

The irruption theory: How mind and matter meet along the route to consciousness

An interview with
Tom Froese

By Matthieu Koroma

Citation: Froese, T. & Koroma, M. (2024). The irruption theory: How mind and matter meet along the route to consciousness. An interview with Tom Froese by Matthieu Koroma. *ALIUS Bulletin*, XX, XX-XX

Tom Froese
tom.froese@oist.jp
OIST, Onna, Japan

Matthieu Koroma
matthieu.koroma@uliege.be
University of Liege,
Liege, Belgium

Abstract

In this interview, Tom Froese and Matthieu Koroma address the irruption theory, an account of how mental actions manifest themselves as unexplained variability in brain and bodily activity. We discuss how this approach can be applied to trace back consciousness in (neuro)physiological signatures such as scale free activity and informational complexity. We highlight the study of altered states of consciousness as an opportunity to disentangle the physiological correlates of consciousness from agency. We conclude by addressing how irruption theory provides insights into the therapeutic effects of psychedelics and the importance of crafting an appropriate context to consciously shape agency and behavior.

Keywords: *irruption theory, agency, free-will, volition, consciousness, psychedelics*

With your team, the Embodied Cognitive Science Unit at the Okinawa Institute of Science and Technology Graduate University (OIST, Japan), you develop an approach of the mind drawing on theories and methodologies from philosophy, physical science, computational modeling, neuroscience and psychology. What is the importance for you of interdisciplinarity in the study of the mind and how is it enacted in the research practice of your team?

We are interested in understanding the human mind in all its complexity, which basically means starting with the recognition that people are physical, biological, psychological, social, cultural, and of course also conscious beings. The human being integrates all of these domains of description, and so our aim is to do justice to this inherent complexity. The very nature of the target phenomenon thereby forces us to find a way of doing science that can put these usually distinct disciplines into a productive dialogue with each other. In practice, this means keeping an open mind, respecting the specificities characterizing different domains of investigation, and finding a way to work with and capture the resulting diversity, and even irreducible ambiguity, without aiming to reduce them to a simple unity. Ambiguity is a positive characteristic of being human, rather than a shortcoming of the interdisciplinary method.

You recently introduced the irruption theory as a principled account of motivated activity and consciousness (Froese, 2023, 2024). Could you explain this theory and how it relates to other theories of the mind and consciousness?

The irruption theory asks the question: *how does mind matter?* How do I know that certain aspects from someone's mental and conscious activity make a difference in the material world? From the perspective of the natural sciences, the first-person point of view of consciousness is an unobservable, rather than an observable, because we cannot directly observe it when we look at the level of description of the brain and the body. Natural sciences have moved over their evolution beyond direct observables already ages ago and deal with non-observables all the time. We have lots of methods for working with unobservables, yet most of cognitive neuroscience for some reason is obsessed about finding observables, matching consciousness with neural events that are supposed to be the basis of the experience, the neural correlates of consciousness.

“ How do I know that certain aspects
from someone's mental and conscious activity
make a difference in the material world? ”

The route to consciousness is still very open but the general idea is the following: while we cannot directly measure the subjective level from brain signals in the usual scientific way, we can still have the ambition to measure how the subjective mind makes a difference in the body described as a physical system and vice-versa. Our approach is a bit like predictive processing where surprise — or prediction error — in brain signals is reflected as a difference between bottom-up and top-down processes rather than only the consequence of one or the other. In that sense, the irruption theory is realist about the physical world making a difference to consciousness and vice-versa. It is neither an identity or physicalist theory that would say that consciousness and the physical body are the same thing nor an idealist or epiphenomenalist theory that would treat the physical world and consciousness as separate. It is a “not one not two kind” of approach like Varela used to say.

“ We can measure how the subjective mind makes a difference in the body described as a physical system and vice-versa. ”

What could distinguish consciousness from other natural processes?

A kind of self-reference that is suggested by certain kinds of material configurations, like autopoietic, adaptive systems, may be the source of an internal perspective on the world from where the subjective point of view could come from. There is also the question whether the origin of these kinds of systems, the origins of life, is also the origin of everything that is qualitative in the universe. If you really want to be realist and non-reductionist about something like “goodness” in a physical system, you have to show that another, purely physical or dynamical account cannot do full justice to the phenomenon described as being good. If there is such an alternative account, then there is no need to appeal to what is good for the system to explain what is happening. And it doesn't help to posit representational content either, because that would simply presuppose what we are trying to explain.

To do better, we need a notion of what it means for a goal or value to make a difference to the system in its own right: we call that a participation criterion (Froese et al., 2023). The base phenomenon is a primitive one: a “doing” rather than a “happening” with normative conditions of success or failure that are making a difference to something. For example, if Earth fell into the Sun because of a fluke in gravity, this would not be a failure of earth’s trajectory because the planet as such is not actively regulating its distance to the sun. Irruption theory is an answer to this challenge of differentiating a “doing” from a mere “happening” (Froese, 2023). The first irruption theory paper was a little bit simplistic, as I only looked in one direction, which was characterizing intelligent behavior and agency, whereas the emergence of conscious experiences actually points in the other direction, which leaves a trace I call absorption (Froese, 2024).

“ The base phenomenon is a primitive one:
a “doing” rather than a “happening”
with normative conditions of success or failure
that are making a difference. ”

What would a signature of consciousness in the material world look like according to the irruption theory?

If we are realist about the mind and we accept a non-reductionist take, we can measure how much being a conscious agent, one engaged in volitional action, actually makes a difference in the neural signal and vice-versa. If we use an objective method to study quantitatively what happens in your subjective mind makes a difference in the world, there will be an unexplained variability in neural signals that is not reducible to the description of the physical system (Froese, 2023). The world-to-mind direction of influence, that is absorption and not irruption, would be a hidden variable, a kind of counterfactual condition, that shows as a decrease rather than increase in variability in the description of the physical system (Froese, 2024). The question is then what kind of neural signals reflect the phenomenon of becoming conscious of the world?

“ If what happens in your subjective mind makes a difference, there will be an unexplained variability in neural signals that is not reducible to the description of the physical system. ”

An example of such a signal that we are investigating is what I like to refer to as “counterfactual chirality.” Chirality refers to situations like handedness with pairs that are symmetric but not identical. If I put my right and left hand on top of each other, they do not match; they are related but not the same. Asymmetry is all over the place in all kinds of biological systems and even in neurodynamics and we are trying to develop a method to test if such symmetry and anti-symmetry in the neural phase space could be a signature for consciousness.

Investigating the dimensionality of signals could be a way of operationalizing this question. One may have a perceptual discrimination task for which people respond whether they saw an image or not. From there, you can investigate the dimensionality of the neural responses. An increase in the dimensionality of the neural signals could be related to the fact that, for example, an image consciously seen triggers memories of a person suddenly opens a whole mental world to the extent that some of that mental world shapes the neural response, which is something that cannot directly be inferred from neural signals.

Some people in my group keep pushing me to say that structured noise, like pink noise, may have long-term dependencies in the signal over time that could well fit a marker of consciousness. Pink noise is already part of the general scientific discussion of the signatures of life and mind: it's a noise signal but the noise is in itself structured temporally in a way that is not independent from what's happened in the past (Van Orden et al., 2005, 2011; Ward & Greenwood, 2007). This seems like a quality that can be studied carefully to investigate whether temporal dependencies possibly reflecting self-reference are associated with the temporal structure of consciousness.

Signatures of scale freeness should, however in my opinion, not be enough to characterize the trace of consciousness. There are many lifelike phenomena in the universe that have self-organizing properties and are adaptively robust. For example, in Japan, we have frequent earthquakes and it is easily observed that earthquakes follow a power law. Small ones are more frequent than large ones, and this relationship can be observed in the slope of the frequency spectrum of seismographs. I don't think that it is because there's consciousness at play deep in the earth, and yet, we observe long range temporal dependency so these properties may be necessary but not sufficient to talk about the efficacy of mind. We must remain careful with identifying consciousness with this and other physical properties.

Bjorn Brembs is a researcher at Freie Universität Berlin that works on free will in flies (Brembs, 2011). One of his experiments is putting a group of flies in a bidirectional box. 80% of them go in one direction while 20% go in the other direction. The 20% that went in one direction and the 80% that go in the other direction are then put in the same box and the experiment is done again. Here again, there is an equal split of 20% and 80%. His argument is that there's a kind of inner noise generator that is producing this 80-20% divide that can be a basis of spontaneous behavior generation. How could the irruption theory characterize such a link between noise generation and spontaneous behavior?

It makes a lot of sense to me to investigate the basis of voluntary processes in insects in terms of such stochastic phenomena. We could go further, even to single cell organisms for which their whole genome and biochemical pathways are mapped. Even at that level, we get surprises such as transcriptional burstiness phenomena in real time measurements of gene expression levels that is a headache for engineering of chemical productions purposes because we would rather have a very stable level of expression. Gene expression is actually stochastic and even in these systems, we can ask what is it about regulation that needs this intrinsic noise aspect to it. With these stochastic processes, there are degrees of freedom that sometimes get tapped and sometimes don't. When they get tapped, they look like stochastic events that are not predetermined.

Could we be able to capture moments when consciousness and the mind would make a significant difference in bodily signals and neural markers of consciousness? How would they then be conceived in the irruption theory framework?

I think we underestimate how intelligent our body actually is without our conscious control. Most of what we consider intelligent goal-directed behaviors is actually unconscious. I don't fall out of this chair because my body keeps me in balance and this remains outside the scope of my awareness. Slightly separating mind and body also gives the body its due. The role for conscious intervention is relatively minor here and more about context switching, for example should I shift my pose, switch conversational topics, get to the next meeting room: those transition moments are probably where most of the conscious mind is required in terms of intervening in the existing bodily trajectories.

Operationalizing these processes in terms of neural signals is hard because systems are complex and there are always multiple possible explanations to account for them. A way of thinking about the literature which resonates with some of the topics we're discussing is the readiness potential, a phenomenon uncovered by Libet describing a ramping up noise term preceding conscious voluntary decision (Libet, 1985).

I like the work in this area of Aaron Schurger on the stochastic accumulator model. The basic idea is that neural activity stochastically fluctuates and eventually crosses a threshold that triggers a decision such as moving a finger when the fluctuations are large enough (Schurger et al., 2012). The philosophical challenge is to provide a story of why that noise should be associated with agency and/or free will (Schurger et al., 2021). Irruption theory provides a broader theoretical framework giving us *a priori* reasons for looking for these kinds of signals that are already there in the data and to explore how they can relate to agency and consciousness.

Can the study of altered states of consciousness be informative about the dissociation between agentic control and consciousness and their relation to markers of consciousness?

A lot of inspiration for the irruption theory comes from what happens in the brain during psychedelic states, where there is an increase of neural information metrics such as entropy and complexity (Carhart-Harris et al., 2014; Schartner et al., 2017). The question whether this rise in neural complexity reflects awareness or increased agency is difficult to disentangle in psychedelics because a rise in awareness in psychedelics also leads to a more challenging experience and thus usually requires more cognitive effort. There's a couple of studies on people undergoing administration of anesthesia who, while slowly falling unconscious, are still able to engage in a cognitive task (Boncompagni et al., 2021). It turns out that neural complexity goes down for people falling unconscious and stopping to work on the task properly. But neural complexity is even higher than during wakeful baseline for people who hang on, despite the fact that they're almost falling unconscious. For me, it suggests that maybe irruption theories are on the right track because they are associating complexity with mental or cognitive effort rather than level of awareness *per se*.

Can this interplay between agency and consciousness as conceptualized in the irruption theory provide insight into the application of psychedelics for psychiatric diseases?

If the psychedelic state is characterized by increased neural entropy, it could explain why they're effective at treating certain kinds of conditions where agency seems to be impaired (Hipólito et al., 2023). What the psychedelics do is loosen up degrees of freedom in the system such that something like irruption can happen more easily, meaning that the intervention of the mind in the bodily basis of behavior generation is facilitated. That would be one way of thinking about that, but more work certainly needs to be done.

Note that it's not just about increasing neural entropy, because too much openness could be counterproductive in the long run. The traditional ritual way of managing these kinds of altered experiences is by creating special occasions that are social events most of the time basically preparing the body and the mind to enter that space and reintegrating them during and after the process would be very important to consider for mental health.

“ It’s not just about increasing neural entropy,
because too much openness could be
counterproductive in the long run. ”

As a concluding remark, I feel that we often forget from a first-person perspective to take into account the context when we think about our own actions. While being conscious allows us to be sensitive to many factors from the context and deciding between possible actions, a lot about our own ability to direct actions is actually driven by constraints that come from the world.

Indeed, we cannot be in charge of everything that we do and happen to us and it's okay, it is just the nature of things, and we shouldn't beat ourselves up if we sometimes do the wrong things. We can't directly be in charge of everything happening in the “here and now” and which affordances of the environment we may follow. There is an irreducible gap between our intention and our behavior; we can open a space of possibilities, but the precise action that emerges is self-organized on the basis of a history of organism-environment interaction. Where our responsibility lies is to create the right conditions along the path: that is a long-term project; it's about the cultivation of habits, about skills, about selecting the right environments, because that's really where the creation of the conditions of behavior generation takes place.

“ Responsibility is a long-term project;
it's about the cultivation of habits, about skills,
about selecting the right environments,
because that's really where the creation of the conditions
of behavior generation takes place. ”

References

Boncompte, G., Medel, V., Cortínez, L. I., & Ossandón, T. (2021). Brain activity complexity has a nonlinear relation to the level of propofol sedation.

British Journal of Anaesthesia, 127(2), 254–263.

<https://doi.org/10.1016/j.bja.2021.04.023>

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: Spontaneous actions and decision-making in invertebrates.

Proceedings of the Royal

Society B: Biological Sciences, 278(1707), 930–939.

<https://doi.org/10.1098/rspb.2010.2325>

Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., & Nutt, D. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, 8.

<https://doi.org/10.3389/fnhum.2014.00020>

Froese, T. (2023). Irruption Theory: A Novel Conceptualization of the Enactive Account of Motivated Activity. *Entropy*, 25(5), 748.

<https://doi.org/10.3390/e25050748>

Froese, T. (2024). Irruption and Absorption: A ‘Black-Box’ Framework for How Mind and Matter Make a Difference to Each Other. *Entropy*, 26(4), 288. <https://doi.org/10.3390/e26040288>

Froese, T., Georgii, K., & Takashi, I. (2023). Making mind matter with irruption theory: Bridging end-directedness and entropy production by satisfying the participation criterion. *Preprint*.

<https://philsci-archive.pitt.edu/22824/>

Hipólito, I., Mago, J., Rosas, F. E., & Carhart-Harris, R. (2023). Pattern breaking: A complex systems approach to psychedelic medicine.

Neuroscience of Consciousness, 2023(1), niador17.

<https://doi.org/10.1093/nc/niador17>

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8(4), 529–539.

<https://doi.org/10.1017/S0140525X00044903>

Schartner, M. M., Carhart-Harris, R. L., Barrett, A. B., Seth, A. K., & Muthukumaraswamy, S. D. (2017). Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Scientific Reports*, 7(1), 46421. <https://doi.org/10.1038/srep46421>

Schurger, A., Hu, P. “Ben,” Pak, J., & Roskies, A. L. (2021). What Is the Readiness Potential? *Trends in Cognitive Sciences*, 25(7), 558–570.

<https://doi.org/10.1016/j.tics.2021.04.001>

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42).

<https://doi.org/10.1073/pnas.1210467109>

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human Cognition and 1/f Scaling. *Journal of Experimental Psychology: General*, 134(1), 117–123.

<https://doi.org/10.1037/0096-3445.134.1.117>

Van Orden, G. C., Kloos, H., & Wallot, S. (2011). Living in the Pink. In *Philosophy of Complex Systems* (pp. 629–672). Elsevier.

<https://doi.org/10.1016/B978-0-444-52076-0.50022-5>

Ward, L., & Greenwood, P. (2007). 1/f noise. *Scholarpedia*, 2(12), 1537.

<https://doi.org/10.4249/scholarpedia.1537>