Qualia Formalism and a Symmetry Theory of Valence

Michael Edward Johnson
mike@opentheory.net

June 16, 2023

Abstract

If consciousness is amenable to formalization, symmetry within the formalism is likely to play a significant compositional, functional, and aesthetic role. We survey the current state of affective neuroscience and consciousness research, then offer a theoretical and experimental paradigm constructed around (1) Qualia Formalism and (2) the symmetry aesthetic inherited from physics and mathematics. The initial pilot of this approach is the Symmetry Theory of Valence (STV): the symmetry of an information geometry of mind corresponds with how pleasant it is to be that experience. STV is the first formal, universal, frame-invariant candidate theory of pain and pleasure.

This paper is a highly condensed and updated version of an earlier book, Principia Qualia [1]

1 We don't know what pleasure is

The biggest mystery hiding in plain sight is what gives experiences valence. Affective neuroscience has been very effective at illuminating the dynamics and correlations of positive & negative affect, but the more details we assemble the further it seems we are from confident statements about what valence 'is'.

Valence research tends to segregate into two buckets: function and anatomy. The former attempts to provide a description of how valence interacts with thought and behaviour, whereas the latter attempts to map it to the anatomy of the human brain. The following are key highlights from each 'bucket':

Valence as a functional component of thought & behaviour:
One of the most common views of valence is that it's the way the brain encodes value:

> Emotional feelings (affects) are intrinsic values that inform animals how they are faring in the quest to survive. The various positive affects indicate that animals are returning to "comfort zones" that support survival, and negative affects reflect "discomfort zones" that indicate that animals are in situations that may impair survival. They are ancestral tools for living – evolutionary memories of such importance that they were coded into the genome in rough form (as primary brain processes), which are refined by basic learning mechanisms (secondary processes) as well as by higher-order cognitions/thoughts (tertiary processes). [2]

Similarly, valence seems to be a mechanism the brain uses to determine or label salience, or phenomena worth paying attention to [3], and to drive reinforcement learning [4].

A common thread in these theories is that valence is entangled with, and perhaps caused by, an appraisal of a situation. Frijda describes this idea as the law of situated meaning: "Input some event with its particular meaning; out comes an emotion of a particular kind" [5]. Similarly, Clore et al. phrase this in terms of "The Information Principle," where "[e]motional feelings provide conscious information from unconscious appraisals of situations" [6]. Within this framework, positive valence is generally modelled as the result of an outcome being better than expected [7], or a surprising decrease in 'reward prediction errors' (RPEs) [8].

Computational affective neuroscience attempts to formalize this appraisal framework into a unified model of cognitive-emotional-behavioural dynamics. A characteristic example is "Mood as Representation of Momentum" [9], where moods (and valence states) are understood as pre-packaged behavioural and epistemic biases which can be applied to different strategies depending on what kind of 'reward prediction errors' are occurring. E.g., if things are going surprisingly well, the brain tries to take advantage of this momentum by shifting into a happier & more active state that is more suited to exploration & exploitation. On the other hand, if things are going surprisingly poorly, the brain shifts into a "hunker-down" mode which conserves resources and options.

However – while these functional descriptions are intuitive, they fail to generalize as metaphysically-satisfying answers when we look closely at edge cases and the anatomy of pain and pleasure.

Valence as a product of neurochemistry & neuroanatomy
The available neuroanatomical evidence suggests that the above functional themes highlight correlations rather than metaphysical truths, and for every functional story about the role of valence there exist counter-examples. E.g.:

Valence is not the same as value or salience

Berridge and Kringelbach [10] find that "representation [of value] and causation [of pleasure] may actually reflect somewhat separable neuropsychological functions." Relatedly, Jensen et al. [11] note that valence & salience are also handled by different, non-perfectly overlapping systems in the brain.

Valence is distinct from preferences and reinforcement learning

Even more interestingly, Berridge et al. [12] find that what we call 'reward' has three distinct elements in the brain: 'wanting', 'liking', and 'learning', and the neural systems supporting each are relatively distinct from each other. 'Wanting', a.k.a. seeking, seems strongly (though not wholly) dependent upon the mesolimbic dopamine system, whereas 'liking', or the actual subjective experience of pleasure, seems to depend upon the opioid, endocannabinoid, and GABA-benzodiazepine neurotransmitter systems, but only within the context of a handful of so-called "hedonic hotspots" (elsewhere, their presence seems to only increase 'wanting'). With the right interventions disabling each system, Berridge et al. suggest brains can exhibit any permutation of these three: 'wanting and learning without liking', 'wanting and liking without learning', and so on. Likewise with pain, we can roughly separate the sensory/discriminative component from the affective/motivational component, each of which can be modulated independently [13].

These distinctions between components are empirically significant but not necessarily theoretically crisp: Berridge and Kringelbach [10] suggest that the dopamine-mediated, novelty-activated seeking state of mind involves at least some small amount of intrinsic pleasure.

Hedonic brain circuits: a useful but loose abstraction

A strong theme in the affective neuroscience literature is that pleasure seems highly linked to certain specialized brain regions/types of circuits:

> We note the rewarding properties for all pleasures are likely to be generated by hedonic brain circuits that are distinct from the mediation of other features of the same events (e.g., sensory, cognitive). Thus pleasure is never merely a sensation or a thought, but is instead an additional hedonic gloss generated by the brain via dedicated systems. ... Analogous to scattered islands that form a single archipelago, hedonic hotspots are anatomically distributed but interact to form a functional integrated circuit. The circuit obeys control rules that are largely hierarchical and organized into brain levels. Top levels function together as a cooperative heterarchy, so that, for example, multiple unanimous 'votes' in favor from simultaneously-participating hotspots in the nucleus accumbens and ventral pallidum are required for opioid stimulation in either forebrain site to enhance 'liking' above normal. [14]

Some of these 'hedonic hotspots' are also implicated in pain, and activity in normally-hedonic regions has been shown to produce an aversive effect under certain psychological conditions, e.g., when threatened or satiated [10]. Furthermore, damage to certain regions of the brain (e.g., the ventral pallidum) in rats changes their reaction toward normally-pleasurable things to active 'disliking' [15, 16]. Moreover, certain painkillers such as acetaminophen blunt both pain and pleasure [17]. By implication, the circuits or activity patterns that cause pain and pleasure may have similarities not shared with 'hedonically neutral' circuits. However, pain does seem to be a slightly more 'distributed' phenomenon than pleasure, with fewer regions that consistently contribute.

Finally, the core circuitry implicated in emotions in general, and valence in particular, is highly evolutionarily conserved, and all existing brains seem to generate valence in similar ways: "Cross-species affective neuroscience studies confirm that primary-process emotional feelings are organized within primitive subcortical regions of the brain that are anatomically, neurochemically, and functionally homologous in all mammals that have been studied." [2] Others have indicated the opioid-mediated 'liking' reaction may be conserved across an incredibly broad range of brains, from the very complex (humans & other mammals) to the very simple (c. elegans, with 302 neurons), and all known data points in between – e.g., vertebrates, molluscs, crustaceans, and insects [18]. On the other hand, the role of dopamine may be substantially different and even behaviourally inverted (associated with negative valence and aversion) in certain invertebrates like insects [19] and octopi.

A key takeaway from the neuro-anatomical literature on valence is that we presently have no certainty around which properties are necessary or sufficient to make a given brain region a so-called "pleasure centre" or "pain centre." We can only say with confidence that some regions of the brain appear to contribute much more to valence than others.

A taxonomy of valence?

How many types of pain and pleasure are there? While neuroscience doesn't offer a crisp taxonomy, there are some apparent distinctions we can draw from physiological & phenomenological data:

- There appear to be at least three general types of physical pain, each associated with a certain profile of ion channel activation: thermal (heat, cold, capsaicin), chemical (lactic acid buildup), and mechanical (punctures, abrasions, etc.) [20].
- More speculatively, based on a dimensional analysis of psychoactive substances, there appear to be at least three general types of pleasure: 'fast' (cocaine, amphetamines), 'slow' (morphine), and 'spiritual' (LSD, Mescaline, DMT) [21].
- Mutations in the gene SCN9A (the 'master gene' for the Nav1.7 ion channel) can remove the ability to feel any pain mediated by physical nociception [22, 23] – however, it appears that this doesn't impact the ability to feel emotional pain [24].

However, these distinctions between different types of pain & pleasure appear substantially artificial:

- Hedonic pleasure, social pleasure, eudaimonic well-being, etc., all seem to be manifestations of the same underlying process. Kringelbach and Berridge [14] note: "The available evidence suggests that brain mechanisms involved in fundamental pleasures (food and sexual pleasures) overlap with those for higher-order pleasures (for example, monetary, artistic, musical, altruistic, and transcendent pleasures)." This seems to express a rough neuroscientific consensus [25], albeit with some caveats.
- Likewise in support of lumping emotional & physical valence together, common painkillers, such as acetaminophen, help with both physical and social pain [26].

A deeper exploration of the taxonomy of valence is hindered by the fact that the physiologies of pain and pleasure are frustrating inverses of each other.

- The core hurdle to understanding pleasure (in contrast to pain) is that there's no pleasure-specific circuitry analogous to nociceptors, sensors on the periphery of the nervous system which reliably cause pleasure and whose physiology we can isolate and reverse-engineer.
- The core hurdle to understanding pain (in contrast to pleasure) is that there's only weak and conflicting evidence for pain-specific circuitry analogous to hedonic hotspots, regions deep in the interior of the nervous system which seem to centrally coordinate all pain and whose physiological mechanics & dynamics we can isolate and reverse-engineer.

I.e., pain is easy to cause but hard to localize in the brain; pleasure has a more definite footprint in the brain but is much harder to generate on demand.

Philosophical confusion in valence research:

In spite of the progress affective neuroscience continues to make, our current understanding of valence and consciousness is extremely limited, and I offer that the core hurdle for affective neuroscience is philosophical confusion, not mere lack of data. I.e., perhaps our entire approach deserves to be questioned. Several critiques stand out:

Neuroimaging is a poor tool for gathering data:

Much of what we know about valence in the brain has been informed by functional imaging techniques such as fMRI and PET. But neuroscientist Martha Farah notes that these techniques depend upon a very large set of assumptions and that there's a widespread worry in neuroscience "that [functional brain] images are more researcher inventions than researcher observations" [27]. Farah notes the following flaws:

- Neuroimaging is built around indirect and imperfect proxies. Blood flow (which fMRI tracks) and metabolic rates (which PET tracks) are correlated with neural activity, but exactly how and to what extent it's correlated is unclear, and sceptics abound. Psychologist William Uttal suggests that "fMRI is as distant as the galvanic skin response or pulse rate from cognitive processes" [28].
- The elegant-looking graphics neuroimaging produces are not direct pictures of anything: rather, they involve extensive statistical guesswork and 'cleaning actions' by many layers of algorithms. This hidden inferential distance can lead to unwarranted confidence, especially since such models may systematically underestimate differences in brain anatomy.
- Neuroimaging tools bias us toward the wrong sorts of explanations. As Uttal puts it, neuroimaging encourages hypotheses "at the wrong (macroscopic) level of analysis rather than the (correct) microscopic level. ... we are doing what we can do when we cannot do what we should do". [28]

Neuroscience's methods for analysing data aren't as good as people think

There's a popular belief that if only the above data-gathering problems could be solved, neuroscience would be on firm footing. Jonas and Kording [29] attempted to test whether the field is merely data-limited (yet has good methods) in a novel way: by taking a microprocessor (where the ground truth is well-known, and unlimited amounts of arbitrary data can be gathered) and attempting to reverse-engineer it via standard neuroscientific techniques such as lesion studies, whole-processor recordings, pairwise & Granger causality, and dimensionality reduction. This should be an easier task than reverse-engineering brain function, yet when they performed this analysis, they found that "the approaches reveal interesting structure in the data but do not meaningfully describe the hierarchy of information processing in the processor. This suggests that current approaches in neuroscience may fall short of producing meaningful models of the brain." The authors conclude that we don't understand the brain as well as we think we do, and we'll need better theories and methods to get there, not just more data.

Subjective experience is hard to study objectively

Unfortunately, even if we improve our methods for understanding the brain's computational hierarchy, it will be difficult to translate this into improved knowledge of subjective mental states & properties of experience (such as valence).

In studying consciousness, we've had to rely on either crude behavioural proxies or subjective reports of what we're experiencing. These 'subjective reports of qualia' are very low-bandwidth, are of unknown reliability and likely vary in complex, hidden ways across subjects, and as Tsuchiya et al. [30] note, the methodological challenge of gathering them "has biased much of the neural correlates of

Valence is often:	... but this isn’t a perfect description, since:
How the brain represents value	It’s only a correlation, and ’value’ is a fuzzy abstraction
A result of getting what we want	‘Liking’ and ‘wanting’ are handled by different brain systems
Involved with reinforcement learning	‘Liking’ and ‘learning’ are handled by different brain systems
Proximately caused by opioids	Injection of opioids into key regions doesn’t always cause pleasure
Proximately caused by certain brain regions	Activity in these regions doesn’t always cause pleasure

Figure 1: Core takeaways of affective neuroscience on valence

consciousness (NCC) research away from consciousness and towards neural correlates of perceptual reports.” I.e., if we ask someone to press a button when they have a certain sensation, then measure their brain activity, we’ll often measure the brain activity associated with pressing buttons, rather than the activity associated with the sensation we’re interested in. We can and do attempt to control for this with the addition of ‘no-report’ paradigms, but largely they’re based on the sorts of neuroimaging paradigms critiqued above.

Affective neuroscience has confused goals

Lisa Feldman Barrett [31] goes further and suggests that studying emotions is a particularly hard task for neuroscience since most emotions are not “natural kinds,” i.e.. things whose objective existence makes it possible to discover durable facts about. Instead, Barrett notes, “the natural-kind view of emotion may be the result of an error of arbitrary aggregation. That is, our perceptual processes lead us to aggregate emotional processing into categories that do not necessarily reveal the causal structure of the emotional processing.” As such, many of the terms we use to speak about emotions have only an ad-hoc, fuzzy pseudo-existence, and this significantly undermines the ability of affective neuroscience to standardize on definitions, methods, and goals.

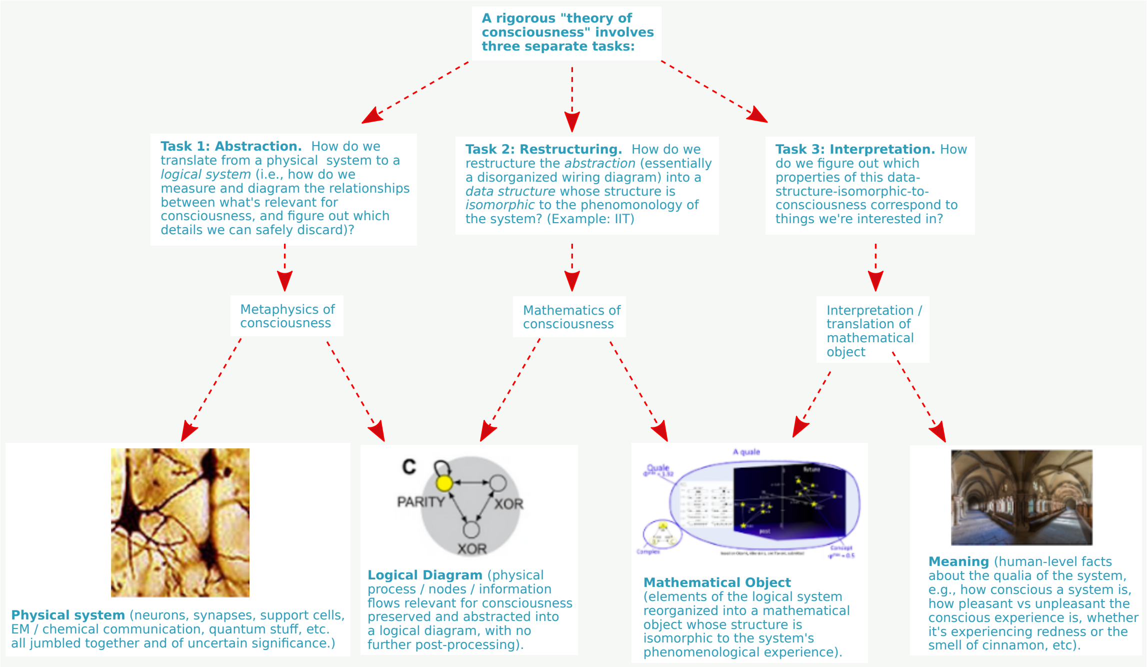
In summary, affective neuroscience suffers from (1) a lack of tools that gather unbiased and functionally-relevant data about the brain, (2) a lack of formal methods which can reconstruct what the brain’s doing and how it’s doing it, (3) epistemological problems interfacing with the subjective nature of consciousness, and (4) an ill-defined goal, as it’s unclear just what it’s attempting to reverse-engineer in the first place.

Fig 1 summarizes some core implications of current neuroscience and philosophical research. In short: valence in the human brain is a complex phenomenon which defies simple description, and affective neuroscience – though it’s been hugely useful at illuminating the shape of this complexity – is unlikely to generate any sudden or substantial breakthroughs on the topic. However, just because the methodology of affective neuroscience isn’t generating crisp insights doesn’t mean there are no crisp insights to be had. Section 2 suggests an alternate way to frame the problem.

2 An information geometry of mind

Giulio Tononi’s Integrated Information Theory (IIT) was the first theory of consciousness that offered an exact specification for what the proper goal of consciousness research should be: to generate an information geometry of mind. IIT attempts to meet this goal by equating consciousness with integrated information (the degree to which activity in each part of a system constrains activity elsewhere in the system) and outlining a formal method for calculating the topology of a system’s integration [32].

An information geometry of mind (IGM) is a mathematical representation of an experience whose internal relationships between components mirror the internal relationships between the elements of the subjective experience it represents. A correct information geometry of mind is an exact representation of an experience. More formally, an IGM is a mathematical object such that there


A rigorous "theory of consciousness" involves three separate tasks:

Task 1: Abstraction. How do we translate from a physical system to a *logical system* (i.e., how do we measure and diagram the relationships between what's relevant for consciousness, and figure out which details we can safely discard)?

Task 2: Restructuring. How do we restructure the *abstraction* (essentially a disorganized wiring diagram) into a *data structure* whose structure is *isomorphic* to the phenomonology of the system? (Example: IIT)

Task 3: Interpretation. How do we figure out which properties of this data-structure-isomorphic-to-consciousness correspond to things we're interested in?

Metaphysics of consciousness

Mathematics of consciousness

Interpretation / translation of mathematical object

Physical system (neurons, synapses, support cells, EM / chemical communication, quantum stuff, etc. all jumbled together and of uncertain significance.)

Logical Diagram (physical process / nodes / information flows relevant for consciousness preserved and abstracted into a logical diagram, with no further post-processing).

Mathematical Object (elements of the logical system reorganized into a mathematical object whose structure is isomorphic to the system's phenomenological experience).

Meaning (human-level facts about the qualia of the system, e.g., how conscious a system is, how pleasant vs unpleasant the conscious experience is, whether it's experiencing redness or the smell of cinnamon, etc).

Figure 2: Factorizing the problem of consciousness

exists an isomorphism (a one-to-one and onto mapping) between this mathematical object and the experience it represents. Any question about the contents, texture, or affective state of an experience can be answered in terms of this geometry.

Many have taken issue with the mathematical formulation of IIT [33], and others have attempted to reformulate IIT into an explicitly physical theory [34]. But I suggest these critiques are almost a distraction from the clarity that comes from focusing on IGMs; by doing so, we can split "the problem of consciousness" into three pieces: choosing what should contribute to consciousness, creating an information geometry from these components, and interpreting the resulting geometry [1].

Nothing about this process is dependent upon the particular math of IIT, nor its conception of integration; Dalton Sakthivadivel, for instance, assembles a candidate information geometry of mind from the mathematics of Markov blankets [35]. The specification of a geometry allows a modular, "divide and conquer" approach to consciousness, though to date methods of constructing information geometries are sufficiently new, technical, and speculative that interpretation tends to be linked to a specific geometry, e.g., [32].

A fork in the road

There are essentially two approaches we can take toward consciousness. The first is that consciousness is real, with 'real' meaning that for any conscious system and conscious experience, there exists an exact information geometry of mind. If consciousness is real, the purpose of consciousness research is to find this geometry and figure out how to interpret it. With a nod to Tononi, I've called this view Qualia Formalism [1]. Alternatively, if consciousness is not real and is more like élan vital, i.e. a leaky reification, the quest for an information geometry of mind is futile, and our proper Wittgensteinian goal is to diagnose and dissolve our linguistic confusion and move on to other topics. This view is commonly referred to as analytic functionalism; see, e.g., [36]. I believe that (a)

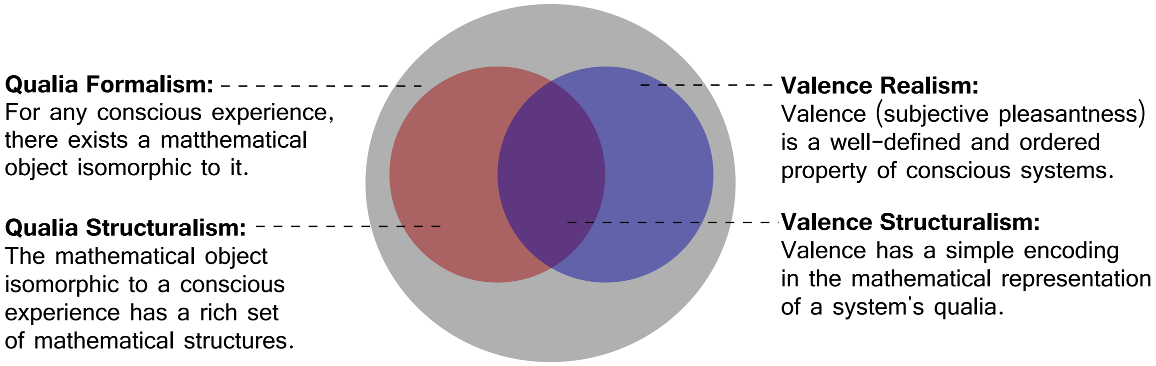

Qualia Formalism:
For any conscious experience, there exists a matthematical object isomorphic to it.
Qualia Structuralism:
The mathematical object isomorphic to a conscious experience has a rich set of mathematical structures.
Valence Realism:
Valence (subjective pleasantness) is a well-defined and ordered property of conscious systems.
Valence Structuralism:
Valence has a simple encoding in the mathematical representation of a system's qualia.

Figure 3: Requirements for an elegant formalism for valence

there is little possibility for a middle ground between these two positions, (b) either could be true, and (c) differences between these positions will cash out empirically. Specifically, the assumption of there being formal structure to consciousness should generate surprising predictions and elegant compressions, much like how the assumption that there is formal structure to electromagnetism allowed Faraday and Maxwell to make surprising predictions. An absence of successful predictions would be weak evidence against the existence of an information geometry of mind [37].

The remainder of this work assumes that a correct information geometry of mind exists but is entirely agnostic about the particular approach used to generate it, instead focusing on a specific subtopic within the interpretation problem: what element, pattern, or property of this geometry corresponds to the pleasantness of the represented experience?

3 The symmetry aesthetic

Objects are mathematically symmetrical insofar as they are unchanged with respect to some transformation; e.g., a square respects eight symmetries (four rotations, two mirror images, and two diagonal flips). Any mathematical object which we can apply operations upon can be evaluated for symmetry, and mathematicians quantify symmetry in terms of the size of the object's symmetry group: how many different operations leave the object unchanged.

The importance of symmetry has long been recognized across disparate fields spanning Ancient Greek philosophy and aesthetics [38], the sacred geometry of most major religions [39, 40], classical and renaissance architecture [41, 42, 43], musical theory [44], intuitive estimations of facial beauty [45] and medical guidelines for plastic surgery [46], with various attempts to formalize a general aesthetic measure based on symmetry [47, 48, 49].

However, one of the most unsung stories is how symmetry has come to absolutely dominate our conceptions of both mathematics [50] and physics [51, 52].

Physics and the symmetry aesthetic

Emmy Noether [53, 54] is most well-known for Noether's theorem, which states that "every differentiable symmetry of the action of a physical system has a corresponding conservation law." In other words, when there's an invariance (in the form of a symmetry) in the equations describing a system, there always exists a corresponding invariance (in the form of a conservation law) in the behaviour of that system, and vice-versa. As physicist Max Tegmark notes:

> German mathematician Emmy Noether proved in 1915 that each continuous symmetry of our mathematical structure leads to a so-called conservation law of physics, whereby some quantity is guaranteed to stay constant [...] All the conserved quantities that we discussed in Chapter 7 correspond to such symmetries: for example, energy corresponds to time-translation symmetry (that our laws of physics stay the same for all time), momentum corresponds to space-translation symmetry (that the laws are the same everywhere), angular momentum corresponds to rotation symmetry (that empty space has no special "up" direction) and electric charge corresponds to a certain symmetry of quantum mechanics. [55]

Noether's theorem sounds simple, but it's arguably one of the most significant pillars of mathematical physics since it provided a focal point for aligning centuries of mathematical research into symmetry (e.g., Pythagoras, Hamilton, Hilbert, Lorentz, Klein & Lie) with centuries of physics research into conservation laws (e.g., Newton, Huygens, Galileo, Einstein, Weyl), offering inspiration for both fields and providing the mathematical basis for gauge theory, the framework modern physics uses to characterize the Strong, Weak, and Electromagnetic forces. Furthermore, Noether's work helps illuminate the elegance aesthetic at the heart of physics, the idea that the laws of reality embody a remarkable conceptual beauty, and this beauty ultimately derives from symmetry. As Nobel laureate Frank Wilczek puts it:

> [...] the idea that there is symmetry at the root of Nature has come to dominate our understanding of physical reality. We are led to a small number of special structures from purely mathematical considerations—considerations of symmetry—and put them forward to Nature, as candidate elements for her design. [...] In modern physics we have taken this lesson to heart. We have learned to work from symmetry toward truth. Instead of using experiments to infer equations, and then finding (to our delight and astonishment) that the equations have a lot of symmetry, we propose equations with enormous symmetry and then check to see whether Nature uses them. It has been an amazingly successful strategy. [56]

Another Nobel laureate, Philip Warren Anderson, goes even further: "It is only slightly overstating the case to say that physics is the study of symmetry." [57] Max Tegmark suggests that all the apparent information in reality could in theory be reconstructed from an account of which symmetries have undergone spontaneous symmetry breaking (SSB), and also that even "broken" symmetries are still respected under higher-dimension analysis: "[in the many-worlds hypothesis] the quantum superposition of field configurations decoheres into what is for all practical purposes an ensemble of classical universes with different density fluctuation patterns. Just as in all cases of spontaneous symmetry breaking, the symmetry is never broken in the bird's view, merely in the frog's view: a translationally and rotationally quantum state (wave functional) such as the Bunch-Davies vacuum can decohere into an incoherent superposition of states that lack any symmetry." [58]

Relevance to formalizing phenomenology

Two formal possibilities worth mentioning, albeit outside the scope of this work, are:

- Noether's theorem may apply simply and literally to phenomenology: if we can represent a conscious system in the same way we represent physical systems (as a Hamiltonian or Lagrangian), then symmetries in these equations should naturally correspond to invariances in phenomenology [1, 59];
- Symmetry and broken symmetry may be the most basic building blocks from which more complex structures are built in phenomenology, just as they appear to be in physics.

My immediate thesis here is more qualitative, that the aesthetics of physics can provide guidance about how to approach formalization in general. As general principles:

- Systematizing a phenomenon often involves pointing to what people are already doing implicitly and describing it explicitly;
- Making a formal connection between two things (in Noether's case, mathematical invariances and conserved quantities) helps make progress on both;
- The invariances (symmetries) of a system are a key entry point for understanding the nature of that system.

Finally, as a focus to bring to the next section, I offer that there is a symmetry aesthetic at the heart of mathematics and physics, and this will matter to the study of information geometries of mind:

- Symmetry is much more than just a 'neat quirk of geometry'; it's one of the few properties that are well-defined on all mathematical objects and functionally and foundationally important to modern physics. By implication, symmetry will be important in any field that inherits a theoretical aesthetic from physics or mathematics.

And so, if we have an information geometry of mind, a symmetry analysis of this geometry is likely to be the single best starting point for understanding what this geometry means. Most concretely, we should expect that the symmetry of a properly-constructed information geometry of mind should correspond with some phenomenological property of foundational significance.

4 The Symmetry Theory of Valence

> *The Symmetry Theory of Valence (STV): the symmetry of an information geometry of mind corresponds with how pleasant it is to be that experience.*

Details and definitions: what kind of thing is STV?
David Chalmers has described consciousness research as a search for "psychophysical laws", or principles which relate the physical world with the phenomenal [60]. Touching back on Section 2, we can divide psychophysical laws into Type 1 laws which deal with generating a correct Information Geometry of Mind (IGM), and Type 2 laws which deal with interpreting IGMs. STV is a Type 2 psychophysical law.

If STV is true, it is exactly true; it is not a loose rule of thumb, but an identity relation – symmetry in an IGM is the same thing as valence in the experience that IGM represents.

As stated above, STV raises three core questions:

1. How do we generate the formalism that STV operates on?
2. What is STV's metric for symmetry?
3. How can we apply STV to brains, and what predictions does it make?

1. Generating the information geometry of mind

STV makes no claim on how to generate an information geometry of mind – merely that a correct information geometry of mind exists and some present or future theory of consciousness may succeed at deriving it. However, see [1, 37, 59] for additional speculation.

An important caveat here is that, even if STV is exactly correct, it will only match reality when applied to the exactly correct information geometry of mind, and it's unclear how much convergence there will be across techniques for generating information geometries of mind. However, the upside is that, if we're confident STV is correct or at least plausible, we can use it as an independent sanity check on techniques for generating information geometries of mind. I.e., if a theory of consciousness is applied to a brain experiencing pleasure and generates an information geometry that is particularly low in symmetry, we can consider that theory less promising.

2. Toward a symmetry metric for complex objects

The size of an object's symmetry group is the gold standard for symmetry metrics and will be the standard for STV. However, calculating this exactly on real information geometries of mind may be epistemologically and computationally intractable for the foreseeable future: we don't have examples to work from, and for large objects, the analysis gets very complicated. One notable dynamic here is that as we add additional nodes to an object, the size of the object's symmetry group tends to plummet rapidly; it only takes one point out of position to break a symmetry. On the other hand, as we add more nodes to an object this tends to increase the underlying dimensionality of the object, and as this dimensionality increases the number of possible symmetries skyrockets because there are radically more available mathematical transformations. It's unclear which process 'wins' this race in higher dimensions, and the dimensionality of human-scale minds is likely to be very high; e.g., the dimensionality of objects in IIT's qualia space is n^2, where n is the number of nodes (and from IIT's formulation may be in the millions).

To hedge against these concerns, we can consider a basket of symmetry metrics for complex shapes, with computational methods trading off hypothetical rigour for tractability and practicality:

- **Formal construction:** if we can calculate the size of the IGM's symmetry group, we should. Depending on how monolithic the construction of the IGM is, it may also make sense to (1) count symmetries combinatorially or only across directly bound nodes, or (2) treat each particle (or other basic unit) contributing to the overall IGM as its own object, evaluate its symmetries, weight by contribution to the magnitude of the IGM, then sum across all such particles.
- **Cost function:** there exist practical approximations for finding partial symmetries in graphs: e.g., Christoph Buchheim's notion of fuzzy symmetry detection, which defines an algorithmic process for finding a 'closest symmetrical figure' and an 'edit distance' for how far the actual figure is from it [1].
- **Compressibility:** the act of data compression can be reframed as finding and encoding symmetries in data. However, standard library-based data compression algorithms (e.g., LZW) only look for a small subset of possible symmetries, and a more general AI-based compression approach may better reflect holistic symmetry [61]; see Section 6.

All that said, the simplest approach would be to assume the details of the construction of an IGM will uniquely determine what symmetry metric should be applied.

3. Qualitative and quantitative evidence for STV

We can consider two classes of evidence: the degree to which STV clarifies unresolved qualitative questions about the brain and mind, and the degree to which STV offers distinguishing quantitative

Symmetry Theory of Valence (STV)	Symmetry Theory of Homeostatic Regulation (STHR)	Symmetry Theory of Aesthetics (STA)
Symmetry in the mathematical representation of an experience corresponds to its pleasure	The brain's primary attractor is symmetry, and symmetry gradients in brain networks are coupled to homeostatic requirements	Symmetry is a core factor in what we find beautiful because symmetry in stimuli often translates to symmetry in internal networks
Law of the universe; exactly true in all conscious systems	Contingent fact (aliens or AIs could work differently)	Contingent fact (aliens or AIs could work differently)

Figure 4: Symmetry Theories

predictions. Section 5 discusses the former and Section 6 the latter.

5 Toward a new synthesis of aesthetics, metaphysics, and neural regulation

The promise of being able to say One True Thing about phenomenology is this should help clarify a broad range of topics because we can factor this thread from complex issues and get combinatorially less confusion. I propose the following factorization:

STA vs STV vs STHR: an ecosystem of symmetry theories

The Symmetry Theory of Valence (STV) is the conjecture that symmetry in an information geometry of mind corresponds with how pleasant it is to be that information geometry of mind. I offer that STV is (exactly) true, as a metaphysical brute fact.

The Symmetry Theory of Homeostatic Regulation (STHR) is the thesis that our homeostatic requirements are coupled to symmetry gradients in the nervous system [62]. I offer that STHR is (approximately) true because symmetry has favourable properties as a self-organization target, and our evolution has leaned on this extensively.

The Symmetry Theory of Aesthetics (STA) is a crystallization and update of Greco-Roman and Enlightenment aesthetics that mathematical symmetry and ratio are central pillars of aesthetic proportion. I offer that STA is (approximately) true because symmetry in stimuli often translates to symmetry in internal networks, which we seek out (STHR) and is intrinsically pleasant (STV) (also see [63]).

The truth of a theory may be judged in terms of falsifiability, but the success of a theory is better modelled by its intuitive and imaginative possibility. The following is an attempt toward a qualitative story of symmetry as an extremely broad and rich organizing principle for intelligent systems.

1. Symmetry is the simplest Schelling point for self-organization

Self-organization depends on the components of a future system being able to follow gradients. The symmetries of a network are frame-invariant and thus always well-defined, so there's always a well-defined gradient of improvement. This gradient is often inferable locally since networks with high symmetry have high redundancy and any development of asymmetry quickly and strongly propagates through the network. The simplicity of this logic allows symmetry-focused networks to self-assemble, and the causal coupling between local and global asymmetries lets systems use these gradients for computational purposes: local changes can be quickly judged as increasing or decreasing global symmetry. We can label networks which perform gradient descent on symmetry as 'symmetry networks'. STHR is the hypothesis that our regulatory networks are symmetry networks whose symmetry gradients are coupled to homeostatic requirements.

An important theme for STHR is that by using symmetry as a normative template, networks get many computational properties 'for free'. Much of the heavy lifting here will already have been done if we can formalize the relationship between symmetry, compression, and free energy minimization: Jürgen Schmidhuber's Compression Drive suggests we seek out novel experiences to improve our brain's tacit compression library [64]; Karl Friston's Free Energy Principle argues we must as a matter of long-term stability minimize short-term surprise [65]; György Buzsáki famously models the brain around conditionally phase-locked rhythmic oscillations [66]; Paul Smolensky and Steven Lehar write about minds built around harmonic principles [67, 68]. My tentative expectation is that under various limits, these can be shown to be part of the same equivalence class.

2. Nervous system as lock; symmetry as the outcome of a successful key

Freud's "pleasure principle" suggests humans seek out pleasure as their primary drive; STV+STHR suggest such a drive is best understood as the brain performing gradient descent on symmetry. We can consider the nervous system as built around a 'lock and key' system: the nervous system's complex topology functions as the tumblers on a lock, and the system performs gradient descent on this lock by seeking out stimuli or 'keys' that combine with this lock to create internal symmetry. STV's thesis is that the success condition is always the same, and very simple. This is consistent with, e.g., Kringelbach and Berridge's observation in Section I that "[t]he available evidence suggests that brain mechanisms involved in fundamental pleasures (food and sexual pleasures) overlap with those for higher-order pleasures (for example, monetary, artistic, musical, altruistic, and transcendent pleasures)" [14]. Or with apologies to Tolstoy, "All happy brains are alike; each unhappy brain is unhappy in its own way."

Insofar as the nervous system has nodes with local autonomy (e.g. Tononi's Maximally Irreducible Conceptual Structure (MICS) [69] or Safron's Self-Organizing Harmonic Modes (SOHMs) [70]), this lock-and-key system will be local to each ganglion's connectome. Essentially, every major nerve cluster is a different lock (complex dynamical landscape whose symmetry gradient is coupled with homeostatic requirements) and requires its own particular key.

Brains naturally seek out skeleton keys, stimuli that will reliably make us happy regardless of our initial physiological state. Likewise, adversaries may also look for skeleton keys that can mesmerize predators or prey. Evolution naturally fights back, and we can understand boredom as a powerful anti-wireheading technology which decreases our lock's tolerance, intended to push us out of pleasant equilibria and into action.

3. The strange pleasantness of music

Pinker has suggested "music is auditory cheesecake, an exquisite confection crafted to tickle the sensitive spots of at least six of our mental faculties" [71]. By describing the success condition of such 'auditory cheesecake' as symmetry in the brain and mind, STV offers a container for understanding how components of music may work to bring this result about.

In Principia Qualia, I offered the "non-adaptedness principle": "Evolutionary psychology lets us roughly estimate how pleasurable a stimulus should be given our evolutionary history, and we can compare this to how pleasurable the stimulus actually is. If we can find a stimulus where there's a large difference between these two quantities, it could be a hint of something interesting happening: some pattern 'directly hacking into' the mental pattern which produces pleasure/pain." [1] Music, flow states, and abstract mathematics are pleasant beyond what we would expect from evolutionary pressures; auditory dissonance is surprisingly unpleasant, and listening to two songs at the same time doesn't sum their pleasure but instead turns the experience negative. STV offers an intuitive story in each case.

STV implies that music is not an aberration but that all pleasurable states are ultimately rhythmic, consistent with Safron‘s observation that "the rhythmic nature of sexual activity is central for understanding the phenomenology of sexual trance and orgasm" [72].

4. Hedonic hotspots as tuning knobs

Much has been written about the brain's "hedonic hotspots" (Section 1), but we lack consensus on how these hotspots mechanistically generate pleasure. If STV is correct, the functional role of these hotspots is likely best described as tuning knobs for harmony in the brain [1].

5. A harmonic economy

Just as an organism is a confederation of cells, a nervous system is a confederation of ganglia. And as cells have their own economy of sticks and carrots meant to identify and prevent cancerous defection, so too each ganglia has its own toolset of harmonic sticks and carrots meant to align disparate organs. Put simply, if an organ isn't getting its needs met and wishes to raise an objection, it can strategically broadcast dissonance into other organs. If it is getting its needs met and wishes to train other organs with positive reinforcement, it can broadcast loud harmony – the stomach is notorious for both. Each organ has different channels available for this, depending on its position in the nervous system's topology.

6. Symmetry breaking is directional, which helps identify threats and needs

A symmetry network offers directional information about where disruptions are coming from, which can be inferred from the pattern of dissonance projected into the network, and this information is almost costless – it's "baked into" the network structure.

A starfish is essentially a ring of neurons; STHR suggests that homeostasis will be represented by an even flow of signals around this ring. If a predator or chemical risk appears – some fish starts nibbling on arm #4 – this flow representing homeostasis will be disrupted into dissonance, the location of the disruption can be inferred at many points in the network by the 'flavour' of the local dissonance, and the network can coordinate around strategies to reduce the dissonance (i.e. escape the danger). The starfish doesn't have to think or plan – it just needs to maintain internal neural harmony, and proper behaviour will follow. (Thanks to Jeff Lieberman for the example of the starfish.)

This 'lensing' property is especially useful when we multiplex different sensory flows together and snap normative judgments need to be made: is this chunk of phenomenology good? Should I move towards or away? Is it good to eat? Will it bite me? By multiplexing information flows into a symmetry network, these complex inferences are built-in: if something is associated with increased local and global symmetry, the control network will default to towardness-type actions; if it reduces internal symmetry, awayness-type actions. Nearly all actions involve tradeoffs along multiple dimensions, but global symmetry is a common currency which allows us to intuitively and rapidly navigate such tradeoffs.

We may classify interoceptive development as learning one's idiosyncratic "dissonance lensing patterns" – learning that when one's stomach hurts in such-and-such way, eating a specific class of food will make it feel better; when this specific flavour of dissonance appears, one should hand off executive control to that specific subagent; etc. Essentially, growing up involves learning what keys (stimuli) fit each of your organ's locks (dissonance patterns).

7. The golden ratio as non-interference technology

Safron writes about Self-Organizing Harmonic Modes (SOHMs) of the brain [70], or a collection of semi-sovereign harmony regimes. Much as each piece of music has a key signature, each SOHM has its own basis set of eigenmodes which selects and emphasizes certain frequencies while deadening others; Safron suggests this preferential selection corresponds to collapsing predictive networks' indecisive “Bayesian blur” of distributions into definite judgments and actions. No SOHM can encode all symmetries, much as no key signature can harmonize with all other key signatures. This leads to an imperative for limiting interactions between SOHMs: interference very likely leads to sensory ambiguity and energetic dissipation.

The golden ratio (approximately 1 : 1.618) is the optimal ratio for non-interference between frequencies: “synchronization of the excitatory phases of two oscillations with frequencies f1 and f2 is impossible (in a mathematical sense) when their ratio equals the golden mean because their excitatory phases never meet” [73]. I suggest this is the source of the golden mean’s aesthetic significance: we can think of stimuli as structured packages of symmetries, and when they are arranged in relation to the golden mean, they are, in a sense, prepackaged for separability and non-interference. Such stimuli should excite the body’s SOHMs without feeding their conflict.

High-frequency SOHMs may function as high-resolution feature detectors: as sensory patterns enter the nervous system’s sensory pipelines, the SOHMs within these pipelines naturally re-encode the sensations into component symmetries (similar to a Fourier decomposition), which can be correlated against a library of interpretations. Lower-frequency SOHMs may function similarly, but in a more non-localized and stochastic fashion: less like a circuit and more like wind chimes. Being in any specific environment will lead to specific patterns of SOHM activation, and by the presence and absence of particular SOHMs and their interactions we obtain a subconscious feeling about what kind of environment we’re in and where its rewards and dangers are. As sensations shift, so do the hum of our SOHMs. Emotions may be thought of as particular bundles of active low-frequency SOHMs, with the list of possible emotions heavily influenced by considerations around non-interference.

8. Biophysics, symmetry considerations, and gauge theories

The finalization of this manuscript coincided with a special issue of Interface Focus for The Royal Society, ”Making and breaking symmetries in mind and life” [74]. From the introduction:

> In this diverse collection of articles, we explored the roles of symmetry in complex adaptive systems across a multitude of scales—from the emergent dynamics of biophysical systems and their underlying mechanisms, to the shaping of adaptations through ontogenic and phylogenetic processes. Across these theoretical and empirical explorations, we hoped to demonstrate that the study of symmetries may illuminate fundamental properties of living and intelligent systems. Towards this end, we considered a broad range of perspectives on (a)symmetries, exploring the extent to which intersections may be found between seemingly disparate phenomena.
>
> One of the most powerful applications of symmetry-related concepts is found in gauge theory [75]. Whenever a physical theory has a redundant quantity, meaning a quantity that leaves a system’s dynamics invariant with respect to local changes in the value of that quantity (a ‘frame of reference’ or ‘gauge’), we can understand that quantity as a kind of abstract symmetry recorded in what is called a gauge field [76, 77, 78]. Deformations of gauge fields are understood as ‘fictitious’ forces; these forces restore the local symmetry of quantities that are dynamically invariant, by recording the system’s interactions with the field of possible gauges for that quantity. Gauge theories provide a general way of modelling physical systems. Notable use-cases include general relativity’s handling of gravity as the curvature of space–time and models of the attractive and repulsive forces in electromagnetic fields. These theories are so far-reaching that the word ‘fictitious’ may potentially be left out of descriptions of these emergent forces, as it may be the case that there are no other kinds [79].
>
> Gauge-theoretic perspectives on biophysics have been suggested in the past, especially in the context of brains as goal-seeking systems, guided by hierarchical information process-

> ing and prediction-error minimization. This is one aspect of the view presented by the free energy principle and active inference (FEP-AI) [76, 80, 65, 81]: the attracting states of nervous systems are understood as entailing predictions, where the consistent realization of these predictions can be viewed as the preservation of goal states, contingent on particular symmetries being enforced by a gauge field governing those dynamics. The importance of symmetry reaches deep into the functional aspects of the brain: mental causation may be understood as a kind of 'fictitious' force over neural dynamics (especially with respect to perception and action) [82], and a 'symmetry theory of valence' has even been proposed, in which pleasure and pain may best be understood as the degree to which mental systems transition to more or less symmetric states in some appropriate sense [1, 8, 83].

My conclusion is that one of the most generative heuristics in biology today is applying symmetry considerations to biophysical frames. This is not proof of STV, but it is suggestive.

6 Empirical predictions

STV is a formal, causal expression of the sentiment that "pleasure is harmony in the mind". If we can empirically measure this harmony and estimate valence, we can test this quantitatively.

We can consider three scenarios for how STV could be true:

1. STV is true, and is true in a legible way;
2. STV is true, but the brain and its relationship to the mind are complex enough that we shouldn't expect a story that is transparent and intuitive;
3. STV is true, but the proxies involved in our neuroimaging stack and their levels of measurement [28] do not lead to a characteristic picture of what's going on in the structure of consciousness.

Investing in scenario (1) is a matter of collecting methods for intuitively decomposing brain activity into components that can be symmetric or dissonant with each other and checking for supporting evidence; addressing scenario (2) requires a willingness to trade off legibility for raw predictive power; addressing scenario (3) requires a willingness to collect a wide portfolio of neuroimaging sources, with a preference toward minimally-processed data or data that is processed in an uncorrelated way vs. traditional methods.

Scenario 1: CSHW+CDNS as intuitive story

To date, attempts at empirically validating STV have built on Selen Atasoy's Connectome-Specific Harmonic Wave (CSHW) framework [84, 85, 86]. CSHW attempts to quantify connectome resonance by inferring a (10k node) connectome through DTI, calculating this connectome's eigenmodes, then determining which harmonic decomposition (distribution of energy across computed eigenmodes) best recreates the observed fMRI activity. Gomez Emilsson created a 'Consonance Dissonance Noise Signature' (CDNS) method which takes this power-weighted list of harmonics, evaluates the pairwise consonance, dissonance, and noise values between each, and sums the result into an overall CDNS 'score', with consonance as a proxy for symmetry [87]. I believe it's a very clever approach, though quantitative results have been inconclusive; my intuition is extending CDNS into DICE (Dimensionality, Integration, Consonance, Energy) may help to further isolate noise.

Scenario 2: dropping the constraint of legibility by focusing on compression

All else being equal, brain states with large amounts of symmetry should be more compressible. In practice this is a very noisy test since conventional compression approaches (a) are geared toward simple data regularities that miss wide classes of symmetries, and (b) may be a particularly poor fit for compressing brain imaging, given the many degrees of freedom in neuroimaging.

To get around these challenges, we can consider:

1. Casali's "Zap and zip" method, which first stimulates a brain with TMS, measures the after-echo with EEG, then tries to compress the result. Casali uses this to infer healthy brain microstructure: coma patients with a more compressible state are more likely to wake up since a more resonant echo implies the preservation of harmonic structure. Analyzing this data rather than resting state may offer a less noisy (or differently noisy) result than simply trying to compress brain activity [88];
2. Machine learning-based compression techniques that can adaptively search for a wide set of symmetries;
3. Applying compression metrics to CSHW;
4. Various combinations of (1-3).

Scenario 3: building a portfolio of neuroimaging data

A. Relative value of data types

There are a wide variety of neuroimaging types: EEG, fMRI, fNIRS, PET, and others, each with a complex set of tradeoffs between time resolution, spatial resolution, spatial depth, proxy for neural activity, and so on. Given that STV is a conjecture about structure, as a rule I expect that the value of each modality for validating STV will be roughly proportional to the value of the modality as a data source for holographic source reconstruction of brain activity.

Positive valence is a better target for study than negative valence for at least three reasons. First, STV is currently more clear about positive valence (although a privileged symmetry metric may also imply a privileged dissonance metric). Second, since harmony is convergent and self-similar, whereas disharmony is divergent, the signature of harmony is more clear. Third, since harmony is more stable over time, it's easier to measure with 'slower' imaging modalities like fMRI.

Self-reported valence may not be the "ground truth", but it's one of the most believable metrics. An alternative to using valence self-reports is drawing from datasets where we have an extremely strong prior that the subject is experiencing very positive valence. Two likely exemplary datasets are from people under the influence of a drug such as MDMA and from experienced meditators practising jhana meditation.

B. Adding wider nervous system data

To date, modern theories of consciousness have focused on data from the brain. However, it's also possible that a significant factor for symmetry in an IGM is the amount of harmony vs. disharmony between the brain and other organs. Likewise, if there are multiple centres of consciousness in the human body, this may be independently interesting [89]. The ideal data situation would be to have organ-level data from, e.g., the heart, stomach, and vagus nerve, rich enough to do realistic source reconstruction and paired (and precisely synched) with any brain data.

C. Electromagnetic metrics (e.g. pressure, stress, interference

Most (though not necessarily all) of our focus should be on tests which involve minimal assumptions about the precise substrate of consciousness and the precise method of constructing an information geometry of mind, although exploring symmetry in specific substrates may be useful (1) if harmony in the brain is self-similar in such a way that it will be robust under many different constructions, and (2) if we can identify flows where symmetry in one substrate is causally upstream of symmetry in others.

In particular, I suggest attempting to paint a picture of EMF dynamics will be helpful for STV for at least two reasons: (1) directly measuring EM phenomena is ideal if consciousness is an EM phenomenon [34, 90, 1, 91], and (2) EMF-centric modelling may offer novel proxies or sanity checks for connectome-level structure – i.e., the macroscopic dynamics of a brain's EM field as driven by aggregated LFPs are likely highly correlated with whatever connectome-level events 'matter' for consciousness and might tell us something interesting about connectome-level dynamics that are difficult to see with current imaging.

A wildcard is that it's possible that STV would be intuitively obvious with high-resolution imaging of the EM field at select frequency bands – how would we know without looking? EMF patterns from everyday activity may be very noisy, but some forms of meditation are reported as being extremely intense yet stable states. A simple story here is that attention lines in phenomenology correspond to field lines in the EM spectrum [92]; a simple experiment would be to ask an accomplished meditator (or several) to focus their attention on a single physical object and measure the EM field around both them and the object. This could be repeated for different types of meditation and for different instructions, e.g., "now try to smooth out the EM field", "create harmony in the EM field". If this produced noticeable differences in how directed or stable the EM field was, we could explore further.

A systematic approach here would involve collecting a standard suite of well-defined properties in the EM field and testing for correlations with valence, with special focus on terms also used in phenomenology. Shinzen Young has suggested stress, interference, and turbulence as promising metaphors, and has suggested that mental processes that generate suffering are analogous to physical processes that generate heat [93]. We might extend this metaphor and attempt to find a proxy or physical measure for STV by looking at correlations between valence and the electromagnetic equivalents of entropy-affecting processes or dynamics that perform thermodynamically irreversible work.

The core challenge for this path is creating the imaging.

A path forward

My preferred path for testing STV blends these principles and scenarios in the following way:

- Start with a compression pipeline that is built around methods and algorithms which can compress a very wide range of symmetries in brain data. The goal wouldn't be absolute compression efficiency but to apply a sufficiently general compression algorithm (sensitive to as many symmetries as possible) such that it's directionally correct in terms of how 'generally compressible' some piece of neuroimaging is. In theory, the best machine learning algorithms should exhibit this sort of sensitivity naturally [61, 94].
- Once we have this compression pipeline, we can throw all manner of data into it and see how much more compressible brain data from pleasant experiences is compared to the baseline and negative experiences.
- We can then add preprocessing that performs intuitive dimensionality reduction such as CSHW, CDNS, and DICE in order to reduce noise.

By using compression as our proxy for symmetry and building a data pipeline around the intention of being 'directionally correct' about compressibility, we de-emphasize most metaphysical & methodological questions and can rapidly cycle through various types and sources of data.

7 Discussion

The Symmetry Theory of Valence is both aggressive and simple. It's aggressive in that it purports to offer an exact, single-factor solution to one of the oldest mysteries in phenomenology, what makes some things feel better than others. It's simple in that it fits in one sentence, merely combining three concepts: (1) information geometries of mind, (2) symmetry, and (3) valence. Each pairwise connection is already strong:

- 1+2: The history of physics and mathematics gives us a strong prior to analyze information geometries of mind in light of symmetry;
- 1+3: By definition, information geometries of mind fully describe a mind's valence;
- 2+3: There's a long cultural history of understanding pleasantness and beauty in terms of symmetry.

As with any novel theory, it would be strange if STV was true. But I hold that it would be even more strange – there would be some deep violation of beauty somewhere – if it wasn't.

An ecosystem of Type 1 and Type 2 psychophysical laws

If we take the formalism path and assume that calculating an information geometry of mind is the correct goal of consciousness research, we can envision a rich future research ecosystem with dozens of distinct methods for calculating an information geometry of mind (Type 1 psychophysical laws), spanning computational, functional, and physical theories of mind, and dozens of different mathematical heuristics for interpreting the meaning of an information geometry of mind (Type 2 psychophysical laws), and people freely mixing and matching between each.

The strength of this ecosystem will be in the surprising predictions and elegant compressions it allows, and the crux will be the value which accrues to those who take information geometries of mind as (1) real and/or (2) serious sources of insight. As with all new dimensions of analysis, this will enable and require new forms of validation and reward careful thought.

I believe STV is a good place to start.

Acknowledgements

A sincere thank-you to Romeo Stevens for his calm interest and presence; Andres Gomez Emilsson for believing in symmetry and a fruitful and enjoyable 7-year collaboration; Quintin Frerichs and Andrew Zuckerman for their time and energy; Randal Koene for kicking off formalizing STV with his interest in collaboration; Sean McGowen, Mackenzie Dion, Patrick Taylor, Elin Ahlstrand, and the rest of QRI's volunteers and donors for their hard work; and Justin Mares, Joe Kelly, Jason Dorsett, Lili Mao, Ivanna Evtukhova, Elena Zaiceva, and Elena Selezneva for many forms of support. Special thanks to Rafael Harth for indispensable help preparing this manuscript; Adam Safron for belief, support, and pointers on the functional role of SOHMs; and my parents for their unconditional support. Section 1 previously appeared in [1]; Section 3 is an adaptation of [95]. Additional thanks to Andres Gomez Emilsson and Quintin Frerichs for conversations about EM theories of consciousness during 2020-2021, and Radhika Dirks for the same during 2016-2017.

References

[1] Michael Edward Johnson. "Principia Qualia". In: *URL https://opentheory.net/2016/11/principia-qualia* (2016).

[2] Jaak Panksepp. “Affective neuroscience of the emotional BrainMind: evolutionary perspectives and implications for understanding depression”. en. In: *Dialogues Clin Neurosci* 12.4 (2010), pp. 533–545.

[3] Jeffrey C Cooper and Brian Knutson. “Valence and salience contribute to nucleus accumbens activation”. en. In: *Neuroimage* 39.1 (Aug. 2007), pp. 538–547.

[4] Amanda Bischoff-Grethe et al. “The influence of feedback valence in associative learning”. en. In: *Neuroimage* 44.1 (Sept. 2008), pp. 243–251.

[5] Nico H Frijda. “The laws of emotion”. In: *American Psychologist* 43 (1988), pp. 349–358.

[6] Gerald L Clore, Karen Gasper, and Erika Garvin. *Affect as information.* Mahwah, NJ, US, 2001.

[7] Wolfram Schultz. “Neuronal Reward and Decision Signals: From Theories to Data”. en. In: *Physiol Rev* 95.3 (July 2015), pp. 853–951.

[8] Mateus Joffily and Giorgio Coricelli. “Emotional Valence and the Free-Energy Principle”. In: *PLOS Computational Biology* 9.6 (June 2013), pp. 1–14. DOI: `10.1371/journal.pcbi.1003094`. URL: `https://doi.org/10.1371/journal.pcbi.1003094`.

[9] Eran Eldar et al. “Mood as Representation of Momentum”. In: *Trends in Cognitive Sciences* 20.1 (Jan. 2016), pp. 15–24.

[10] Kent C Berridge and Morten L Kringelbach. “Neuroscience of affect: brain mechanisms of pleasure and displeasure”. en. In: *Curr Opin Neurobiol* 23.3 (Jan. 2013), pp. 294–303.

[11] Jimmy Jensen et al. “Separate brain regions code for salience vs. valence during reward prediction in humans”. en. In: *Hum Brain Mapp* 28.4 (Apr. 2007), pp. 294–302.

[12] Kent C Berridge, Terry E Robinson, and J Wayne Aldridge. “Dissecting components of reward: ’liking’, ’wanting’, and learning”. en. In: *Curr Opin Pharmacol* 9.1 (Jan. 2009), pp. 65–73.

[13] Adam Shriver. “The unpleasantness of pain for humans and other animals”. In: *Philosophy of pain* (2018), pp. 147–162.

[14] Morten L Kringelbach and Kent C Berridge. “Towards a functional neuroanatomy of pleasure and happiness”. en. In: *Trends Cogn Sci* 13.11 (Sept. 2009), pp. 479–487.

[15] H C Cromwell and K C Berridge. “Where does damage lead to enhanced food aversion: the ventral pallidum/substantia innominata or lateral hypothalamus?” en. In: *Brain Res* 624.1-2 (Oct. 1993), pp. 1–10.

[16] Kyle S Smith et al. “Ventral pallidum roles in reward and motivation”. en. In: *Behav Brain Res* 196.2 (Oct. 2008), pp. 155–167.

[17] Geoffrey R O Durso, Andrew Luttrell, and Baldwin M Way. “Over-the-Counter Relief From Pains and Pleasures Alike: Acetaminophen Blunts Evaluation Sensitivity to Both Negative and Positive Stimuli”. en. In: *Psychol Sci* 26.6 (Apr. 2015), pp. 750–758.

[18] V E Dyakonova. “Role of Opioid Peptides in Behavior of Invertebrates”. In: *Journal of Evolutionary Biochemistry and Physiology* 37.4 (July 2001), pp. 335–347.

[19] Bruno Van Swinderen and Rozi Andretic. “Dopamine in Drosophila: setting arousal thresholds in a miniature brain”. en. In: *Proc Biol Sci* 278.1707 (Jan. 2011), pp. 906–913.

[20] Jeremiah D Osteen et al. “Selective spider toxins reveal a role for the Nav1.1 channel in mechanical pain”. en. In: *Nature* 534.7608 (June 2016), pp. 494–499.

[21] Andres Gomez Emilsson. “State-Space of Drug Effects: Results”. In: *Qualiacomputing: The Blog of Andres Gomez Emilsson* (2015).

[22] Danica Marković, Radmilo Janković, and Ines Veselinović. "Mutations in Sodium Channel Gene SCN9A and the Pain Perception Disorders". In: *Advances in Anesthesiology* 2015 (Feb. 2015), p. 562378. ISSN: 2356-6574. DOI: 10.1155/2015/562378. URL: https://doi.org/10.1155/2015/562378.

[23] Joost P H Drenth and Stephen G Waxman. "Mutations in sodium-channel gene SCN9A cause a spectrum of human genetic pain disorders". en. In: *J Clin Invest* 117.12 (Dec. 2007), pp. 3603–3609.

[24] Justin Heckert. "The Hazards of Growing Up Painlessly". In: *New York Times* (Nov. 2012).

[25] Todd B. Kashdan, Robert Biswas-Diener, and Laura A. King. "Reconsidering happiness: the costs of distinguishing between hedonics and eudaimonia". In: *The Journal of Positive Psychology* 3.4 (2008), pp. 219–233. DOI: 10.1080/17439760802303044. eprint: https://doi.org/10.1080/17439760802303044. URL: https://doi.org/10.1080/17439760802303044.

[26] C Nathan Dewall et al. "Acetaminophen reduces social pain: behavioral and neural evidence". en. In: *Psychol Sci* 21.7 (June 2010), pp. 931–937.

[27] Martha J Farah. "Brain images, babies, and bathwater: critiquing critiques of functional neuroimaging". en. In: *Hastings Cent Rep* Spec No (Mar. 2014), S19–30.

[28] William R. Uttal. *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience*. The MIT Press, 2011. ISBN: 9780262526654. URL: http://www.jstor.org/stable/j.ctt9qf99r (visited on 12/12/2022).

[29] Eric Jonas and Konrad Paul Kording. "Could a Neuroscientist Understand a Microprocessor?" In: *PLOS Computational Biology* 13.1 (Jan. 2017), pp. 1–24. DOI: 10.1371/journal.pcbi.1005268. URL: https://doi.org/10.1371/journal.pcbi.1005268.

[30] Naotsugu Tsuchiya et al. "No-Report Paradigms: Extracting the True Neural Correlates of Consciousness". In: *Trends in Cognitive Sciences* 19.12 (2015), pp. 757–770. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2015.10.002. URL: https://www.sciencedirect.com/science/article/pii/S1364661315002521.

[31] Lisa Feldman Barrett. "Are Emotions Natural Kinds?" en. In: *Perspect Psychol Sci* 1.1 (Mar. 2006), pp. 28–58.

[32] Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0". In: *PLOS Computational Biology* 10.5 (May 2014), pp. 1–25. DOI: 10.1371/journal.pcbi.1003588. URL: https://doi.org/10.1371/journal.pcbi.1003588.

[33] Scott Aaronson. "Why I am not an integrated information theorist (or, the unconscious expander)". In: *Shtetl Optimized: The Blog of Scott Aaronson* (2014).

[34] Adam Barrett. "An integration of integrated information theory with fundamental physics". In: *Frontiers in Psychology* 5 (2014). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00063. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00063.

[35] Dalton A R Sakthivadivel. *Weak Markov Blankets in High-Dimensional, Sparsely-Coupled Random Dynamical Systems*. 2022. DOI: 10.48550/ARXIV.2207.07620. URL: https://arxiv.org/abs/2207.07620.

[36] Daniel C Dennett. *Consciousness explained*. Penguin uk, 1993.

[37] Michael Edward Johnson. "Against functionalism: why I think the Foundational Research Institute should rethink its approach". In: *Opentheory: The Blog of Michael Edward Johnson* (2017).

[38] Christopher Bamford. *Homage to Pythagoras: Rediscovering sacred science*. SteinerBooks, 1994.

[39] Robert Lawlor. *Sacred Geometry: Philosophy and Practice*. 1982.

[40] Fra Luca Pacioli. *Divina proportione*. Ripol Klassik, 1969.

[41] Andrea Palladio. *The four books on architecture*. Mit Press, 2002.

[42] Morris Hicky Morgan, Herbert Langford Warren, et al. *Vitruvius: the ten books on architecture*. 1914.

[43] Leon Battista Alberti. *On the art of building in ten books*. Mit Press, 1991.

[44] Hermann LF Helmholtz. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Cambridge University Press, 2009.

[45] G Rhodes et al. "Attractiveness of facial averageness and symmetry in non-western cultures: in search of biologically based standards of beauty". en. In: *Perception* 30.5 (2001), pp. 611–625.

[46] Harpal Harrar, Simon Myers, and Ali M Ghanem. "Art or Science? An Evidence-Based Approach to Human Facial Beauty a Quantitative Analysis Towards an Informed Clinical Aesthetic Practice". en. In: *Aesthetic Plast Surg* 42.1 (Jan. 2018), pp. 137–146.

[47] George David Birkhoff. "Aesthetic measure". In: *Aesthetic Measure*. Harvard University Press, 2013.

[48] Veronika Douchová. "Birkhoff's aesthetic measure". In: *Auc Philosophica Et Historica* 2015.1 (2016), pp. 39–53.

[49] Manil Suri. "What happens if we make the Mona Lisa more symmetrical?" In: *Psyche* (Nov. 2022).

[50] Noson S. Yanofsky and Mark Zelcer. "The Role of Symmetry in Mathematics". In: *Foundations of Science* 22.3 (Mar. 2016), pp. 495–515. DOI: 10.1007/s10699-016-9486-7. URL: https://doi.org/10.1007%2Fs10699-016-9486-7.

[51] Katherine Brading and Elena Castellani. *Symmetries in Physics: Philosophical Reflections*. 2003. DOI: 10.48550/ARXIV.QUANT-PH/0301097. URL: https://arxiv.org/abs/quant-ph/0301097.

[52] David J. Gross. "The role of symmetry in fundamental physics". In: *Proceedings of the National Academy of Sciences* 93.25 (1996), pp. 14256–14259. DOI: 10.1073/pnas.93.25.14256. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.93.25.14256. URL: https://www.pnas.org/doi/abs/10.1073/pnas.93.25.14256.

[53] Emily Conover. "In her short life, mathematician Emmy Noether changed the face of physics". In: *ScienceNews* (June 2018).

[54] Chris Quigg. *Colloquium: A Century of Noether's Theorem*. 2019. DOI: 10.48550/ARXIV.1902.01989. URL: https://arxiv.org/abs/1902.01989.

[55] Max Tegmark. *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Penguin UK, Jan. 2014.

[56] Frank Wilczek. *A beautiful question: Finding nature's deep design*. Penguin, 2016.

[57] P. W. Anderson. "More Is Different". In: *Science* 177.4047 (1972), pp. 393–396. DOI: 10.1126/science.177.4047.393. eprint: https://www.science.org/doi/pdf/10.1126/science.177.4047.393. URL: https://www.science.org/doi/abs/10.1126/science.177.4047.393.

[58] Max Tegmark. "The Mathematical Universe". In: *Foundations of Physics* 38.2 (Nov. 2007), pp. 101–150. DOI: 10.1007/s10701-007-9186-9. URL: https://doi.org/10.1007%2Fs10701-007-9186-9.

[59] Michael Edward Johnson. "Taking Monism Seriously". In: *Opentheory: The Blog of Michael Edward Johnson* (2019).

[60] D J Chalmers. "The puzzle of conscious experience". en. In: *Sci Am* 273.6 (Dec. 1995), pp. 80–86.

[61] Henry W. Lin, Max Tegmark, and David Rolnick. "Why Does Deep and Cheap Learning Work So Well?" In: *Journal of Statistical Physics* 168.6 (Sept. 2017), pp. 1223–1247. ISSN: 1572-9613. DOI: 10.1007/s10955-017-1836-5. URL: https://doi.org/10.1007/s10955-017-1836-5.

[62] Michael Edward Johnson. "Why we seek out pleasure: the Symmetry Theory of Homeostatic Regulation". In: *Opentheory: The Blog of Michael Edward Johnson* (2017).

[63] Andres Gomez Emilsson. "Harmonic Society: 8 Models of Art for a Scientific Paradigm of Aesthetic Qualia". In: *Art Against Art* 6 (2019). URL: https://www.artagainstart.com/p/issue-6.html.

[64] Jürgen Schmidhuber. "Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes". In: *Anticipatory Behavior in Adaptive Learning Systems*. Ed. by Giovanni Pezzulo et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 48–76. ISBN: 978-3-642-02565-5.

[65] Friston Karl. "The free-energy principle: a unified brain theory?" In: *Nature Reviews Neuroscience* 11.2 (Feb. 2010), pp. 127–138.

[66] György Buzsáki. *Rhythms of the Brain*. Oxford University Press, Oct. 2006.

[67] Paul Smolensky. "Harmony in Linguistic Cognition". In: *Cognitive Science* 30.5 (2006), pp. 779–801. DOI: https://doi.org/10.1207/s15516709cog0000_78. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0000_78. URL: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0000_78.

[68] Steven M Lehar. *The world in your head: A gestalt view of the mechanism of conscious experience*. Psychology Press, 2003.

[69] G Tononi. "Integrated information theory of consciousness: an updated account". en. In: *Arch Ital Biol* 150.2-3 (June 2012), pp. 56–90.

[70] Adam Safron. "An Integrated World Modeling Theory (IWMT) of Consciousness: Combining Integrated Information and Global Neuronal Workspace Theories With the Free Energy Principle and Active Inference Framework; Toward Solving the Hard Problem and Characterizing Agentic Causation". en. In: *Front Artif Intell* 3 (June 2020), p. 30.

[71] Steven Pinker et al. *How the mind works*. Vol. 524. New York Norton, 1997.

[72] Adam Safron. "What is orgasm? A model of sexual trance and climax via rhythmic entrainment". en. In: *Socioaffect Neurosci Psychol* 6 (Oct. 2016), p. 31763.

[73] Belinda Pletzer, Hubert Kerschbaum, and Wolfgang Klimesch. "When frequencies never synchronize: the golden mean and the resting EEG". en. In: *Brain Res* 1335 (Mar. 2010), pp. 91–102.

[74] Adam Safron et al. "Making and breaking symmetries in mind and life". In: *Interface Focus* 13.3 (Apr. 2023). DOI: 10.1098/rsfs.2023.0015. URL: https://doi.org/10.1098/rsfs.2023.0015.

[75] Bomark NE. "Teaching gauge theory to first year students". en. In: (Sept. 2020). URL: https://arxiv.org/abs/2009.02162.

[76] Erik D. Fagerholm et al. "Conservation laws by virtue of scale symmetries in neural systems". In: *PLOS Computational Biology* 16.5 (May 2020). Ed. by Peter Neal Taylor, e1007865. DOI: 10.1371/journal.pcbi.1007865. URL: https://doi.org/10.1371/journal.pcbi.1007865.

[77] Maldacena J. “The symmetry and simplicity of the laws of physics and the Higgs boson”. en. In: *Eur. J. Phys* 37 (2015), p. 015802.

[78] Rovelli C. “Gauge is more than mathematical redundancy”. In: *One Hundred Years of Gauge Theory*. Ed. by Silvia De Bianchi, Claus Kiefer, et al. Springer, 2020.

[79] Carroll S. *The big picture: on the origins of life, meaning, and the universe itself*. en. New York, NY: Penguin, 2016.

[80] Biswa Sengupta et al. “Towards a Neuronal Gauge Theory”. In: *PLOS Biology* 14.3 (Mar. 2016), e1002400. DOI: 10.1371/journal.pbio.1002400. URL: https://doi.org/10.1371/journal.pbio.1002400.

[81] A Tozzi et al. “Gauge fields in the central nervous system”. In: *The physics of the mind and brain disorders: integrated neural circuits supporting the emergence of mind*. Ed. by Ioan Opris and Manuel F Casanova. Springer, 2017.

[82] Safron A. “The radically embodied conscious cybernetic Bayesian brain: from free energy to free will and back again Entropy”. en. In: *Entropy* 23 (June 2021), p. 783. DOI: doi:10.3390/e23060783.

[83] Hesp C et al. “Deeply felt affect: the emergence of valence in deep active inference”. en. In: *Neural Comput* 33 (Feb. 2021), pp. 398–446. DOI: doi:10.1162/neco_a_01341.

[84] Selen Atasoy, Isaac Donnelly, and Joel Pearson. “Human brain networks function in connectome-specific harmonic waves”. In: *Nature Communications* 7.1 (Jan. 2016), p. 10340. ISSN: 2041-1723. DOI: 10.1038/ncomms10340. URL: https://doi.org/10.1038/ncomms10340.

[85] Selen Atasoy et al. “Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under LSD”. In: *Scientific Reports* 7.1 (Dec. 2017), p. 17661. ISSN: 2045-2322. DOI: 10.1038/s41598-017-17546-0. URL: https://doi.org/10.1038/s41598-017-17546-0.

[86] Selen Atasoy et al. “Harmonic Brain Modes: A Unifying Framework for Linking Space and Time in Brain Dynamics”. en. In: *Neuroscientist* 24.3 (Sept. 2017), pp. 277–293.

[87] Andres Gomez Emilsson. “Quantifying Bliss: Talk Summary”. In: *Qualiacomputing: The Blog of Andres Gomez Emilsson* (2017).

[88] Adenauer G Casali et al. “A theoretically based index of consciousness independent of sensory processing and behavior”. en. In: *Sci Transl Med* 5.198 (Aug. 2013), 198ra105.

[89] Giulio Tononi and Christof Koch. “Consciousness: here, there and everywhere?” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1668 (2015), p. 20140167. DOI: 10.1098/rstb.2014.0167. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2014.0167. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2014.0167.

[90] J. McFadden. “The Conscious Electromagnetic Field: The Hard Problem Made Easy?” In: *Journal of Consciousness Studies* (2002).

[91] Susan Pockett. “Consciousness Is a Thing, Not a Process”. In: *Applied Sciences* 7.12 (2017). ISSN: 2076-3417. DOI: 10.3390/app7121248. URL: https://www.mdpi.com/2076-3417/7/12/1248.

[92] Michael Edward Johnson. “The Binding Problem: principles and conjectures”. 2021.

[93] Michael Edward Johnson. “Shinzen Young interview: stress, equanimity, & sensory clarity”. In: *Opentheory: The Blog of Michael Edward Johnson* (2021).

[94] Alexandre Défossez et al. *High Fidelity Neural Audio Compression*. 2022. DOI: 10.48550/ARXIV.2210.13438. URL: https://arxiv.org/abs/2210.13438.

[95] Michael Edward Johnson. “Emmy Noether and the symmetry aesthetic”. In: *Opentheory: The Blog of Michael Edward Johnson* (2022).