Integrated Information Theory: Some Philosophical Issues

David Chalmers

Two Issues

- What is consciousness, according to IIT?
  - Are the axioms/postulates correct and complete?
- What is the metaphysical status of IIT?
  - materialism, dualism, idealism, panpsychism, Russellian monism?

Axioms

- Existence: Consciousness exists (intrinsically).
- Composition: Consciousness is structured (each experience is composed of many elements).
- Information: Consciousness is differentiated (each experience differentiated from other possible exps)
- Integration: Consciousness is integrated (each experience is irreducible to components)
- Exclusion: Consciousness is definite (each experience excludes other experiences)

Correctness

- Are these axioms correct?
- I think: appropriately interpreted, each of them is correct.
- I’m committed to versions of the first three in work on the double-aspect theory of information (1996) and versions of the last two in work on unity of consciousness (2003).

Completeness

- Do these axioms completely characterize consciousness?
- Arguably not. Further potential axioms:

Missing Axioms?

- Consciousness has phenomenal character (there's something it is like to have it)
- Consciousness is knowable/introspectible/ accessible (available to a subject)
- Consciousness is representational (it presents a world)
- Consciousness is temporal?
- Consciousness has subjective character and qualitative character?

Using Missing Axioms

- Worry behind e.g. expander case: prima facie, lots of systems are composed/differentiated/ integrated/etc without being conscious.
- Appropriately developing further axioms might exclude these cases.

From Axioms to Postulates

- Tononi translates his five axioms (governing phenomenology) to five postulates (governing physical basis of consciousness).
- Lots of slack in between! (Hard to just read postulates off axioms.)

Translation Manual

- Basic hypothesis: physical locus of consciousness is cause-effect structure (CES).
- Specifically: a certain sort of cause-effect structure (conscious CES) specified by the axioms.
- So translate consciousness -> conscious cause-effect structure throughout?

Postulates

- Existence: Conscious CES exists (intrinsically).
- Composition: Conscious CES is structured (each CES composed of many CESs).
- Information: Conscious CES is differentiated (each CES differentiated from other possible CESs)
- Integration: Conscious CES is integrated (each CES is irreducible to component CESs)
- Exclusion: Conscious CES is definite (each CES excludes other CESs)

Differences between Postulates

- The first three seem trivially true of all CES.
- The last two are nontrivial and serve to distinguish the consciousness-relevant CESs from others.
- Why the difference?

Translation to Mathematics

- Existence, composition, information: construe CES and cause-effect graphs
- Integration says system CES is not reducible to parts' CES. I.e. any two parts have holistic effects. Phi > 0.

Exclusion and Phi-Max

- Exclusion says system CES has maximal phi (higher than any of its parts).
  - Q1: Where does “maximal” come from (axioms don’t say anything about consciousness being maximally integrated!).
  - Q2: Exclusion axiom says experience of subject excludes alternative experiences by that subject — not by other part-subjects.

Identity Postulate

- Amount of consciousness = phi
- Specific state of consciousness (quale) = Maximal irreducible cause-effect structure

- What is the status of this identity? It looks a little like a classical mental-physical identity (e.g. pain = C-fiber firing) and subject to similar worries.

Deriving Experience

- Can we derive the phenomenal character of an experience from a MICS?
- It looks nontrivial to derive even the structural character of an experience from a MICS, let alone its specific qualitative character?
- Does one need some further specific bridging principles to cross the bridge from MICS to experience?

Conceivability

- Prima facie, one can conceive any MICS without an associated experience (e.g. in a zombie system)
- Also any experience without an associated MICS (e.g. in a ghost system?).
- If so, hard to see how exp = MICS can be an *identity*.
- Alternative: exp = MICS is a *psychophysical law.*

Intermediate View

- Giulio says: one can't derive experience from the physical — but maybe one can derive physical properties from experience?
- Suggests: zombies are conceivable, but ghosts are not conceivable. Experience necessarily has certain causal powers.
- Highly reminiscent of the phenomenal powers view (Hedda, Schopenhauer).
- Consistent with panpsychism, Russellian monism.

Metaphysics of IIT

- What is the metaphysical status of IIT?
- Is it materialism, dualism, etc?

Type-A Materialist IIT

- Type-A Materialism: functioning is all that needs explaining (Dennett, etc)
- Type-A IIT: Consciousness is wholly explainable in terms of the dynamics of information integration
    - Explain integration dynamically, nothing else needs explaining, zombies inconceivable.
- Doesn’t seem to be Tononi’s view.

Type-B Materialist IIT

- Type-B Materialism: consciousness ontologically reudible to physical, though not conceptually/ epistemically reducible (Block, Balog, etc).

- Type-B IIT: Consciousness is identical to and reducible to integrated information

    - a primitive theoretical identity, as with classic mind-brain identity theory?

Dualist IIT

- Property dualism: physical properties and consciousness are distinct but lawfully linked (by fundamental laws).

- Dualist IIT: Integrated information and consciousness are distinct but linked. IIT as fundamental psychophysical law.

Interactionist and Epiphenomenalist IIT

- Q for Dualist IIT: Is physical/informational dynamics closed? Does consciousness play a role?

- Interactionist IIT: Consciousness plays causal role in physical dynamics. Phi-induced consciousness collapses the wave function? (Chalmers/McQueen, Kremnizer/Ranchin)

- Epiphenomenalist IIT: Info dynamics closed, consciousness plays no causal role.

Panpsychism

- (Russellian) Panpsychism: Consciousness serves as the intrinsic nature of physical processes.
- (Russellian) Panprotopsychism: Protoconsciousness serves as the intrinsic nature of physical processes.
- Constitutive Pan(proto)psychism: these micro experiences combine to form our macroexperiences.

Panpsychist IIT

- Russellian IIT: Consciousness is the intrinsic nature of integrated information (see Morch, Grasso).

- Problem for IIT: integrated information looks like a nonfundamental quantity. How can its intrinsic nature play a causal role in fundamental physics?

- Maybe its intrinsic nature only plays a causal role in nonfundamental physics. But how can we derive this role from microphysical causal roles that don't involve consciousness?

Is Integrated Information Fundamental?

- Giulio sometimes says: integrated information is fundamental.

- But: phi is defined by an equation in physical terms, supervening on fundamental physics. So physics and II can't both be fundamental.

- Is II fundamental and physics derivative?

- Or is consciousness and its powers fundamental and physics derivative? (Idealism!)

Combination Problem

- Combination problem for pan(proto)psychism: How do the (proto)microexperiences add up to larger macro experiences?

- For IIT: How do zombie microentities add up to conscious macroentities?

Mental Causation

- How does consciousness play a causal role in IIT (given dualism, panpsychism, Russellian monism)?

- Worry: all physical causation can be explained by unconscious atoms, no need for consciousness to do any work,

- Tononi has unorthodox theory of causation where macro can trump micro — can this really remove exclusion problems? (See Goff, Hoel, Marshall),