The Meta-Dynamic Nature of Consciousness

John Barnden
(Emeritus Prof. of AI)

School of Computer Science
University of Birmingham
UK

at

Models of Consciousness
University of Oxford

9–12 September 2019

Nature of Talk

- Physicalist proposal, albeit involving extended physics.
  Radical, but analogous to some other suggestions made at this conference.

- Outline of a philosophical argument that phenomenal consciousness (subjective experience) is (at least partly) constituted of (some particular form of) *physical meta-causation*.

  [My approach: Barnden 2014, 2018a,b, 2019.]
  [For (rare) discussion of related meta-causation notions: Ehring 2009, Koons 1998, Kovacs 2019.]
  CAUTION: There are other meanings for "meta-causation."

- Take "causation" = fundamental physical dynamism.
  Meta-causation = *meta-dynamism*.

- Provisional, small steps towards formalization of meta-dynamism and meta-dynamic physical laws.

"Meta-Causation" In General

Whatever your notion/theory of "causation" in general,
meta-causation is when a causing [instance of causation]
itself, directly causally influences something ((and/or is causally influenced)).

Cases and Heuristic examples:

Left-handed 1: "<Donald's causing democracy to collapse> *caused* Vlad to like him."

Left-handed 2: "<The way Donald caused democracy to collapse> *caused* Dom to like him."

((Tower of meta:)) There can be meta-...-meta-causation up to any level, in principle.

((Not addressed (much) in this talk:))

Right-handed 1: "Petulance *caused* <Donald to cause democracy to collapse>."

Right-handed 2: "Greed *affected the way* <Donald caused democracy to collapse>."

Right-handed 3: "Non-reelection *prevented* <Donald causing democracy to collapse>."

Ambidextrous: a causing causes/modifies/inhibits a causing.

Outline of Philosophical Argument for Consciousness requiring Meta-Causation

- Consciousness is a property of physical processes:
  "genuine" ones (properly causally-linked, internally).

- A minimal reflexivity assumption, about a conscious process needing to be sensitive to itself.

  [Related to the popular but contentious *pre-reflective self-consciousness*: see e.g., Gallagher & Zahavi 2015, Guillot 2017, Kriegel 2009, Sebastián 2012, Williford 2015, Zahavi 2005.]

- Then, the process must objectively pick itself out *as a particular causal process*.

- A good (only?) way of achieving this is for the process's internal causation
  1. not to be reducible to the process's trajectory of states and
  2. to have direct, i.e. meta-causal, influence on states in the process.

Additional Conjectures

- Some suitable form and arrangement of meta-causation is sufficient for, indeed constitutes, the phenomenal core of consciousness.
- In isolation it constitutes something like contentless conscious awareness, or a basic sense of continuing existence, or a basic sense of (fat) Now-ness [cf. Yakov Kremnitzer's and Michael Silberstein's talks].
- This arrangement is an episode of meta-causation meta-causally affecting itself; the whole episode is just that reflexive meta-causation.

A Little More Detail of the Necessity Argument

Starting Assumption: A Type of Reflexivity:

Every [conscious] experience continuously, constitutively involves
PRE-REFLECTIVE AUTO-SENSITIVITY (PRAS).

- That is, the genuine [i.e., causally-unfolding] physical processing constituting the experiencing is *pre-reflectively* sensitive to that experiencing/processing itself [where this sensitivity is not *assumed* to amount to self-consciousness].

Problems achieving adequate PRAS:

- Problems with having a state within the process REPRESENTATIONALLY identify its recent causal history as being its own recent causal history.
- Problems with possible suggestion that the sensitivity to own past history is fully accounted for by causal reaction to, merely, own past STATES as opposed to their causal linking. [Thought-experiment in Barnden (2018a,b).]

Meta-causation provides a good way out:

- At every moment throughout a conscious process, the causing within some sub-interval of the process up to that moment META-CAUSES SOME "SUITABLE" EFFECT (TBD) on the state/state-trajectory/causation at/from that moment.

A Radical Anti-Humeanism

The universe's regularities are not there just by brute fact—it *must* unfold as it does. It is not just its particles, forces, momenta, curvatures, fields, etc., just happening to interact in a regular way.

The extra is the dynamism.

> It is the *lawful pushing-forward (ongoing lawful necessitating)* of the universe.
>
> [Dynamism as "oomph" — Demarest 2017, Kutach 2014, Schaffer 2016. I am influenced by Maudlin 2007, though I stress lawful pushing-forward rather than laws themselves as the extra, non-Humean ingredient.]

A spatiotemporally localized instance ("chunk") of dynamism is *NOT* just a *TRAJECTORY* of localized kinematic states over time:
it's the lawful pushing-forward responsible for providing that trajectory.

The dynamism is a "first-class citizen" of the physical universe:
((a fully physical aspect of it, not just some metaphysical addendum or ground)).

For me: "causation" = dynamism. Underlies any dynamic effect, such as from entanglement, quantum collapse, etc., even if not usually regarded as "causal."

> [So my causation is not as usually discussed in philosophy of causation (causal links between time-separated events or property instances or facts or ....).]

Meta-Dynamism as Meta-Causation, Possibly Going Beyond Consciousness

Dynamism Instances, as *bona fide* physical entities, can dynamically affect ((and be affected by)) other physical entities. Such affecting is meta-dynamism. Governed by a new, meta-dynamic physical laws that explicitly mention dynamism.

I only propose that *some* particular form and regimentation of meta-dynamism, TBD, is required by—and *may* be sufficient for—consciousness.

- The universe conceivably contains other forms, possibly completely unrelated to consciousness.
- These other forms may be anything from rare to ubiquitous.

Basic Tools towards Formalization of Meta-Dynamism (in General) in a non-quantum, non-relativistic framework for now

An Assumption in my approach: Dynamism is not a coherent notion as an attribute of points of space-time, but only of (appropriate) regions of non-zero spatial and temporal extent.

AR = the set of all "appropriate" regions.

$\text{PTASSR}[\bar{x}, t] \quad = \quad \{r \in \text{AR} \,|\, \bar{x} \in \text{SpatialPointsIn}(r) \,\&\, \text{TimesIn}(r) = [., t)\}.$

(PTASSR for "Preceding Temporally-Abutting Spatially-Surrounding Regions.")

DynIs = a domain of *dynamism instances* (unanalysed for now). New physical entities, irreducible to ones in current physical theories.

$\mathcal{D}[\bar{x}, t]$: $\text{PTASSR}[\bar{x}, t] \longrightarrow$ DynIs.

Home: DynIs $\longrightarrow$ AR.

$\mathcal{B}[\bar{x}, t]$ delivers a complete "base-level" or "ordinary" state at $[\bar{x}, t]$ (electric field potential, momentum density, ...).

Basic Tools contd: Meta-Dynamic Laws

Base-level laws: Existing physical laws. They mention only base-level quantities. Their operating constitutes base-level dynamism, but this dynamism is of course not explicitly mentioned in the laws.

Meta-dynamic laws: mention dynamism itself in some way, via *dynamism quantities* of some sort. The operating of these laws constitutes *meta*-dynamism.

Current plan for general nature of dynamism quantity expressions & meta-dynamic laws:

- A *dynamism-quantity expression* is of form $\mathcal{Q}(\mathcal{B}[\bar{x}, t], \mathcal{D}[\bar{x}, t])$ for some function $\mathcal{Q}$.
- For some or all base-level laws L (assumed to apply to any give single instant t),
  L is replaced by a meta-dynamic version DynL, which is L with
  the addition of extra terms, factors, etc. that are dynamism-quantity expressions.
- E.g., suppose some base-level L is of form

$$u[\bar{x}, t] \quad = \quad v[\bar{x}, t] \; . \; w[\bar{x}, t]$$

  then DynL might be of form

$$u[\bar{x}, t] \quad = \quad v[\bar{x}, t] \; . \; w[\bar{x}, t] \; . \; \mathcal{Q}_1(\mathcal{B}[\bar{x}, t], \mathcal{D}[\bar{x}, t]) \; + \; \mathcal{Q}_2(\mathcal{B}[\bar{x}, t], \mathcal{D}[\bar{x}, t])$$

Meta-Dynamic Laws, contd

Intention: in "normal" systems that are operating, in effect, entirely according to existing physical laws, each such $\mathcal{Q}(\ldots)$ expression's value is *ineffectual*, e.g. zero if the element is an extra term or 1 if its an extra factor.

$\mathcal{B}[\ldots]$ in $Q(\mathcal{B}[\ldots], \mathcal{D}[\ldots])$ serves to allow current base-level state to help "gate" or scale the effect of Dynamism Instances in $\mathcal{D}[\ldots]$ on other quantities on the law.

Meta-Dynamic Laws, contd

The idea of adjusting each current law is just one possible route. *Perhaps* could get by with separate new laws interacting with existing laws.

> In moving to a QT framework, could have adjustment to the quantum-state evolution function analogous to those proposed by Yakov Kremnitzer, but not involving / confined to collapse.

From the point of view of the universe unfolding forward through time:

> a DynL is *left-handedly* meta-dynamic in describing the effect of past Dynamism Instances forwards in time on the current base-level state.

There may also be a need for laws that are *right-handedly meta-dynamic* in describing the effect of present state instances (and/or past Dynamism Instances) forwards in time on mentioned Dynamism Instances that are ongoing or starting now. [Ongoing work.]

> Also, might have laws that cause changes in laws (roughly cf. Adrian Kent's talk). Counts as right-handed meta-dynamism?

Meta-Dynamic Contributions to Existing Consciousness Theories?

- *Orch OR's objective reduction* [Hameroff & Penrose 2016]?
  - Orch OR proposes that objective quantum-state reduction (collapse), which can "result in moments of conscious awareness and/or choice," takes place in AVERAGE time τ inversely proportional to the gravitational self-energy E_G of the superimposed states.
  - So, room for additional, e.g. meta-dynamic, influences on the specific times taken.
  - And, the "conscious" collapses are the meta-dynamically affected ones???
- *Integrated Information Theory* [IIT, Oizumi, Albantakis & Tononi 2014]?
  - Add (suitable) meta-causation (meta-dynamism) as an extra requirement for consciousness, over and above potential requirements such as a high enough Φ value?
  - Involve meta-causation in the integrative aspect of the Φ measure?
    $\longrightarrow$ not all "integration" is appropriate for consciousness?
  - Thereby alleviate perceived IIT problems: e.g.: over-liberal attributions of consciousness; question of nesting of potentially conscious systems within others.
- *Objective quantum collapse guided by Integrated Info* [Kremnitzer & Ranchin 2015]?
  - Dynamics of collapse is controlled by measure Φ of "quantum Integrated Information" (QII).
  - Make Φ definition partly meta-dynamic using dynamism-quantity expressions?
  - Consciousness is related to collapse via QII.
    My consciousness-meta-dynamism link $\rightarrow$ could refine / solidify this role.

Concluding Summary

- THANKS FOR LISTENING!
- ANY SUGGESTIONS WELCOME, NOW OR LATER.

Auto-Sustaining Meta-Dynamism

Meta-dynamism, without any explict right-handedness, can be *auto-sustaining*:

after the operating of a DynL at time t,

the aspect of meta-dynamism consisting of that operating is now reflected in Dynamism Instances detectable by $\mathcal{Q}(\ldots, \mathcal{D}[, t'])$ for later times t',

making the resulting $Q(\ldots)$ values (in DynL or other meta-dynamic laws) effectual when they wouldn't otherwise have been.

Connection of Meta-Dynamism to Temporal Non-Locality

Meta-dynamic laws are temporally (and spatially) non-local by virtue of dealing with explicitly non-local physical quantities.

Get a form of temporal non-locality

- in a new, natural way
- from considerations (about consciousness) not themselves initially focussed on nature of physics or time
- going beyond specific previously proposed forms such as those reviewed in Adlam (2018)

Meta-dynamism is arguably a natural further step on a path that temporal non-locality starts.

Temporal Non-Locality in (Current) Physics

Adlam (2018) surveys various suggestions for temporally non-local models of physics, and presses for more attention to the notion, and points out its disruptive potential in particular to quantum theory.

Her notion of Temporal Locality:

> *Intuitively:* "[i]n a temporally local world there would be 'no action at a temporal distance', i.e., all influences on a measurement outcome would be mediated by the state of the world immediately prior to the measurement."
>
> *Her definition* (re-notated, simplifed to isolate intuitive idea):
>
> Alice and Bob perform measurements on a shared physical system. At some time t_a, Alice performs a measurement obtaining outcome A. Similarly Bob at later time t_b, outcome B. Let $W(t)$ be the whole world-state at time t. Then

$$p(A, B \mid W(t_a), W(t_b)) \quad = \quad p(A \mid W(t_a)) \quad . \quad p(B \mid W(t_b)).$$

Minimal Auto-Sensitivity

I define here: a sort of sensitivity to own dynamism that is a minimal requirement if a (genuine) process is to have PRAS of the particular sort (TBD) that constitutes consciousness.

The minimal auto-sensitivity here is NOT being claimed to be sufficient for consciousness.

(And I need a careful definition of "genuine process" in terms of Dynamism Instances)

Defn.:

- A Dynamism Instance I *is minimally auto-sensitive throughout* if for all $[\bar{x}_1, t_1]$ in its Home region, the state at t_1 (consisting of base-level state together with Dynamism Instances starting at t_1 and including $\bar{x}_1$) is affected by one or more Dynamism Instances in $\mathcal{D}[\bar{x}_1, t_1]$ that are contained Home-wise within I.

- The state at $[\bar{x}_1, t_1]$ is *affected* by a Dynamism Instance J if there is some law K applying at $[\bar{x}_1, t_1]$ that contains a $\mathcal{Q}(\mathcal{B}[\bar{x}, t], \mathcal{D}[\bar{x}, t])$ expression where:
  - this expression's value at $[\bar{x}_1, t_1]$ is effectual
  - $J \in \mathcal{D}[\bar{x}_1, t_1]$, and
  - there exists a set E of Dynamism Instances such that

    $$\mathcal{Q}(\mathcal{B}[\bar{x}_1, t_1], E) \quad \neq \quad \mathcal{Q}(\mathcal{B}[\bar{x}_1, t_1], \mathcal{D}[\bar{x}_1, t_1])$$

    and E is identical to $\mathcal{D}[\bar{x}_1, t_1]$ except that J is replaced by a different possible instance (and concomitant changes are made to instances whose Home region intersects J's).

References

Adlam, E.(2018). Spooky action at a temporal distance. *Entropy, 20*, 41 [online]; doi:10.3390/e20010041

Barnden, J.A. (2014). Running into consciousness. *J. Consciousness Studies, 21* (5–6), pp.33–56.

Barnden, J.A. (2018a). Phenomenal consciousness, meta-causation and developments concerning casual powers and time passage. Poster presented at *22nd Conference for the Association for the Scientific Study of Consciousness,* 26–29 June 2018, Kraków.

Barnden, J.A. (2018b). Consciousness, causal oomph and meta-oomph. Poster presented at *26th Conference of the European Society for Philosophy and Psychology,* 10–13 Sep. 2018, Rijeka, Croatia.

Barnden, J.A. (2019). Consciousness and meta-causation. Talk at *Joint Session of the Aristotelian Society and the Mind Association*, University of Durham, UK, 19–21 July 2019.

Beebee, H. (2009). Causation and observation. In H. Beebee, C. Hitchcock & P. Menzies (Eds), *The Oxford Handbook of Causation.* pp.471–497. Oxford: Oxford University Press.

Demarest, H. (2017). Powerful properties, powerless laws. In J.D. Jacobs (Ed.), *Causal Powers,* Chapter 4. Oxford: Oxford Scholarship Online.

Ehring, D. (2009). Causal relata. In H. Beebee, C. Hitchcock & P. Menzies (Eds), *The Oxford Handbook of Causation.* pp.387–413. Oxford: Oxford University Press.

Ellis, B. (2013). The power of agency. In R. Groff & J. Greco (Eds), *Powers and Capacities in Philosophy: The New Aristotelianism,* pp.186–206. New York and London: Routledge.

Gallagher, S. & Zahavi, D. (2015). Phenomenological approaches to self-consciousness. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = ¡http://plato.stanford.edu/archives/spr2015/entries/self-consciousness-phenomenological/¿, accessed 8 October 2015.

References, contd

Guillot, M. (2017). *I me mine:* on a confusion concerning the subjective character of experience. *Review of Philosophy and Psychology, 8*, pp.23–53.

Hameroff, S.R. (2018). Anaesthetic action and "quantum consciousness." *Anaestheisology, 129*(2), pp.228–231.

Hameroff, S.R. & Penrose, R. (2016). Consciousness in the universe: an updated review of the "Orch OR" theory. In R.R. Poznanski, J.A. Tuszynski & T.E. Feinberg (Eds), *Biophysics of Consciousness: A Foundational Approach*, pp.517–599. Singapore: World Scientific.

Koons, R.C. (1998). Teleology as higher-order causation: A situation-theoretic account. *Minds and Machines, 8*: 559-585.

Kovacs, D.M. (2019). The question of meta-causation. Talk at *FraMEPhys/Gothenburg Conference on Metaphysical Explanation in Science,* co-organized by Philosophy department, University of Birmingham. Birmingham, 10–11 Jan 2019.

Kremnizer, K. & Ranchin, A. (2015). Integrated information-induced quantum collapse. *Foundations of Physics, 45*, pp.889–899.

Kriegel, U. (2009). *Subjective consciousness: A self-representational theory.* Oxford University Press.

Kutach, Douglas (2014). *Causation.* Cambrige, UK: Polity Press.

Maudlin, Tim (2007). *The metaphysics within physics.* Oxford: Oxford University Press.

Oizumi, M., Albantakis, L. & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology 10*(5): e1003588. See also *J. Consciousness Studies, 26*(1–2), Jan/Feb2019: Special Issue entitled *Reflections on Integrated Information Theory*.

References, contd

Schaffer, J. (2016). The metaphysics of causation. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), online at http://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/

Sebastián, M.A. (2012). Experiential awareness: Do you prefer *it* to *me*? *Philosophical Topics, 40*(2), pp.155-177.

Williford, K. (2015). Representationalisms, subjective character, and self-acquaintance. In T. Metzinger & J.M. Windt (Eds), *Open MIND*: 39(T). Frankfurt am Mein: MIND Group.

Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective.* Cambridge, Mass. and London, UK: MIT Press (a Bradford Book).