

Taking Monism Seriously

Michael Edward Johnson^a

Qualia Research Institute

San Francisco, CA

MIKE@QUALIARESEARCHINSTITUTE.ORG

On weak vs strong dual-aspect monism

Sometimes, the solutions to the hardest problems are embarrassingly simple, and all that's lacking is collecting people with the right research aesthetic, and *actually trying*. I often feel this way about formalizing consciousness.

Thus far, our approach at QRI has focused on mathematizing pleasure/pain. The flagship outputs have been the Symmetry Theory of Valence [4], which is an information-theoretic definition of what pleasure is, and CDNS, which is a paradigm for applying this theory to the brain. Basically, STV says that *how pleasant an experience is is encoded in the symmetry of its mathematical representation*, and CDNS says that we can build a *great proxy* for this symmetry with a specific fMRI technique that can measure harmony in the brain. We're cautiously optimistic that we're right, and incredibly excited about the prospects of leveraging this to build better neuroscience and neurotechnology (see: Quantifying Bliss [1]; A Future for Neuroscience [3]).

But there's another frame that I think could be *at least* as generative, and is *even simpler*: taking dual-aspect monism seriously.

Dual-aspect monism?

Dual-aspect monism (*aka* 'neutral monism') essentially argues the physical and the phenomenal are ultimately different aspects of the same thing. Two sides of the same coin, or as I like to describe it, different shadows (projections) cast by the same object. The Stanford Encyclopedia of Philosophy suggests that this model originated with Ernst Mach, William James, and Bertrand Russell; today it has a small but devoted following [7]. Personally, I think that if consciousness is *something to be formalized*, as opposed to a 'God of the Gaps' frame where it's *a leaky abstraction* that will shrink to nothing as we sharpen our vocabulary, then dual-aspect monism seems like the most coherent frame (see: Against Functionalism [2]).

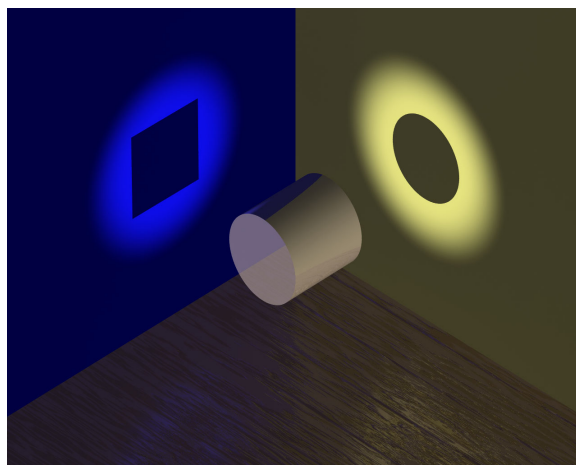


Figure 1: Monism: one source, two projections (‘shadows’)

Today, though, I want to make a further distinction: there are two versions of dual-aspect monism available to us, the weak version and the strong version. And I think an enormous amount of progress in consciousness research will happen when we start to take the strong version seriously.

Weak vs Strong monism

Weak dual-aspect monism frames its argument as a *philosophical* solution to consciousness: under this framework, there’s the physical and the phenomenal, and in some sense they’re the same, and this clarifies issues of ontology and causality in the study of consciousness. The resolution of metaphysical confusion about these issues is seen as the victory.

Strong dual-aspect monism, on the other hand, says we shouldn’t stop there. Namely, if the physical and the phenomenal *really are mathematical projections from the same object*, they’ll have *an identical deep structure*, and we can ‘port’ theories from one projection to the other. I’d offer this as the meta-theorem of monism: *every true theorem in physics will have a corresponding true theorem in phenomenology, and vice-versa*. Literally speaking—if we go through a textbook on physics and list the theorems, ultimately we’ll be able to find a corresponding truth in phenomenology for every single one. I don’t know of any ‘strong dual-aspect monists’ out there doing this—but there should be.

An analogy: much as two shadows ultimately represent the same object, but might display different features, both aspects of the monism will ultimately derive from the same structure, though each may have a different set of explicitly-represented vs hidden variables. However, as we progress on determining invariants of how the shadows behave under different conditions (alternate lighting angles, different rotations of the object, *etc.*), we’ll ultimately settle on mathematically-equivalent laws for each shadow. Essentially, anything we figure out about one shadow should likewise constrain how the other shadow will behave. And

this is far from hypothetical: we already have an extensive, elegant, and highly predictive framework for describing the dynamics of one shadow-physics!

Noether's theorem

An obvious place to start here is Noether's theorem: that for every symmetry in the mathematical structure which describes our laws of physics, there will be a conserved (invariant) quantity in our physical reality. Conservation of linear momentum, for instance, corresponds to a symmetry (invariance) in how space behaves; conservation of angular momentum to a symmetry in rotational dynamics. If Noether's theorem is true in physics, and if the *meta-theorem of monism* is true, this theorem should also apply to phenomenology and should be a gold-mine of heuristics for discovering the invariants of qualia space (as I suggested in Principia Qualia [6]).

Conservation Law	Respective Noether symmetry invariance	Number of dimensions
Conservation of mass-energy	Time invariance	1 translation about time axis
Conservation of linear momentum	Translation symmetry	3 translation about x,y,z position
Conservation of angular momentum	Rotation invariance	3 rotation about x,y,z axes
CPT symmetry (combining charge conjugation, parity and time reversal)	Lorentz invariance	1 + 1 + 1 (charge inversion $q \rightarrow -q$) + (position inversion $r \rightarrow -r$) + (time inversion $t \rightarrow -t$)
Conservation of electric charge	Gauge invariance	1 ⊗ 4 scalar field (1D) in 4D spacetime (x,y,z + time evolution)
Conservation of color charge	SU(3) Gauge invariance	3 r,g,b
Conservation of weak isospin	SU(2) _L Gauge invariance	1 weak charge
Conservation of probability	Probability invariance ^[1]	1 ⊗ 4 total probability always = 1 in whole x,y,z space, during time evolution

Figure 2: Table of Noether symmetries from Wikipedia

Looking for more people to actually try

We see a lot of these threads, and don't have the resources to pull them all. I'd encourage physicists interested in phenomenology to literally open up their favorite physics textbook, start writing down the theorems inside, and strategize about how each could be 'ported' to phenomenology. (Arguably, the same could apply to Chemistry too—see, *e.g.*, QRI's work on neural annealing [5]—although I doubt the meta-theorem of monism is as true as in physics.)

And if you find something interesting, especially something that leads to better predictions, we'd love to hear about it.

References

- [1] A. G. Emilsson. Quantifying Bliss: Talk Summary, June 2017. URL <https://qualia.computing.com/2017/06/18/quantifying-bliss-talk-summary/>.
- [2] M. Johnson. Against functionalism: why I think the Foundational Research Institute should rethink its approach, . URL <https://opentheory.net/2017/07/why-i-think-the-foundational-research-institute-should-rethink-its-approach/>.
- [3] M. Johnson. A Future for Neuroscience, . URL <https://opentheory.net/2018/08/a-future-for-neuroscience/#>.
- [4] M. Johnson. Heuristics For Interpreting The Output Of Formal Panpsychist Theories Of Consciousness, . URL <https://www.youtube.com/watch?v=CzUf0QIIxxA>.
- [5] M. Johnson. Neural Annealing: Toward a Neural Theory of Everything, . URL <https://opentheory.net/2019/11/neural-annealing-toward-a-neural-theory-of-everything/>.
- [6] M. Johnson. Principia Qualia, . URL <https://opentheory.net/2016/11/principia-qualia/>.
- [7] H. H. Mørch. Is Matter Conscious?, Apr. 2017. URL <http://nautil.us/issue/47/consciousness/is-matter-conscious>.