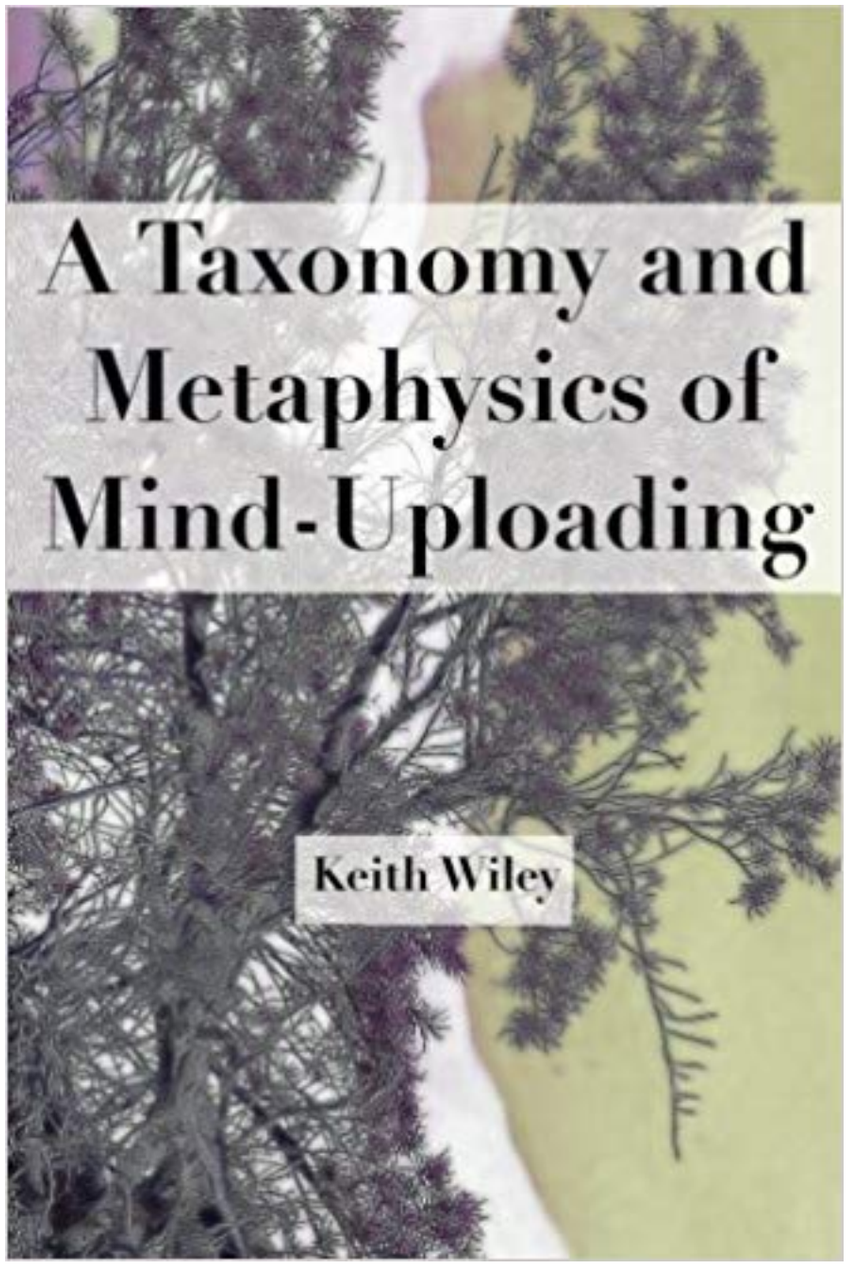
Mind uploading

Sushrut Thorat

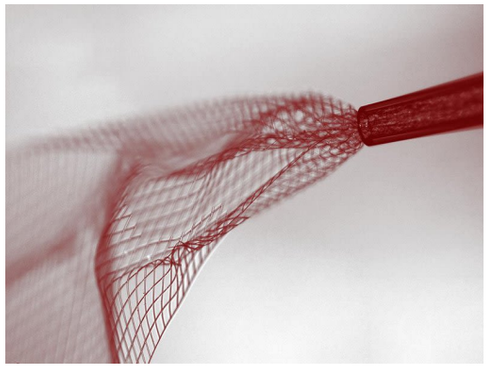


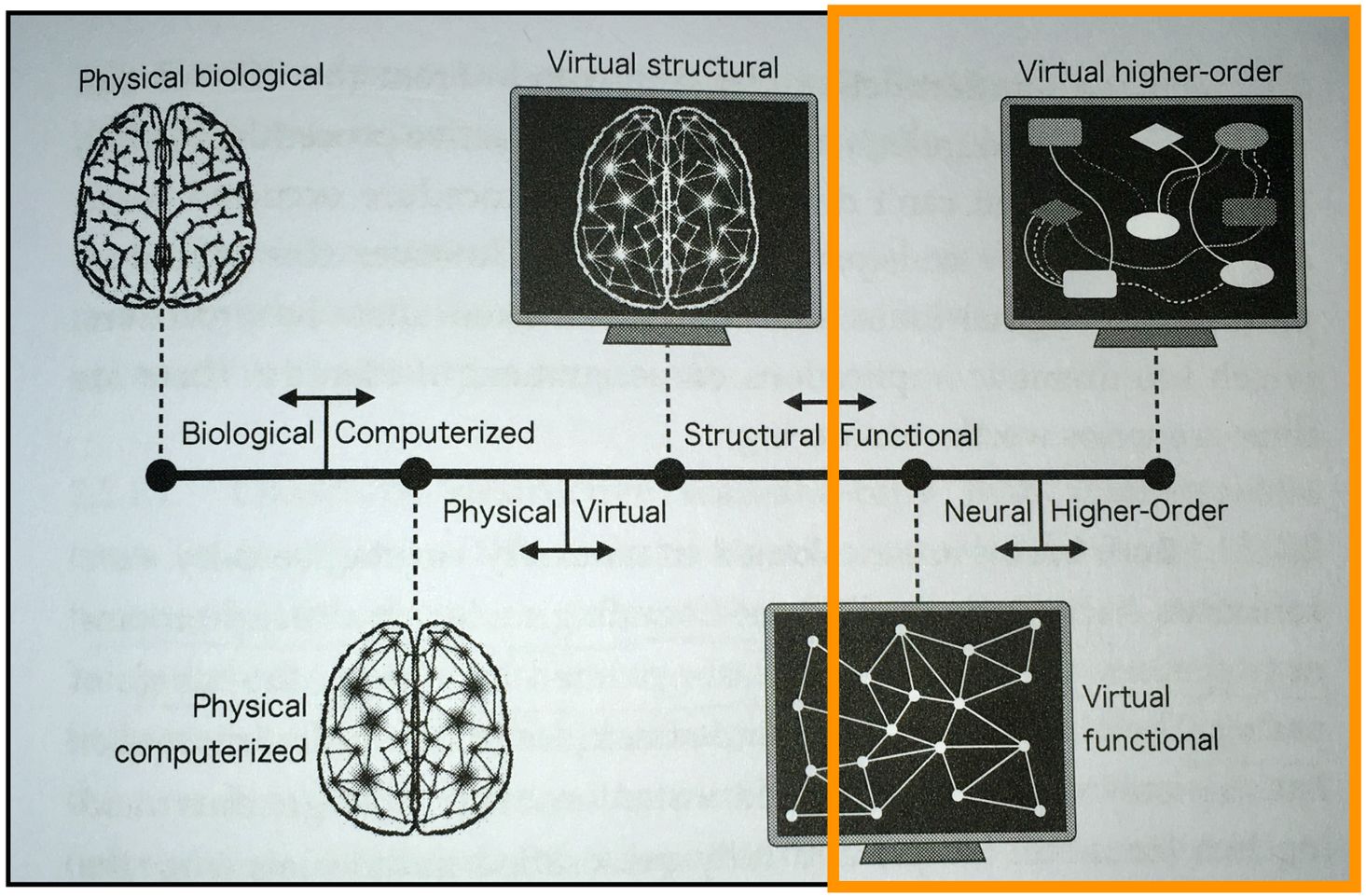
We are all beginners to this discussion.

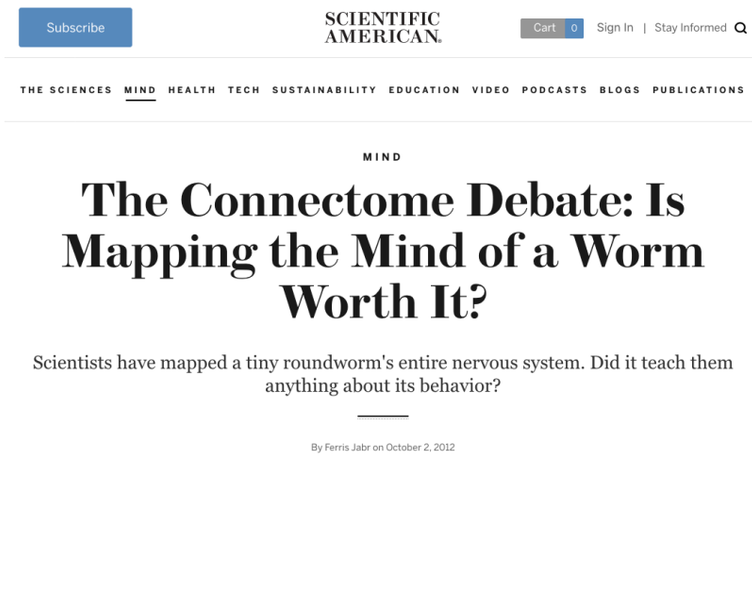
Some “mind uploading” scenarios

More brain copying/manipulation

1. Destructive gradual replacement
    1. e.g. neural lace based
2. Scan & duplicate
    1. Frozen scan & duplicate
        1. Does the connectome have sufficient information to simulate a brain?
            1. Can we “kickstart” a brain given the connectome?
    2. Else scan at what level?







Subscribe SCIENTIFIC AMERICAN. Cart 0 Sign In | Stay Informed

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS PUBLICATIONS

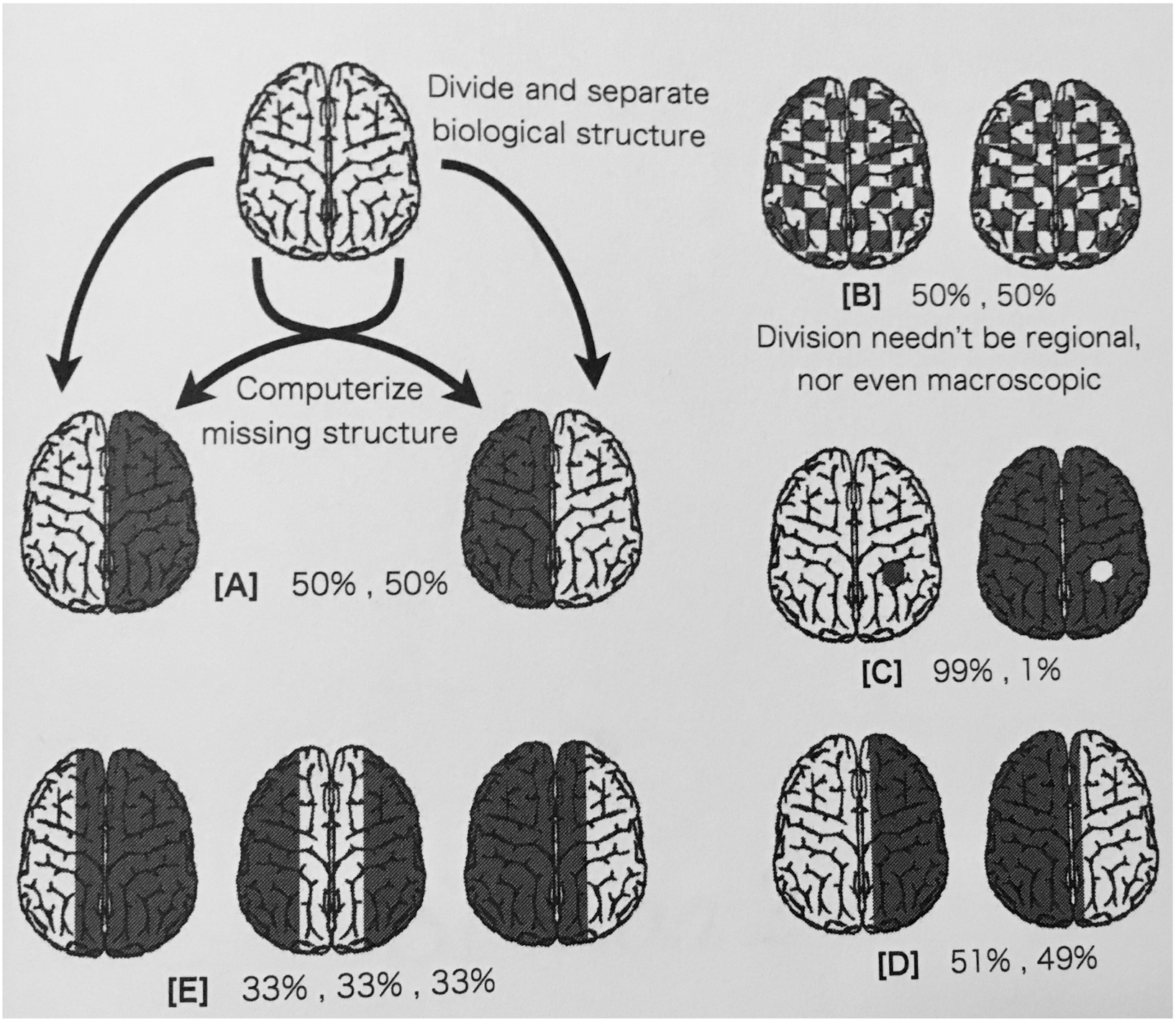
MIND

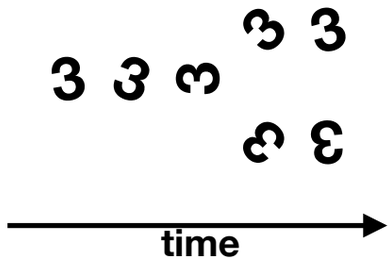
The Connectome Debate: Is Mapping the Mind of a Worm Worth It?

Scientists have mapped a tiny roundworm's entire nervous system. Did it teach them anything about its behavior?

By Ferris Jabr on October 2, 2012

“Identity” ownership issue - Brain division





Which is the original 3?

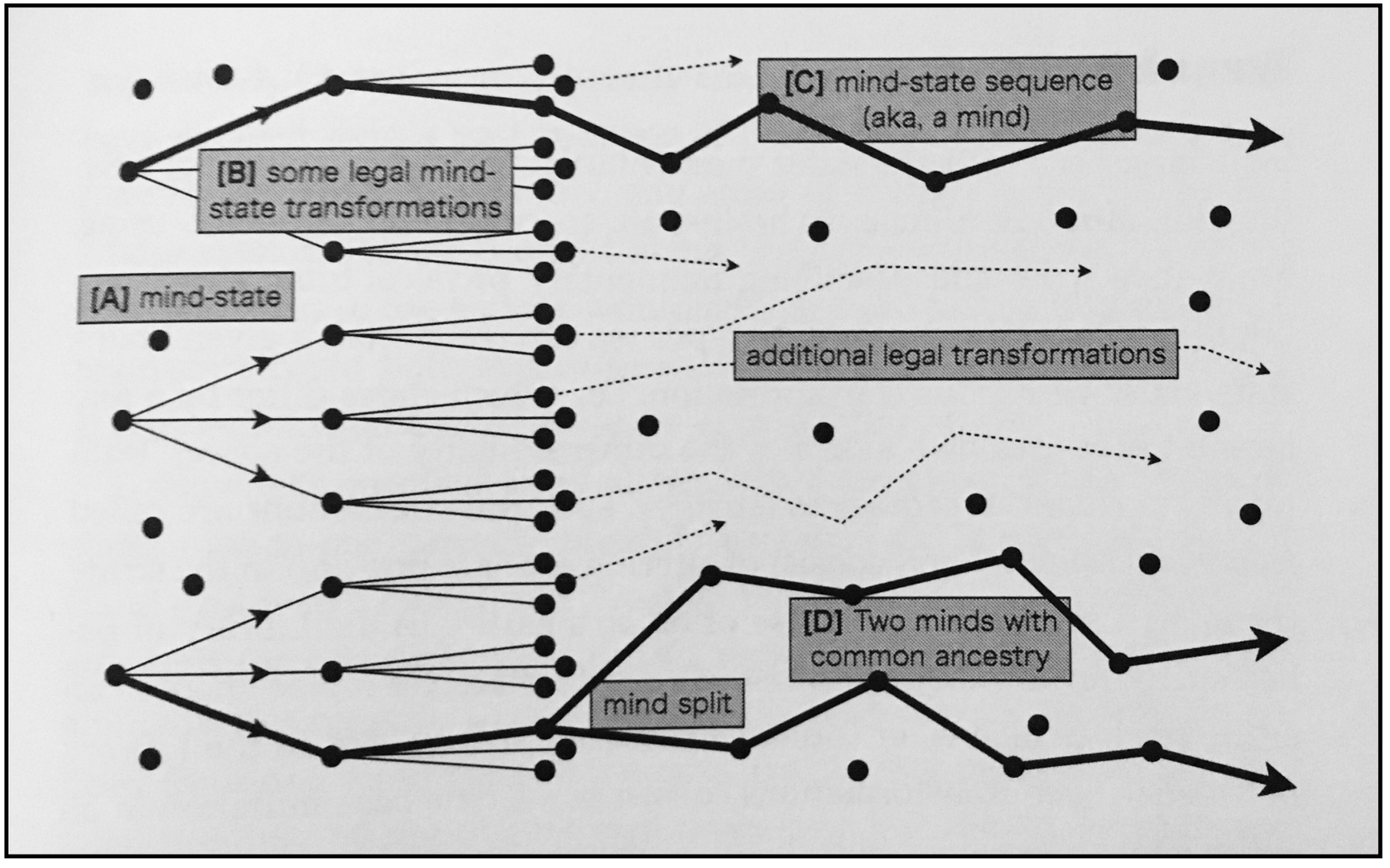
Ownership assignment is non-trivial!

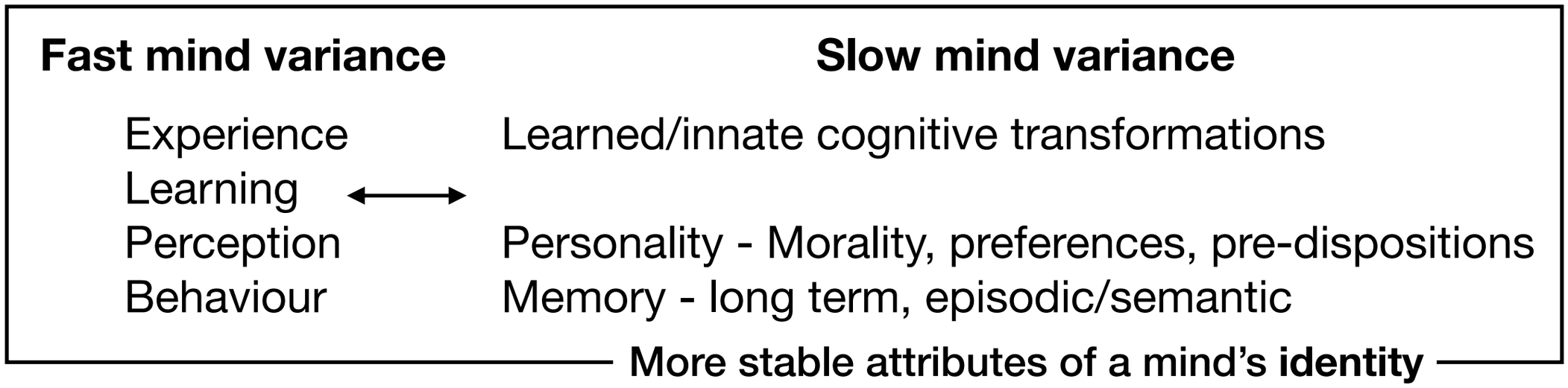
What is a mind in this formalism?

- "Dynamic processes cannot be meaningfully considered in transitory static states" - snapshot of "my parrot in flight" - cannot shy away from dynamics
- A mind is a sequence of mind-states
- A mind-state represents the totality of mental phenomena at a moment in time -> life experiences & memory, and cognitive states (*to be explored*)
- Mind states are isomorphic with brain states (neural activations at a point in time or a window of time)

- Epsilon transformations
  - E.g. bit flips in a string, APs in brain state space
- Two mind states cannot map onto the same mind state
  - Memory issues
- "Mind approximations"

Mind-state space (The Library of Anaxagoras)





Levels of approximating a mind in a system:

- Cognition ~ 1, Parameters ~ 0 : clean slate AGI, Importance ~ 0
- Cognition = 0, Parameters >> 0 : (Auto-)Biography, Importance >> 0
- Cognition ~ 1, Parameters ~ 1 : Good "mind upload", Importance ~ 1
- **Cognition >> 0, Parameters >> 0, Importance >> 0 : ?**

Example approximations of a mind:

1. A CNN for categorizing objects which makes errors like a given human and has similar representations in FC7 as in LO/VTC of that human Style transfer
    1. Cognition ~ 0, Parameters > 0, Importance ~ 0
2. A chatbot which can communicate like a given human
    1. Cognition > 0, Parameters >> 0, Importance >> 0
    2. Req - Conversational priors, manipulable "identity latent space", **TBD?**

- Cognition >> 0 will require much more work in both AI and neuroscience
- Parameters ~ 1 is partially dependent on high cognitive abilities

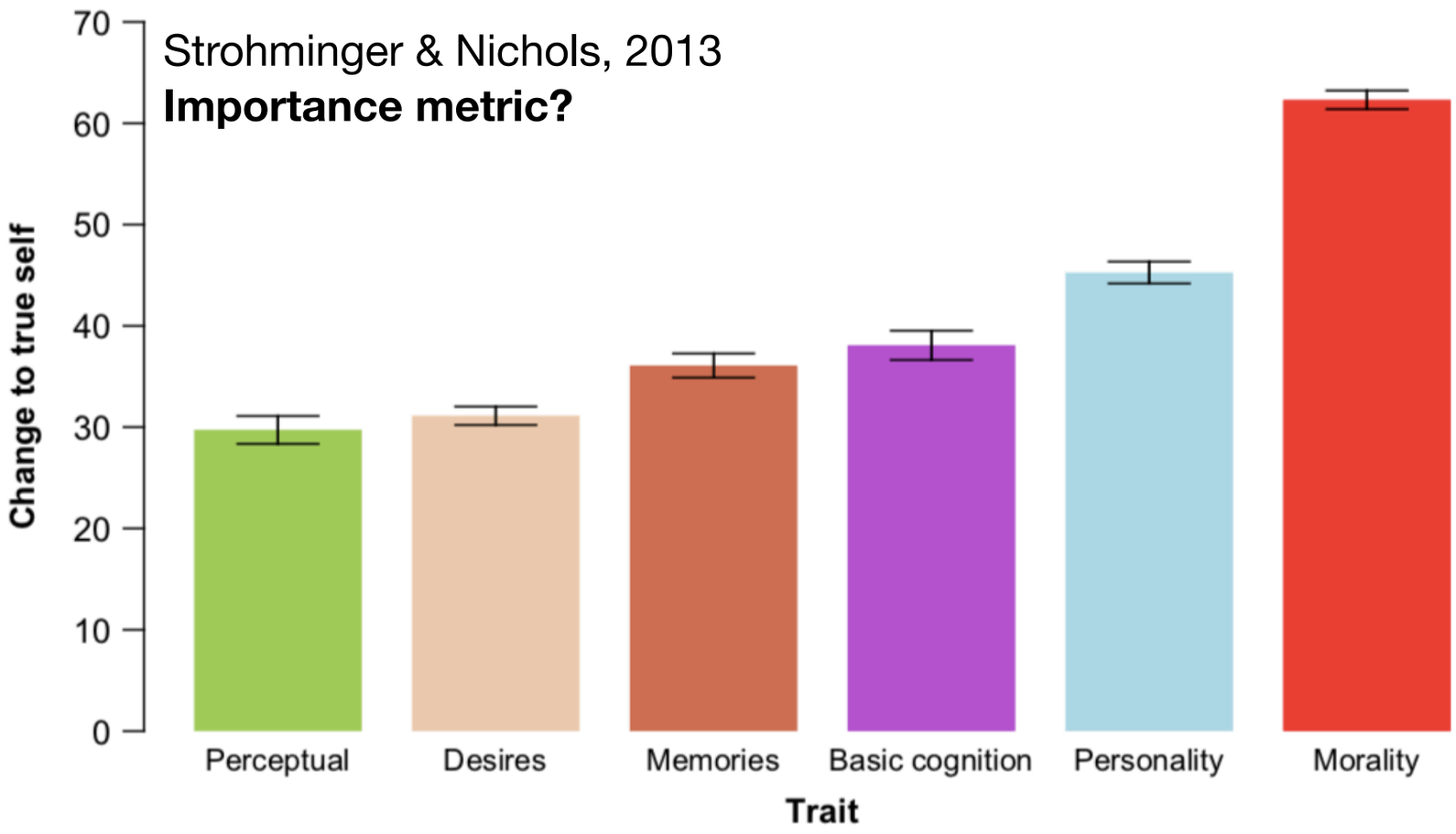
Gauging the Importance of an approximation

- Turing test for Human-ness —> Turing test for Identity
- How do we model a mind (and the self)?
  - Theory of Mind
    - Modeling input-output functions of a mind given priors from modeling the self with the same machinery? (Consciousness and the social brain - Graziano)
    - Would self-reports be trustworthy in gauging the identity of the self?
      - Assuming so,

How can we formalize such an Importance metric?

Embodiment might help

(e.g. instantiation in a low-cognition substrate - chatbot)



Going all the way…

- General idea: crude dissociation b/w Cognition and Parameters
- Higher the cognitive abilities, more the possible expressivity of parameters?
  - Start thinking about training artificial agents in open world scenarios to attain the level of human cognition
    - Task for both ML and neuroscience - cannot shy away from the complexities of cognition!
- On the other hand,
  - Start building systems with low cognition but high approximation in parameters
    - Understand character creation in literature
    - Build chatbots which can take on identities and improve them iteratively

Summary

- Scenarios of uploading minds
- Identity ownership issues
- What is a mind?
- Approximations to a mind
- What will it take to get to full-fledged mind uploading?

Any suggestions?