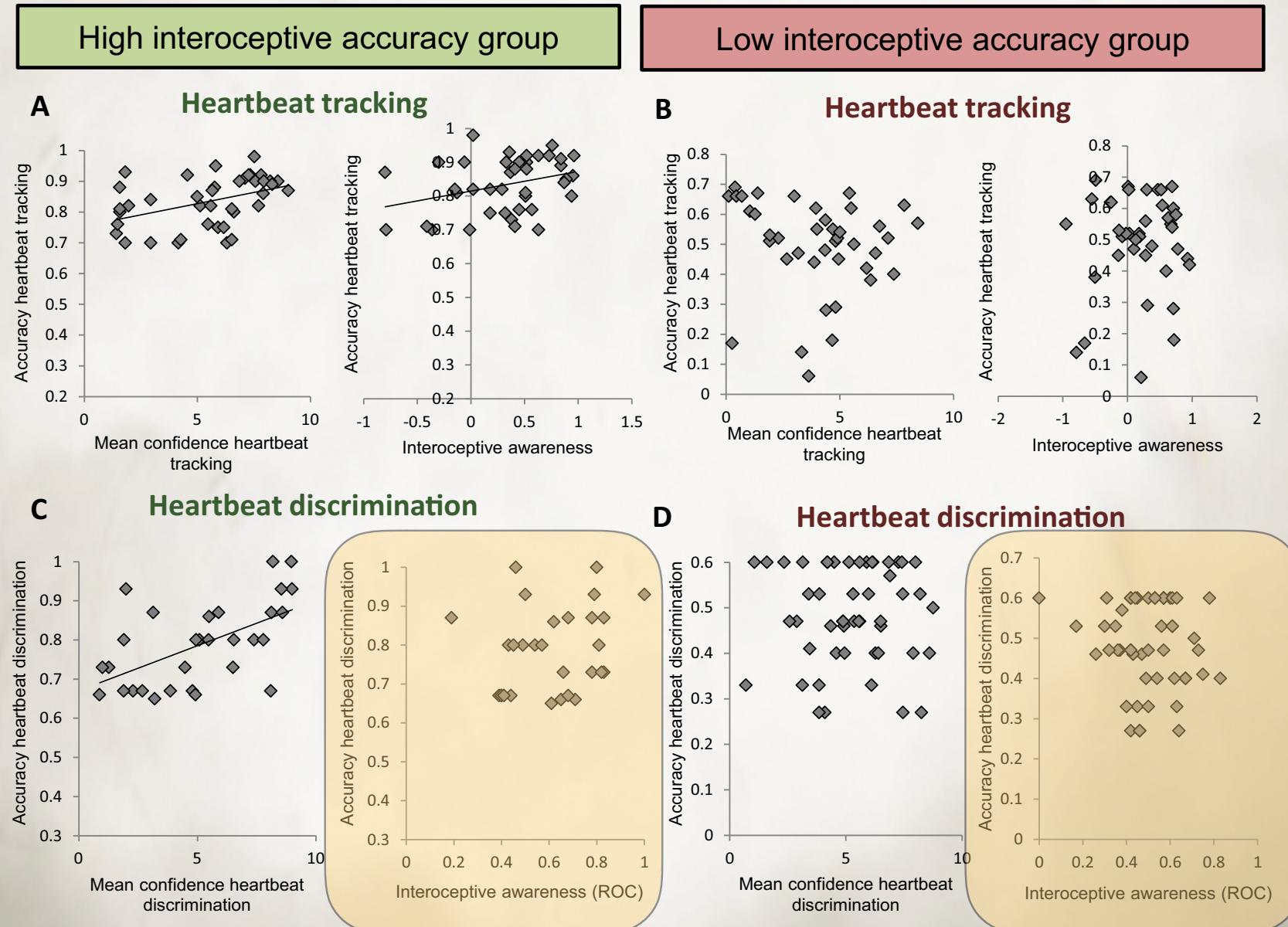
Fletcher & Frith (2009) *Nat Rev Neuro*

Pellicano & Burr (2012) *Trends Cogn Sci*

## Interoception and psychosis

- Examined interoceptive performance in 40 people with first episode psychosis (FEP) & 20 controls
- Initial indications are that FEP is associated with greater alignment in interoceptive dimensions.

# Accuracy, awareness, and sensibility



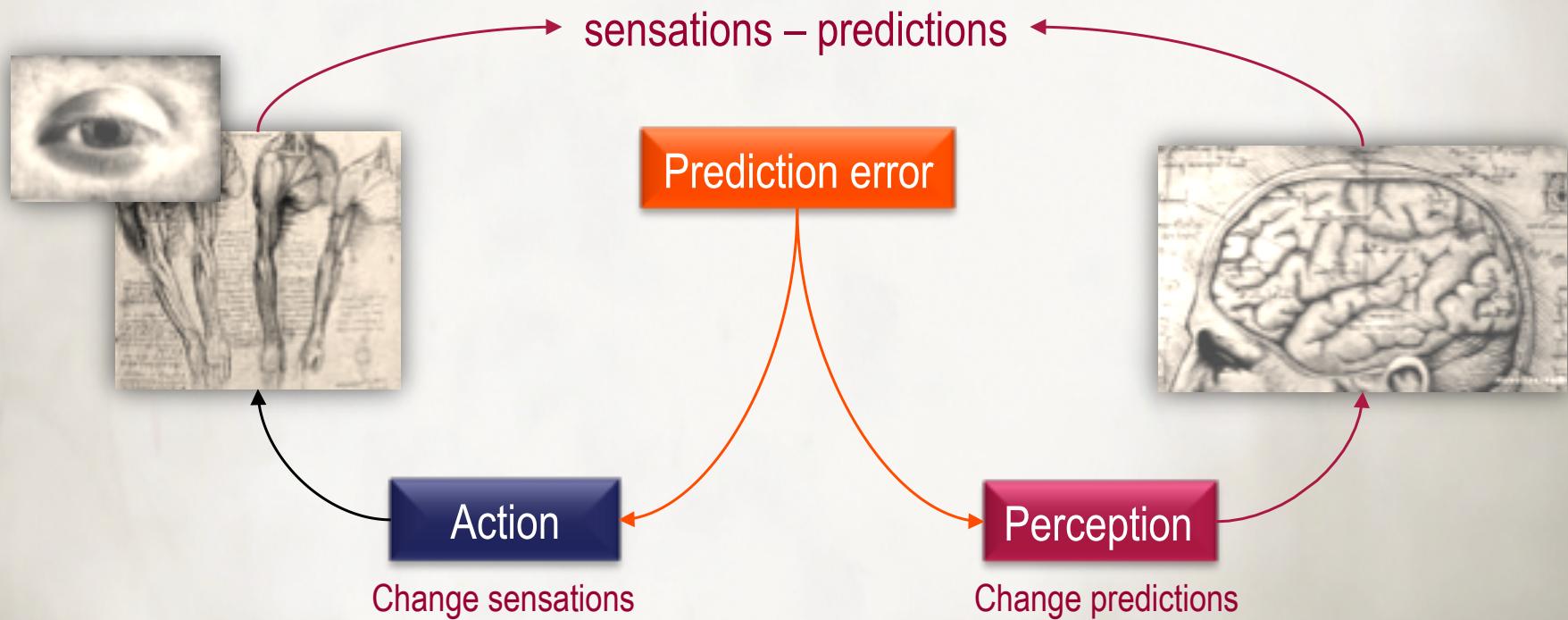
## Open questions

- Do concepts of aberrant (prior) precision in ASC and psychosis translate to the interoceptive domain?
- For example:
  - Does reduced reliance on prior (in ASC) for exteroceptive signals apply also to interoception?
  - Or, does reduced prior precision in one domain lead to an increase in another?
- ASC and psychosis are not unitary conditions and may not admit unitary explanations.

# **counterfactual processing and perceptual presence**

# Active inference

Minimizing prediction error



# Counterfactual predictive processing

- Active inference implies counterfactual processing.
- A counterfactual predictive (generative) model encodes how sensory inputs would change given a repertoire of possible actions, even if those actions are not performed.

Being able to predict what is currently seen also enables us to predict fictive sensations that we could experience from another viewpoint

(Friston et al., 2012, p. 17)

## Perceptual presence

- Perceptual presence is the phenomenological property that obtains when the objects of perception are experienced as real, i.e. experienced as existing as objects in the world.



# Perceptual presence

How can it be true, as I think it is, that we are perceptually aware, when we look at a tomato, of the parts of the tomato which, strictly speaking, we do not perceive. This is the puzzle of perceptual presence.

(Noë, 2006, p. 414)



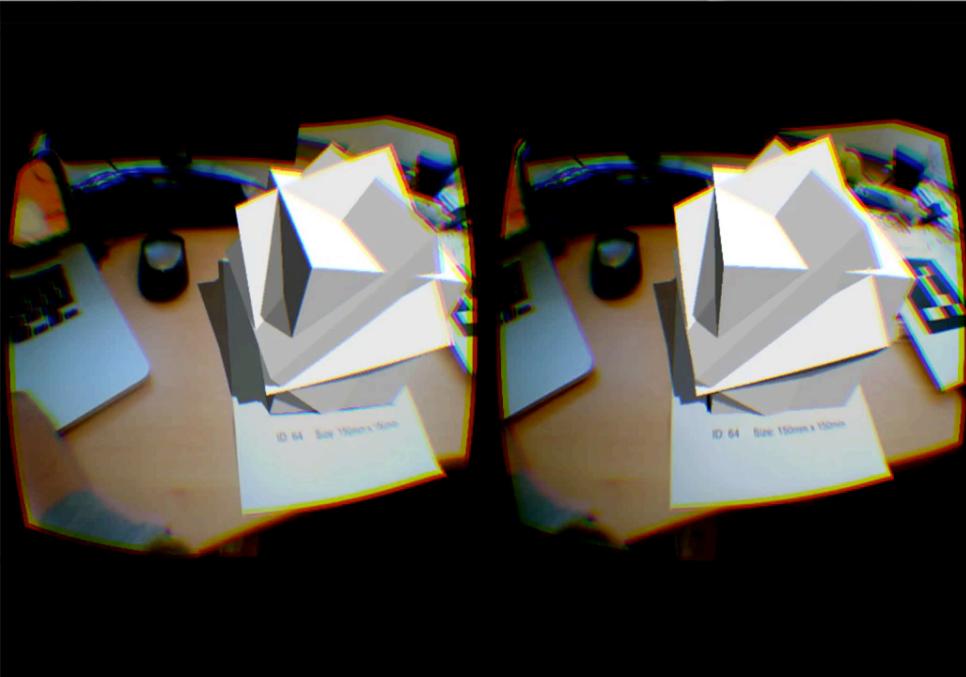
# Sensorimotor theory and perceptual presence

- Sensorimotor theory **explains** perceptual presence as resulting from an implicit knowledge (mastery) of the sensorimotor contingencies governing how sensory signals respond to actions.
- **Example:** the experience of a tomato as a (perceptually present) object is given by implicit (sub-personal) knowledge of (e.g.) the ways in which moving our eyes and bodies would reveal additional sensory information about the tomato.

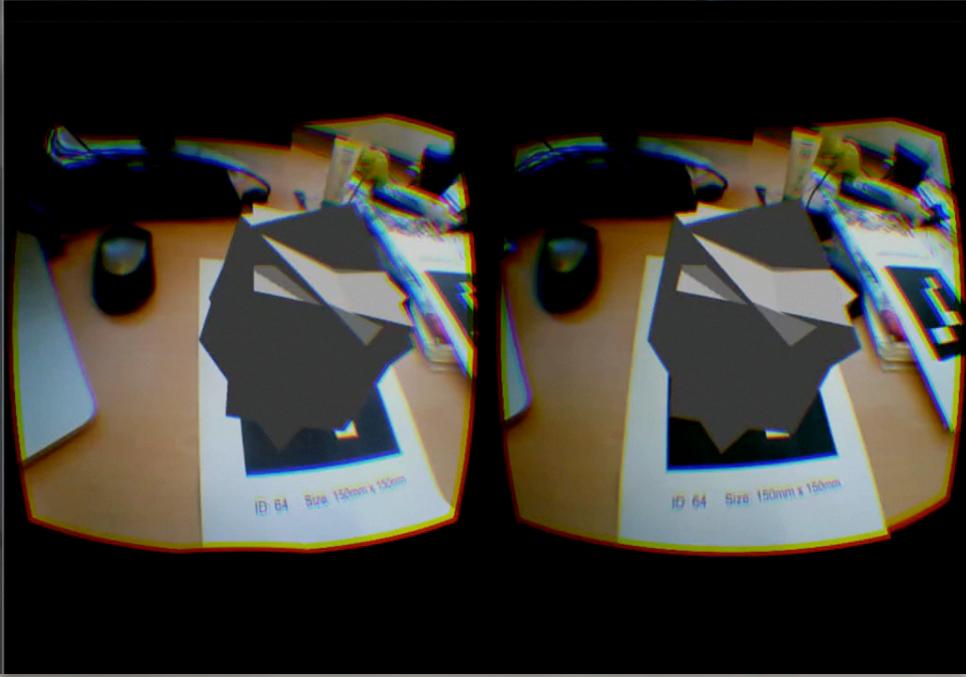


# Predictive processing of sensorimotor contingencies

- Mastery of sensorimotor contingencies is equivalent to active inference via a counterfactually rich generative model.
- If so: perceptual presence will be determined by the richness of counterfactual probability densities.
- That is, perceptual objects will appear subjectively real to the extent that the underlying predictive models encode a rich repertoire of counterfactual sensorimotor dependencies.



veridical



false SMCs

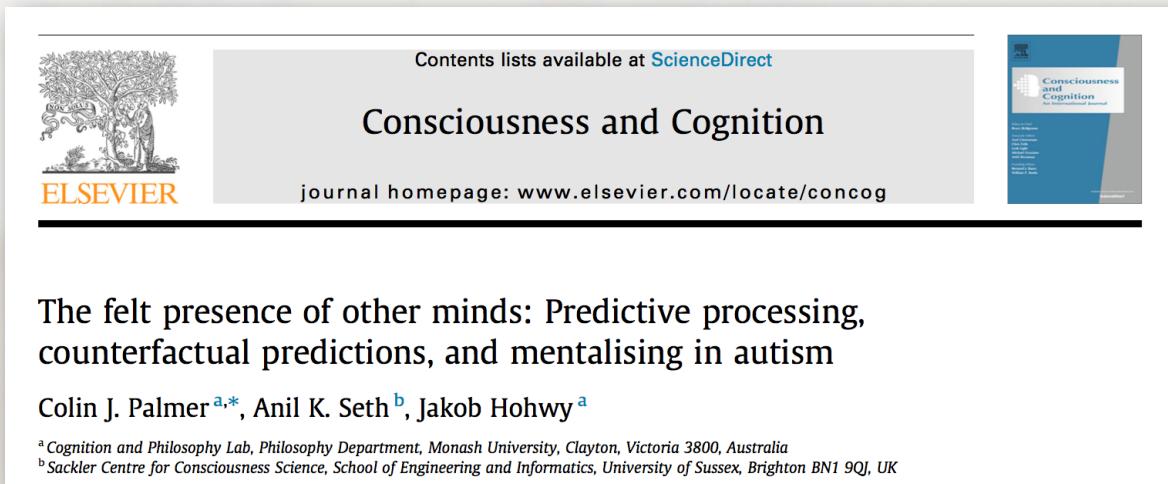
Suzuki & Seth (in progress)

## Synesthesia: an absence of presence

SYNESTHESIA  
0123456789

# ASC: another absence of presence?

- Autism spectrum disorder may involve diminished counterfactual predictions about others' mental states, leading a lack of 'felt presence of other minds').



- But, what about counterfactual richness of interoceptive inference and the 'presence' of emotional and self-related states?

**summary**

## The abductive brain

- Abduction = inference to the best explanation.
- Control-oriented abduction ~= inference to the best prediction ~= instrumental active inference.
- A fundamental cybernetic imperative: maintenance of essential variables within viability bounds.
- From this imperative flows instrumental and then epistemic active inference.
- The balance between instrumental and epistemic inference plays out differently in interoceptive and exteroceptive settings.

## Final thoughts

- Active inference in the cybernetic Bayesian brain is (abductive) **inference to the best prediction**, where the “best” predictions are those which enable control and homeostasis under a broad repertoire of perturbations.
- The experience of being a self, including ‘presence’ of selfhood, rests on counterfactually-rich predictive models geared towards (physiological) control and regulation.
- **Presence** is a potentially powerful phenomenological target for (computational) psychiatry.
- Emerging interoceptive data from ASC and psychosis adds to this picture, though with many open questions.

# Acknowledgments



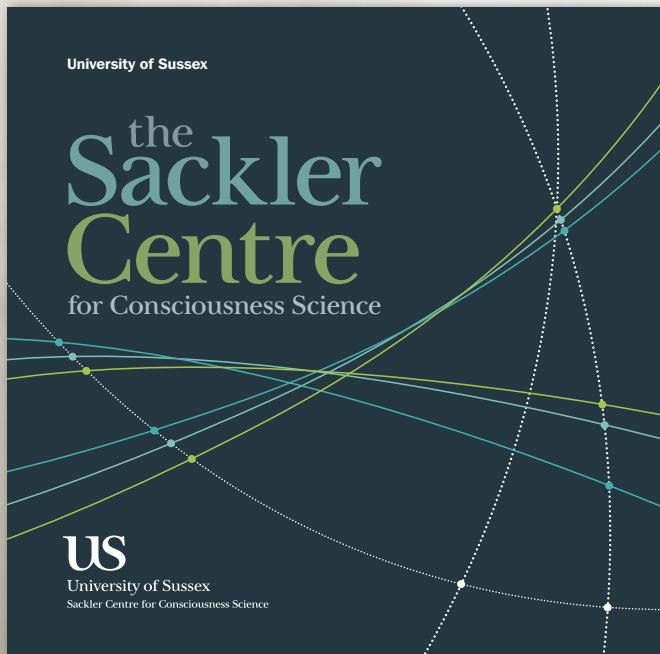
Sarah Garfinkel



Hugo Critchley



Keisuke Suzuki



Funded Ph.D. positions now available! Visit  
[www.sussex.ac.uk/sackler](http://www.sussex.ac.uk/sackler)

Funding:  
Dr. Mortimer and Theresa Sackler Foundation  
EU FP7 project CEEDS (ICT-2009-5-258)  
ERC Advanced Investigator (Critchley/Garfinkel)