2-6-2018 11:30 AM

Phase transitions of Integrated Information in the Generalized Ising Model of the Brain

Sina Khajehabdollahi, *The University of Western Ontario*

Supervisor: Soddu, Andrea, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Physics

Abstract

The bold framework of the Integrated Information Theory of consciousness are explored in this thesis in the context of the generalized Ising model of the brain. Small 5-node networks are simulated on the Ising model with Metropolis transitions where the fitting parameter T is fit to empirical functional connectivity matrices of healthy human subjects. Fitting to criticality, results indicate that integrated information undergoes a phase transition at the critical temperature T_c. The results are interpreted in the context of an emerging perspective of the science of complexity and perhaps even the philosophy of science; the universe as a self-organizing critical system undergoing cascades of phase transitions into complexity.

Keywords: integrated information, consciousness, ising model, criticality, phase transitions, emergence, complexity

Co-Authorship Statement

This thesis has been written by Sina Khajehabdollahi under the supervision of Dr. Andera Soddu. The work in chapter two is in preparation to be submitted to PNAS (Proceedings of the National Academy of Sciences of the United States of America).

Acknowledgments

I wish to express my gratitude and appreciation to my colleagues and supervisor, Dr. Andrea Soddu, for spending countless hours toiling over this research with me and allowing me to express and sharpen my thoughts and ideas in a critical environment. I wish to give my gratitude to Larissa Albantakis and Will Mayner, both of which took their time to help guide me through many of the difficult technical details of IIT and greatly accelerated my research and understanding of the theory. I wish to express my love and appreciation to my mother, father and family for their immeasurable support and guidance. Finally, I wish to thank my dear friend Mahnaz for her friendship and kindness through difficult times. This thesis could not come together in any way without the cumulative effort of all these people and many not mentioned and for that I can only express my humble gratitude to the unimaginable cooperation of the universe around me.

Table of Contents

List of Figures

List of Acronyms

IIT	Integrated Information Theory
MIP	Minimum Information Partition
MICE	Maximally Irreducible Cause-Effect Repertoire
MICS	Maximally Irreducible Conceptual Structure

Brain Networks

AUD	Auditory
DMN	Default Mode
ECNr	Executive Control Right
ECNl	Executive Control Left
SAL	Salience
SEN	Sensorimotor
VISl	Visual Lateral
VISm	Visual Medial
VISo	Visual Occipital

Chapter 1

Consciousness

How do a few billion neuronal cells wired together in complex networks of communities give rise to consciousness? Why does it feel like something to be this interacting conglomeration of matter? The first step to understanding is imitation and to that end, in this thesis we use a generalized Ising model to model the brain as a critical system. Integrated information, a measure of consciousness proposed by the **Integrated Information Theory (IIT)** of consciousness is measured alongside the traditional statistical measures of the model. IIT claims that the measurable quantity 'integrated information' *quantifies* consciousness in a system and an object called the 'conceptual structure' of the system describes the *quality* of the conscious experience. In this thesis the bold ideas of IIT are combined with the emerging physical perspective of the critical brain to explore the interaction of these ideas. Our results indicate that consciousness, an abstract macroscopic object or order parameter, seems to undergo a phase transition at the critical temperature of the Ising model. At the onset of this phase transition, the integrated information (consciousness) generated by the system is maximally susceptible. We interpret our results in light of recent arguments that complexity, life, and consciousness all tend to be critical systems through evolution due to the adaptive capabilities of critical systems (Goldenfeld and Woese, 2011; Hopfield, 1994; Mora and Bialek, 2011). While the numeric results of our experiment are simple and succinct, contextualizing our methods and then justifying our conclusions from these results requires a journey back and forth through several different disciplines, a recurring theme in the physics of meta-cognition. Each discipline has its own unique language, its dialect, and a regime of relevance. Each semi-independent piece must be stitched at its boundaries to some other discipline to form a unified puzzle made from components that may look nothing like each other. This puzzle is the causal structure of our universe. This thesis is mostly interested in the 'brain and consciousness' patch of this puzzle, though a holistic perspective of the puzzle is still useful, if not necessary. To understand the brain at a fine scale, one my look at the individual anatomy of a single neuron, its action potentials, ion channels, metabolic network and all, and attempt to describe that as accurately as possible and construct upwards from there. Here, differential equations, organic chemistry, and biology are the useful and relevant languages one would have to learn to discuss these subjects. At a larger scale, the brain is a non-equilibrium thermodynamic system where billions of similar (but crucially non-identical) cells interact both in cooperation and segregation to give rise (or self-organize) to new qualities (order parameters) that did not exist before (e.g. pain, thought, consciousness). At this scale we employ the languages and methods of statistics, statistical mechanics and thermodynamics to explore these ideas. At an even larger scale, sets of brains interact and once again self-organize to construct even larger and newer qualities that did not exist before (communities, societies, cultures, economies). These scales are predominantly classified under 'complex systems', an accelerating frontier of scientific research which has been a catalyst of interdisciplinary research. By acknowledging this variable scope, a new sort of meta-concept begins to emerge: the iterative act of traversing scales, coarse-graining, zooming out, mixing disciplines. The concept of traversing scales is in some idealized ways explained by the physics of renormalization group flow, a relatively modern mathematical apparatus which has among other accolades helped explain the ubiquity of universality for systems undergoing phase transitions. Universality as a general phenomenon is the observation that many disparate systems with dissimilar microscopic origins can give rise to identical meso- or macroscopic properties. A phase transition as a general phenomenon is the observation that many systems seem to have abrupt or discontinuous changes in their defining characteristics. Perhaps not so coincidentally, from gaining a better understanding of the phenomena of phase transitions many new ideas also condense out, such as the informational/-computational capabilities of critical systems, or their capabilities to adapt/evolve in changing environments. Furthermore, universality can give much-needed relief in helping predict a variety of complex systems by classifying 'micro-complexes' into 'macro-simplexes', to put it simply. Universality classes can act like attractors for dynamics, and give hope that even complex systems like the brain might one day be understood more succinctly as a realization of a particular universality class(es) whose properties can be simply described by a set of critical exponents which describe the scaling relations of important parameters of the system. There is still, however, an elephant in the room. From an idealistic perspective, why did the cumulative branches of knowledge, the disparate disciplines, emerge the way that they did? From a material perspective, why did the universe after the big bang undergo a cascade of phase transitions as new modes of being and order condensed out of the initially symmetric universe. Why is it that in a universe that we expect to tend towards the mundane of disorder we see such diverse modes of existence and complexity? Is the emergence of consciousness particularly special in this grand cascade? This thesis does not pretend to have answers to these questions but instead present them as vectors of exploration in the endeavor to understand complexity. While the symmetric origins of our cascading universe remain mysterious, the observation of this pattern helps understand how a bridge between the ideal and material world might be built. In the paradigm of the critical brain hypothesis, the critical brain is a consequence of a complex, heterogeneous world. The brain is the system which has evolved over time towards criticality to best imitate, predict and be adaptive to, the dynamic, critical environment in which it resides. This thesis hopes to motivate, among a number of other supporting ideas, the notion that a self-organizing critical brain is an evolutionarily attractor for reasons that seems to penetrate into the heart of meta-cognition, physics, and the philosophy of science as a whole.

1.1 What is Consciousness?

Describing consciousness initially seems like it could be a relatively simple task. After all, it is the thing we know and experience most intimately. A relatively unrestricted definition may be proposed as ‘what it feels like to be something’. However even such a simple description only poses more questions. How do we define the concept of ‘feeling’, or ‘being’. These questions are traditionally approached by philosophers interested in ontology, however if we ever hope to explore the ideas of consciousness or the theories that arise attempting to explain it, it seems inevitable that we must brave the winds and leave the comforts of our home disciplines and venture into foreign lands. What follows in the next two sections is a digression on the ideas of ‘being’ and ‘feeling’ which the author of this thesis hopes to prime the mind to the subjectivity and intrinsic quality of the integrated information theory of consciousness. These sections can also be skipped without loss of continuity into other sections.

1.1.1 To be or not to be

If the fundamental particles of this universe are only ever things like electrons and quarks, is it meaningful to think of a macroscopic object which is simply composed of these fundamental particles as existing in any distinct way from its constituent components? The famous thought experiment of the Ship of Theseus presents an interesting variation of this idea.

The thought experiment poses the question that if the components of the ship are slowly replaced one by one until the ship has been completely refurbished, is it still the same ship? In response to the paradox that arises from the thought experiment, one can argue that it's not the constituent components that matter, but the relationship between the components. So the individual planks of wood in the ship do not define the ship, it is the relationship that the planks of wood have with each other. However, one can argue that the Ship of Theseus as an individual object is contingent on the history of the matter it is composed of; it is not simply a 3 dimensional object in space, but a 4 dimensional object in time. If the definition of the object includes its history, modifying the ship can be seen as an act of modifying just 3 dimensions of this 4 dimensional object, so the ship's identity is preserved. Conversely if we ignore its history, replacing the ship's parts can be seen to be equivalent to creating a replica of the ship, that is, a separate but congruent object to the original ship.

It is here the concept of the ship as an individual entity begins to blur as one can ask: at what point in the replacement process does the ship stop being the Ship of Theseus?. If this ship is a 4 dimension object, can it ever split into two objects, and at what point? Clearly, this paradox highlights the ambiguity of the abstract macro-objects we as humans in our human language use to describe the world. In fact, one conclusion to the Ship of Theseus paradox is that what we call the "Ship of Theseus" doesn't actually exist in any physical way. It simply exists in our minds as a virtual object, a smoothed out idealization of reality that acts as a useful tool to approximate the rough edges of our uncertain world. This idealization can include the history of the ship or it can ignore it. The exact boundaries of our idealizations arise from the utility of our idealizations and the contexts in which they are relevant. This implies that while our idealizations aren't completely arbitrary or random, they may not necessarily be universal or unique.

Yuval Noah Harari, author of Sapiens: A Brief History of Humankind (Harari and Perkins, 2014) argues that the human ability to believe in virtual objects and stories are what separated humans from other species on Earth. The ability for humans to coordinate and form large communities arose from the fact that we were all capable of believing the same stories and through such unification we built order and organization. So the objects in our universe that 'exist' are not always exactly material, do not always exist in the same way for different entities, and exist insofar as they have an influence on some entities and are not contingent on some crudely defined sense of 'objectivity'. This abstraction of what 'can be' is captured in IIT by the intrinsic notion of information which is defined more technically in the Axioms and Postulates sections. The motivations behind the demand that *any* theory of consciousness must have such an intrinsic quality deserves its own digression which is discussed in the following section.

1.1.2 Subjectivity and Objectivity

The distinction between subjectivity and objectivity is hard to physically pin-point. For instance, how could one possibly know if something is objectively true? Even our scientific methods depend on the reproduction and verification of previous results. We only ever call something a scientific fact when it has been repeatedly verified to the point that it being a statistical anomaly is sufficiently small (where the definition of 'sufficiently small' is essentially an arbitrarily small number). In other words, the scientific method is contingent on society's self-organizing, error-correcting capabilities. Only observations that are agreed upon in an inter-subjective way are considered to be objective in science. Any individual observation is akin to a subjective experience and objectivity is reserved for a large number of sufficiently similar, non-contradictory subjective observations. This is all to say that what we call objectivity is simply the accumulation of a large number of subjective experiences that agree with each other. Objectivity exists insofar as an abstract but useful idealization/summary statistic of subjectivity; ‘what is unlikely to be an accidental experience’.

I have not commented on the potential objectivity of mathematics. One can argue that even if there are no objective experiences, surely there are some objective relationships and self-consistent properties one could imagine. The answer to this question is likely hidden in the answer to the following question: “Is our universe a mathematical one?” After all, mathematics may be yet another idealization that our universe simply does not obey. So far it has been one of the most useful, fundamental, and self-correcting idealizations we as humans have come up with, and there is good reason to believe that our universe is indeed a mathematical universe. If we *are* residing in a mathematical universe, there is still quite a large space for subjectivity to reign, if only from a general and special relativity perspective, or a from a quantum information perspective. In general, it seems quite natural to imagine in the 21st century that the subjective nature of our universe, from the subjectivity in defining entropy, to the relativity of simultaneity, or even the indecision of coherent, ‘unobserved’ entangled system, that a holistic theory of the emergence of our universe will have to acknowledge the subjective quality of existence. The axioms (or the self-evident truths) of consciousness in IIT are therefore phenomenologically derived from experience and offer a ‘top-down’ approach to defining consciousness. Instead of asking how neural mechanisms can en masse give rise to consciousness, IIT defines consciousness and asks how and when neural mechanisms can satisfy such conditions.

1.1.3 To feel

Simply put, to feel something is to be affected by its presence; its presence makes a difference to the trajectory of the object’s phase space in an informative way from the intrinsic perspective of the object. A photo diode can ‘feel’ light, an electron can ‘feel’ a proton, but are these objects conscious? So far we defined objects, what ‘can be’, from the perspective of the forces they impart. However an opposing perspective can fill in the negative space if we attempt to define how and in what context these objects are felt. This will be the duality between mechanisms and their purview which will be discussed in more detail in the Integrated Information Theory section. In short, to feel something means to be sensitive to its supposed existence such that counterfactually, its lack of existence would change the trajectory of the object in a way that no other force could mimic. It must make a difference that makes a difference.

Integrated Information Theory attempts to define the boundaries of existence for macroscopic objects by analyzing the causal structure of the containing system. Only causal structures that are irreducible or have irreducible components exist intrinsically. For something to be irreducible means that it is not sufficiently describable if one looks at the system in partitions; the system considered as a whole contains much more information than when it is partitioned. Therefore in IIT objects exist insofar that their causal structure as a whole is irreducible. Consciousness arises as causal structures becomes more and more irreducible. Irreducible causal structures are a ubiquitous concept, in mathematics as prime numbers, in physics as elementary particles, or in language as words or letters. IIT aims to measure the irreducibility of systems, claiming that the larger the irreducibility, the greater the conscious experience. Finally, it is in these irreducible structures that the concepts of 'feeling' and 'being' are found in a tangled knot, where the objects 'being' are only defined as such if they are capable of feeling themselves. In short, to be is to interbe (Schindler).

1.1.4 Loss of Consciousness & Brain Injury

Losing or altering consciousness is one of the easiest ways to become aware of its existence in the first place. We lose our consciousness every night when we sleep, we alter our consciousness with drugs and alcohol and food, and we even seem to witness our consciousness explore itself as we experience our dreams and lucid sleep states.

Even more strangely, certain anesthetics can lead to loss of consciousness whereas others keep it intact (Alkire et al., 2008). Measuring, quantifying and describing these fluctuations is another frontier of research in the grind towards understanding consciousness. New methods aiming to capture brain complexity have been measured to decrease in these altered states of consciousness which we phenomenologically expect to be a decrease in consciousness (Casali et al., 2013; Sarasso et al., 2015). Furthermore, the safe usage of non-invasive techniques such as transcranial magnetic stimulation (TMS) can allow researchers to probe and perturb the brain and spark another major avalanche in the research on consciousness and brain dynamics in general.

1.2 Integrated Information Theory

Integrated Information Theory makes the bold claim that integrated information *is* consciousness. The quantity of consciousness is represented by the value of Φ and the quality of consciousness is described by what is called the conceptual structure. The theory is built from phenomenological reports of consciousness. A set of axioms are built that should in principle be agreed upon by all conscious humans capable of expressing their experiences. The axioms are meant to represent self-evident truths. A set of postulates follow from the axioms that attempts to translate the axioms into a mathematical language. The postulates are assumptions on the nature of physical reality. If the axioms of IIT are indeed "self-evident truth" then criticisms of IIT must be centered on the postulates, for everything else arises from these set of rules. The theory aims to make predictions about consciousness that are otherwise ambiguous such as the quality and quantity of consciousness in humans that cannot report their experiences, in infants, animals, neural networks, etc. This section aims to review the basic principles behind IIT but does not attempt to give a comprehensive account of all its details which is given in the newest version by Oizumi et al. (2014).

1.2.1 Axioms

Existence

Consciousness as a phenomenon exists. Descartes' statement "I think therefore I am" puts it most succinctly.

Composition

Consciousness is composed of multiple concepts arranged in distinct ways. For example the experience of seeing the colour red does not preclude the experience of seeing the colour blue elsewhere at the same time.

Information

Consciousness is informative by specifying a particular experience out of all possible experiences. To have any one particular experience means that other experiences are not being had and this is informative.
NOTE: The definition of information used in IIT is notably different from the definition of information that was introduced by Shannon (Shannon, 2001). The information in IIT is described as intrinsic as opposed to the extrinsic nature of Shannon information. Shannon information is prescribed by an extrinsic observer who is capable of assessing the statistical dependencies between the inputs and outputs of their system of interest. Furthermore, this information is meaningless. The extrinsic observer is assumed to know what the meaning of the output is but the quantified information has no dependence on the actual meaning of the observations.

In IIT information is defined through 'causal' relationships and not just statistical relationships. The cause of a system in a particular configuration is assessed by perturbing the system into all possible configurations and calculating its transition probabilities. In physics, this would be equivalent to solving the partition function of a system. Therefore, the information in IIT is intrinsic to the self-interactions a system has with itself. This also ensures that the information and meaning are congruent in IIT. The meaning of the system is encoded in the exact shape of its conceptual structure, which when quantified is expressed as the [integrated] information.

Integration

A conscious experience is integrated and irreducible. For example, if one looks at a red apple, the experience of red and the experience of apple are not distinct from each other. You cannot separate the whole experience into two experiences, one of the apple, the other of the colour red. This axiom aims to capture the sensation of the holistic experience that cannot be separated into independent but simultaneous components. This axiom is at the very heart of IIT.

An irreducible system cannot be separated into independent components whose combined dynamics recreates the whole system. An irreducible system has some unique behaviour (causal structure) that is above and beyond the 'sum of its parts'. Integrated systems have emergent properties that the components of the system were not capable of causing by themselves. In IIT, only integrated information contributes to the conscious experience. Information on its own is not enough, the information needs to arise from an integrated causal structure.

Exclusion

Conscious experiences are exclusive in the sense that one cannot have some superposition of experiences. Furthermore, experiences have boundaries in both space and time. Experience seems to have a distinct scale in time so that conscious percepts have a well-defined temporal grain. This axiom bounds the extent of the conscious complex so that two complexes never overlap. For example your sense of self does not exist in superposition to the sense of self of the left and right hemispheres independently. Temporally, the experience of the flow of time has a particular cadence. Looking at moving clouds for 10 minutes is not equivalent to a timelapse of those same clouds compressed into 10 seconds.

1.2.2 Postulates

For each axiom there follows a postulate that aims to translate the axiom into a mathematical assumption on the nature of reality. Here, only 3 of the 5 postulates will be discussed. The information, integration, and exclusion postulates contain most of the important mathematical concepts that will be used throughout IIT.

Definitions

Some vocabulary that is necessary to describe the postulates are defined below. Their context and utility should become clear as the postulates are reviewed.

Mechanism: A mechanism is any component (or set of components) of a system that has a causal role. Specifically it is a subset of the components of a system. In the brain, a single neuron or a set of neurons can be a mechanism. For example in an Ising model of N spins, the set of two spins (s_1, s_2) are one from a power-set of possible mechanisms.

Cause-effect repertoire: The cause-effect repertoire is the probability distributions of the possible causes and effects of a system constrained to a particular state $\vec{s_t}$ at time t. if one knows the transition probabilities for a particular configuration of the system then the probabilities for past or future states can be calculated by constraining the state of a mechanism in a system. The cause-effect repertoire is calculated for the purview of interest. For example if we constrain the two spins:

$$\vec{s_t} = (s_1, s_2) = \uparrow, \downarrow \tag{1.1}$$

and calculate probabilities for the past or future states of the purview:

$$\begin{aligned} \text{cause repertoire} &= p\left((s_1, s_2, s_3)_{t-1} | \vec{s}_t\right) \\ \text{effect repertoire} &= p\left((s_1, s_2, s_3)_{t+1} | \vec{s}_t\right) \end{aligned} \tag{1.2}$$

Purview: A purview is, like a mechanism, a subset of components in a system from which the cause-effect repertoire is calculated for. The mechanism defines the components that are to be constrained to a state and the purview defines the components of the system whose cause and effect repertoire is to be calculated. In the example given in equation 1.2, the purview is composed of the components (s_1, s_2, s_3).

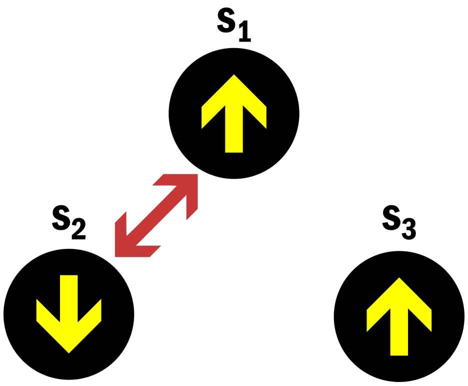

Figure 1.1: Example system of 3 spins. Spins 1 and 2 interact with each other and are disconnected from spin 3.

Information

If a system is constrained with respect to some mechanism in a state $\vec{s}$, and if this constraint yields probabilities for the past/future that is more informative than the 'unconstrained' repertoire p^{uc}, then the mechanism is informative. The unconstrained repertoire is the probability distribution of the past/future when no mechanisms are constrained to a state. In the case of the unconstrained cause repertoire, the maximum entropy distribution corresponds to the uniform distribution. In the case of the unconstrained effect repertoire, this will depend on the dynamics of system where the probabilities of the future are calculated assuming all the past states are equally likely. The comparison between the two distributions can be defined in different ways where in IIT 3.0 the Earth Mover's Distance is used. Formally, it can be written as:

$$\begin{aligned} \text{cause information} &= D\left(p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t\right) \ || \ p^{uc}((s_1, s_2, s_3)_{t-1})\right) \\ \text{effect information} &= D\left(p\left((s_1, s_2, s_3)_{t+1} \mid \vec{s}_t\right) \ || \ p^{uc}((s_1, s_2, s_3)_{t+1})\right) \\ \text{cause-effect information} &= \min(\text{cause information, effect information})) \end{aligned} \tag{1.3}$$

If constraining a mechanism to a state does not constrain its cause-effect repertoire any better than the unconstrained repertoire, then constraining the mechanism doesn't 'feel' like anything to the system. Only "differences that make a difference" are informative and exist from the intrinsic perspective of the system. On the other hand if constraining a mechanism to a state strongly specifies the cause-effect repertoire of a purview (by sharpening the probability distributions of the causes/effects) then the constraint mechanism is informative and can contribute to consciousness.

Integration

Mechanisms can only contribute to consciousness if they are irreducible. If the cause-effect repertoire specified by a mechanism in a state can be reduced into a product of independent components, then the information is not integrated. For example in Figure 1.1 spins s_1, s_2 are causally disconnected from spin s_3, where the 3 spins define the whole system. We could decompose the cause/effect repertoire of this system as the product of independent partitions A and B:

$$p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t\right) = p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t / A, B\right) \tag{1.4}$$

where

$$p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t/A, B\right) \equiv p\left((s_1, s_2)_{t-1} \mid \vec{s}_t^{A}\right) \times p\left((s_3)_{t-1} \mid \vec{s}_t^{B}\right) \tag{1.5}$$

If equation 1.4 is true, no integrated information is generated by this particular mechanism in a state. If it is not true, then the distance between the partition-product cause-effect repertoire (right hand side of equation 1.5) and the unpartitioned cause-effect repertoire is taken. Generalizing to some partition P we can write the integrated information ϕ as:

$$\phi^{P}_{cause} = D\left(p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t\right) \;||\; p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t/P\right)\right) \tag{1.6}$$

The partition that minimizes this distance is defined as the minimum information partition.

$$MIP = \arg\min_{P} \phi^{P} \tag{1.7}$$

Integrated information is the distance between the cause-effect repertoire of the unpartitioned mechanism and its minimum information partition (MIP). The Earth Mover's Distance is once again used for these calculations as of IIT 3.0. When we talk about ϕ we are usually talking about the ϕ calculated under the MIP unless otherwise specified. An example MIP is illustrated in Figure 1.2.

$$\phi^{MIP}_{cause} = D\left(p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t\right) \;||\; p\left((s_1, s_2, s_3)_{t-1} \mid \vec{s}_t/\mathrm{MIP}\right)\right) \tag{1.8}$$

Conceptually what we are doing here is approximating the holistic cause-effect repertoire by the probability product of its partitioned mechanisms' cause-effect repertoires. The partitioned repertoires are marginalized versions of the holistic repertoire, so these probability products at best recreate the holistic repertoire, or at worst information is lost about the holistic repertoire. The amount of information lost by

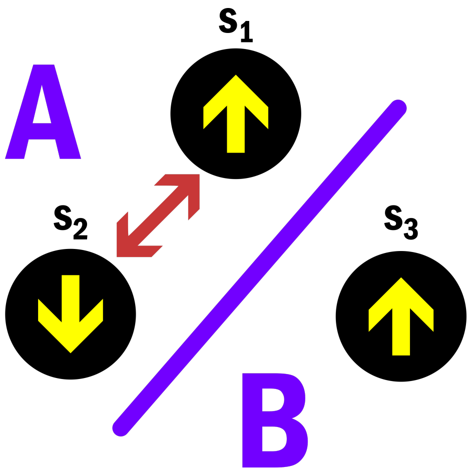

Figure 1.2: The MIP for the 3 spin system. Since spins s_1 and s_2 are causally disconnected from s_3, partitioning these elements makes no difference to the cause/-effect repertoires and the mechanism $(s_1, s_2, s_3 = \uparrow, \uparrow, \downarrow)$ is therefore not integrated. However, the mechanism $(s_1, s_2 = \uparrow, \downarrow)$ generates integrated information because partitioning that mechanism would sever a causal interaction that would therefore make a difference to the cause/effect repertoire of this mechanism.

this act of partitioning is equal to the integrated information of the system. If a particular partition is found such that the information lost is minimized, we have found the most natural partitioning that best decomposes the holistic system. Since we are only interested in measuring irreducible causal structures, we always calculate ϕ with respect to the MIP as it defines the cruelest cut for the partition. In other words, the irreducibility of a causal structure is prodded out of the system cut after cut until the cruelest cut, the MIP is found, where by analogy, the information that is bled after each cut is a measure of this irreducibility.

For a system with N components there are 2^N possible subsets that can be made so the number of possible partition combinations scales very quickly with the number of components. To naively check if a mechanism is integrated, all possible partitions must be checked until the MIP is found to measure the irreducibility of the mechanism. This process is one of the first super exponential computational jumps in IIT.

Exclusion

Each mechanism can only contribute a single cause-effect repertoire to consciousness. For each mechanism only one cause and one effect can exist, called the **core cause/effect**.

When constraining a mechanism to a state, all possible purviews of the whole system are used to find the purviews which contain the core cause and core effect. For each mechanism in a state, only one purview can contain its core cause or core effect, the purview which maximizes the generation of integrated information ϕ^{max}. For all mechanism/purview pairs, a **maximally irreducible cause-effect repertoire (MICE)** must be found following the steps from the integration postulate. For each of the $2^3 = 8$ possible mechanisms in our simple 3 particle system there are also $2^3 = 8$ possible purviews that that could contain its core cause or core effect. This is illustrated in Figure 1.3. This same diagram can be made to illustrate all possible mechanisms.

This postulate is motivated by the idea that mechanisms must only ever have a singular cause or effect and not a superposition of causes. For example the core cause of a thrown ball falling to the ground is due to Earth's gravity, not Earth's gravity plus the Moon's gravity plus all the other negligible forces that interact with the ball. In this example, the purview containing the core cause/effect would be from Earth's gravitational interactions with all other irrelevant interactions kept outside of the purview.

IIT claims that each mechanism can only have one cause/effect therefore for each mechanism. Therefore one must search through all possible purviews in order to find the one that gives maximal integrated information. "The core cause[-effect] of a mechanism from the intrinsic perspective is its maximally irreducible cause[-effect] repertoire". The MICE is an extension of the concept of the MIP from the integration postulate if one imposes the constraint that every mechanism only has a

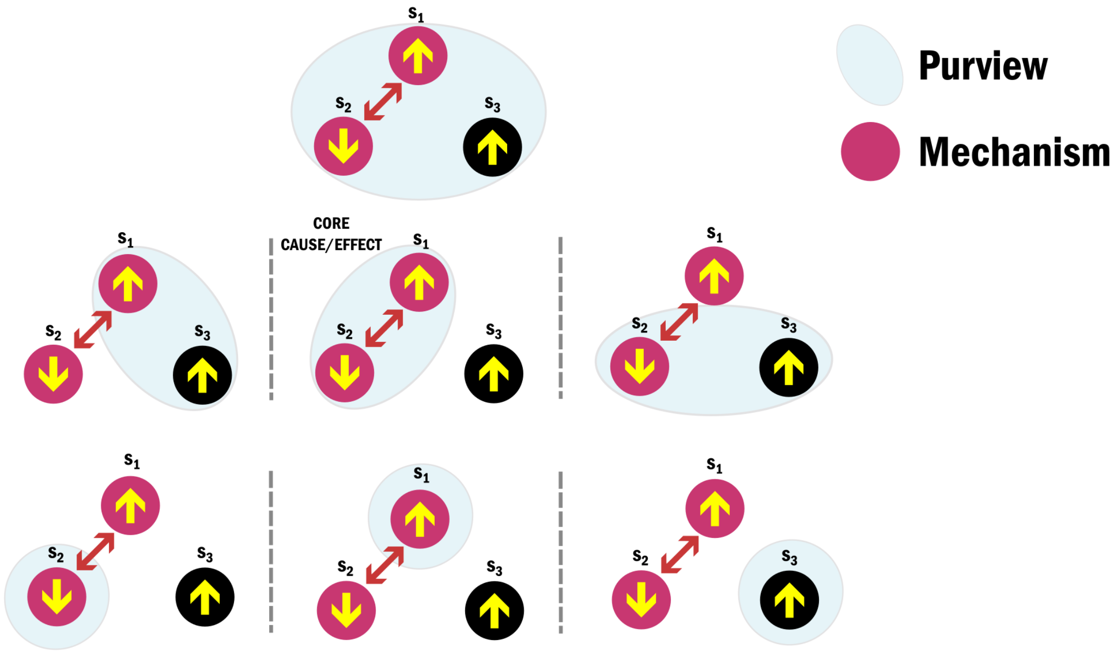

Figure 1.3: All the possible purviews for the 3 spin system are shown by the ellipses filled in blue. The mechanism is set by the pink spins in the state $\vec{s}_t = (s_1, s_2)_t = \uparrow, \downarrow$. The core cause/effect is specified as purview $P = s_1, s_2$. In this case, since these two spins are only interacting with each other, the probabilities for their future and past are entirely contingent on the state of the two spins at the current time. Since the spins are not connected to any other spins, their core effects will also only be on each other. The core cause/effect generate maximal ϕ^{Max} with respect to the other possible causes/effects.

single cause/effect. The MIP defines what to look for in order to reduce a system to its components. Then the Exclusion postulate demands that the largest irreducible objects be found (MICE), or in other words to find the core causes/effects.

In short, for each mechanism, there is a search over all possible purviews and all possible partitions in order to define a single **concept**. "A concept: a mechanism that specifies a maximally irreducible cause-effect repertoire" Oizumi et al. (2014). Finally, by sweeping through all possible mechanisms all the concepts in a system can be found; not all mechanisms necessarily have core causes/effects (and these mechanisms therefore generate no integrated information). Here we are already seeing the beginnings of the exponential growth of the computational complexity of IIT as the postulates of the theory are constructed. It is like trying to find all the prime numbers by first counting all the composite numbers and seeing what's left. A further iterations of these postulates on *sets* of mechanisms gives rise to the notion of **conceptual structures** and, analogous to the MICE, the **maximally irreducible conceptual structure (MICS)**. **Summary:** The mechanism is the set of elements that constrains the system. The purview is the set of elements whose past/future probability distributions are conditioned by the mechanism. The set of elements that define the mechanism and the set of elements that define the purview can overlap but don't have to. If a mechanism in a state does not constrain the past and future probability distributions of a purview, then that mechanism is not informative in that purview. It did not "make a difference that makes a difference". The information lost by partitioning a particular mechanism/purview pair quantifies how integrated a mechanism is with the elements across the partition. To give a grim analogy, this is like measuring the life of something by cutting it and measuring its spilled blood. Partitioning a mechanism amounts to 'noising' across the connections of the partition which can only decrease or leave unchanged the information generated by the mechanism. The act of finding the minimum information partition (MIP) for a particular mechanism/purview pair tells one the partition which maximally reduces the mechanism. The MIP can be intuitively thought of as the 'natural borders' of the cause/effect structure of a system. In general, the MIP tends to separate disconnected or weakly connected elements in a system, however for complex systems the MIP can be quite abstract and non-trivial to find. At the moment the simplest strategy for finding the MIP is brute force though new algorithms are emerging to considerably speed this process up (Kitazono et al., 2017). A concept is specified by a particular mechanism and its corresponding core cause/effect (if they exist). Core causes and effects are found by iterating over the power set of purviews for a particular mechanism. The particular purview that maximizes integrated information is the one that specifies the core cause/effect of that mechanism. There, each concept specifies a mechanism, a purview, and a MIP associated with that mechanism/purview pair.

Note: Sets of Mechanisms

In IIT 3.0, the current newest version of the theory, these ideas are extended to systems of mechanisms. The ideas described above are in the context of a single mechanism in a state. For a single mechanism in a state, the MIP and MICE must be found in order to describe the concept associated with the mechanism. A hierarchy of concepts can be generated from the power set of the mechanisms in the whole system. Again, not all of the mechanisms in the power set of mechanisms will give rise to concepts, but some may, and this set of concepts is referred to as the **conceptual structure**. Once again, there is a systematic process of checking if this conceptual structure is informative or integrated. This layer of complexity is not addressed in this thesis but it is worth acknowledging the vast combinatorial space one faces in the endeavor of measuring integrated information.

1.2.3 IIT Calculations and Methods

As detailed in the section above, calculating the integrated information that a particular system in a particular state generates is a combinatorially tedious task. First, from the entire set of elements in the system one can generate a power set of mechanisms. For each of these mechanisms in the power set one must then look across the power set of purviews in order to find the purview which maximizes integrated information. To calculate integrated information, the minimum information partition for each possible mechanism/purview pair must be found, which itself requires the exploration of all possible partitions. If one uses a brute force technique for all these steps, one must explore the entire combinatorial space across 3 different power sets of all the elements in the system. This is clearly not a computationally trivial task and is one of the largest problems one faces when attempting to test IIT. This characteristic difficulty of calculating integrated information can be summarized as a matrix factorization problem (Tegmark, 2016).

Transition Matrix Factorization

Tegmark (2016) summarizes this process succinctly: All physical processes can be defined with a Markov matrix or transition matrix $\mathbf{M}$. This transition matrix completely defines the dynamics of the system. (For real physical systems one simply needs to ensure that the time-step is sufficiently small in order to use a transition matrix to accurately portray the system.) The transition matrix $\mathbf{M}$ describes everything about the dynamics of a system and combined with knowledge of the system's state at a particular time, one can make use of the transition matrix to calculate probability distributions $\mathbf{p}$ for future or past states. This is an equivalent rephrasing of the cause/effect repertoires discussed in the sections above. Some example transition matrices for the Ising model at different temperatures are illustrated in Figure 1.4.

The process of calculating integrated information is then equivalent to attempting to find factors for the matrix $\mathbf{M}$ in the form of $\mathbf{M}^A \otimes \mathbf{M}^B$. For systems that are composed of independent components, there will be a natural factorization that will separate the independent components such that the full matrix $\mathbf{M}$ can be reduced to independent factors $\mathbf{M}^A$ and $\mathbf{M}^B$. For systems that are integrated, there will be no factorization possible. For such systems, the strategy then is to find approximate factors such that their product *approximately* resembles the original whole system. This is the process of finding the minimum information partition. Once approximate factors are defined, a new approximated transition matrix $\mathbf{M} \approx \mathbf{M}' = \mathbf{M}^A \otimes \mathbf{M}^B$ is generated. Then, given some constraint, for example when the system is in a particular state, one can calculate the cause/effect repertoires of these two transition

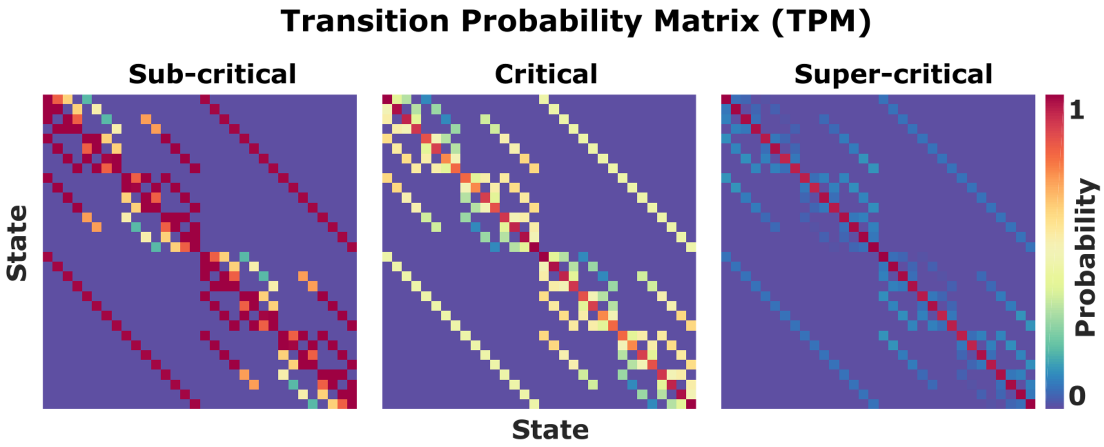

Figure 1.4: The **transition probability matrix (TPM)** of three different temperatures of the Ising model, each characteristic of its particular phase. The deterministic nature of the sub-critical regime, the exploratory nature of the critical regime, and the equiprobable nature of the super-critical regime are clearly visualized in these matrices.

matrices. For the approximate transition matrix $\mathbf{M}'$, we can call this probability distribution $\mathbf{q}$.

Integrated information is then the distance, by some metric, between the vectors $\mathbf{p}$ and $\mathbf{q}$ which demonstrates the amount of information lost by the act of factoring (partitioning) the system. For each possible approximate factorization there exists some value of integrated information that is generated. The combinatorial task of finding the factorization which minimizes the integrated information generated is the same process outlined in the previous section of finding the minimum information partition.

Tegmark (2016) has outlined a taxonomy of methods that integrated information theory can employ. In his review the calculations involved in IIT are separated into 4 steps:

1. Defining an approximate factorization method.

2. Defining the nature of the probability distributions $\mathbf{p}$ and $\mathbf{q}$ to compare. This choice is analogous to the choice of purviews that one may take from the section

above, though it is also slightly more general.

3. Defining what is known about the system when calculating the probability distributions in step 2. This is analogous to the choice of constraint one imposes on the system from the section above. For example if it is its current state that is known, its past state that is known, or a probability distribution of states that is known, etc.

4. Defining the metric by which the probability distributions **p** and **q** are compared. In older version of IIT, the K-L divergence was used, in IIT 3.0 the Earth Mover's Distance is used.

Clearly, the methods involving calculating integrated information are still being developed. For these reasons at the onset of this research project a simplified version of the IIT calculations was implemented. This simplified version written in MATLAB is notably faster than the more complete version that is made accessible by the authors Oizumi et al. (2014) in the python library pyPhi. The distinctions and simplifications implemented in our code are outlined below. The Python documentation of the python library pyPhi explains with more detail the intricacies of implementing IIT calculations (Oizumi et al., 2014).

MATLAB code

In order to better understand the methodology of IIT's calculations, a simplified version of the algorithm was written in MATLAB. Though the final results of this thesis utilize the python library of IIT 'pyPhi' (Oizumi et al., 2014), it is a fruitful exercise to recreate the more basic processes involved in IIT. The distinctions between the simplified version of the algorithm and that of pyPhi are outlined below.

1. The MATLAB IIT code does not consider sets of mechanism in the way that IIT 3.0 defines. The MATLAB IIT code is based on a simpler version of the

theory from (Oizumi et al., 2014). In this version only individual mechanisms are analyzed.

2. The probability distributions that are compared in the MATLAB code are those of the cause repertoire. The effect repertoire is not considered. In principle it is relatively straight-forward to implement the forward direction of the simplified IIT algorithm, but due to computational constraints this is ignored. The choice one makes in defining the probability distributions that are compared is non-trivial and are reviewed by Tegmark (2016) in step #2 of the taxonomy of IIT.

3. Mechanisms and purviews are not considered separately as they are in IIT 3.0. In IIT 3.0, the distinction between mechanisms, which constrain the system, and purviews, which define the cause/effect of that mechanism, are an integral part of the notion of 'concepts' and the 'maximally irreducible cause-effect repertoire'. In the simplified algorithm, when a mechanism is partitioned in search of the MIP, its purview is always the same elements as the mechanism. No distinction is made between mechanisms and purviews. This has the result that if a mechanism whose core cause/effect is part of a purview that does not contain itself, our simplified algorithm will not be able to detect such integration. Therefore our method of calculating ϕ does not find the MICE and therefore may give false positives or negatives when trying to assess if a mechanism generates integrated information.

4. In IIT 3.0, sets of mechanisms give rise to sets of concepts called a conceptual structure. Since our algorithm does not find the MICE, concepts are not defined. Furthermore, no such higher-order structure of concepts are defined either in the simplified version. The only mechanism that is analyzed is the mechanism of the whole system. That is, the state of the entire system is what is used

> to constrain the probability distributions. Therefore the simplified algorithm is not sensitive to mechanisms in the system that are smaller than the whole system. This has the result that the simplified algorithm can sometimes miss smaller-order integrated mechanisms.

Overall, the simplified algorithm mainly revolves around finding the MIP of a system. It is not particularly sensitive to the higher-order structures that may exist in the system and as such can be considered as a first approximation of the integrated information that a system may generate.

Example Calculations

An example is given to illustrate the process of calculating integrated information. First we begin by defining our system, which we choose to be an Ising model with $N = 5$ nodes. A connectivity matrix inspired by the Default Mode Network (DMN) of the brain is used (see chapter 2 for details and the appendix for region labels). This connectivity matrix is illustrated in Figure 1.5-A. A temperature close to criticality is chosen for the model (where the temperature is the only fitting parameter). In general, this is well approximated by the temperature where its susceptibility curve peaks (Binney et al., 1992; Stanley, 1971) as illustrated in Figure 1.5-B.

Setting $T = T_c$, we can now calculate the transition probabilities for the system. At temperatures approaching 0, the system behaves increasingly deterministic. All initial conditions eventually fall into one of two possible minimum energy states after a small number of iterations. (For larger networks with more complicated, perhaps modular, hierarchical or frustrated graphs with negative interaction weights, there may be a number of meta-stable states as well.) At temperatures approaching infinity, the system behaves stochastically and is dominated by thermal fluctuations/noise; the interactions between spins becomes negligible.

Once a TPM is generated, IIT calculations can begin. First, the system must

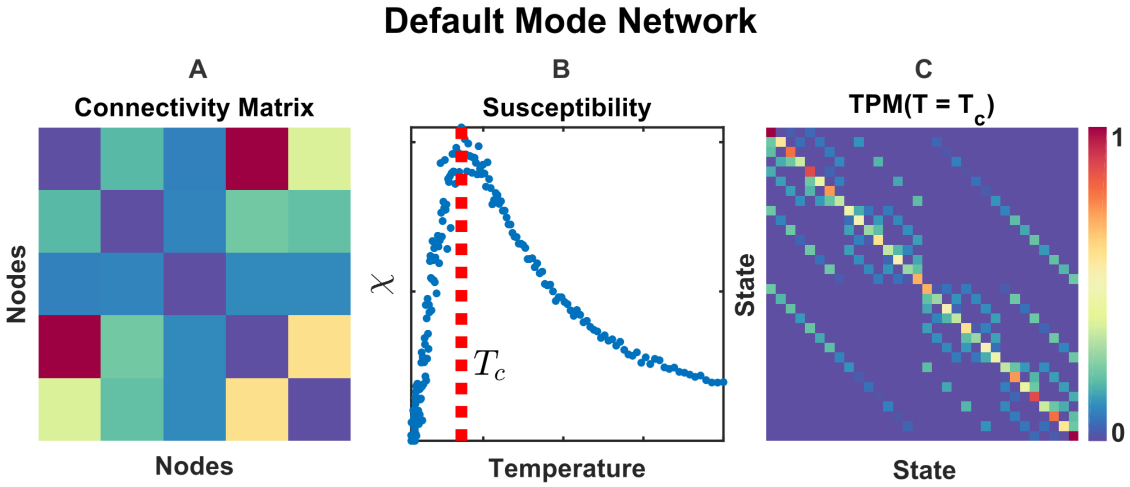

Figure 1.5: **A:** Connectivity matrix J_{DMN} of the default mode network. **B:** Average magnetic susceptibility of this network in the Ising model as a function of temperature is plotted in the middle panel. The critical temperature T_c is marked by the peak of the susceptibility. **C:** The TPM for this system at the critical temperature. These three figures illustrate the basic methodology of the simulations in this thesis. Starting from a connectivity matrix, a simulation on the Ising model is run and the corresponding statistics and transition probabilities are calculated. The TPMs are then fed into the ϕ algorithm to compute the expected integrated information generated by this system as a function of temperature.

be constraint to a state in order to calculate the cause repertoire. We can choose one at random or allow the Ising model to thermalize using the Metropolis algorithm and allow the simulation to pick a state for us. This is what is done when analyzing the Ising model with IIT in chapter 2, but for the sake of simplicity we will simply choose a state at random: $\vec{s}_t = \uparrow, \uparrow, \downarrow, \downarrow, \downarrow$. The mechanism defines which column of the TPM is the cause repertoire where each column in the TPM represents the probability distribution for past states conditioned on the state that it is currently in $p(\vec{s}_{t-1}|\vec{s}_t)$.

When a mechanism is partitioned there will be a number of degenerate states that correspond with that constraint from the perspective of the partitioned mechanism. For example if we partition the system such that nodes $P_A = s_1, s_2$ are contained in partition A and nodes $P_B = s_3, s_4, s_5$ are contained in partition B, the mechanism of partition B no longer constrains the cause repertoire of partition A and vice versa.

Therefore, noise is 'injected' across the partitions where the mechanism of partition A will have $2^3 = 8$ degenerate states associated with it and the mechanism of partition B will have $2^2 = 4$ degenerate states associated with it. This is visualized in Figure 1.6-**i** where all the possible cause repertoires of the first-order mechanism $\vec{s}_{1,t} = \uparrow$ are shown with respect to the original TPM. Since these degenerate states are equivalent from the perspective of the partitioned mechanism, all are treated equally and their cause repertoires are averaged. This is referred to as noising across the partition which is illustrated in Figure 1.6-**ii-iii**.

However, before averaging across these columns in the TPM one final correction must be done. In order to avoid correlations from common inputs, for example if the two nodes in partition A have a common connection to a node in partition B, the mechanisms in partition A must be imposed independently in order to avoid spurious correlations since we are only interested in the cause-information of partition A independently of partition B. This can be dealt with by separating higher-order mechanisms into their constituent mechanisms. For example the mechanism $(s_1, s_2)_t = \uparrow, \uparrow$ can be split into two first-order mechanisms $s_1 = \uparrow, s_2 = \uparrow$. Then the probabilities of the cause-repertoire for the higher-order mechanisms can be calculated as the product of the probabilities of the individual mechanisms (Equation 1.9). Please refer to (Oizumi et al., 2014) for a more detailed explanation of this correction with the introduction of 'virtual elements'.

$$p\left((s_1, s_2)_{t-1} \mid (s_1, s_2)_t = \uparrow, \uparrow\right) = p\left((s_1, s_2)_{t-1} \mid s_1 = \uparrow\right) \times p\left((s_1, s_2)_{t-1} \mid s_2 = \uparrow\right) \tag{1.9}$$

Finally, in order to calculate the probabilities of $p\left((s_1, s_2)_{t-1}\right) \mid (s_1, s_2)_t = \uparrow, \uparrow)$, one must marginalize over the probabilities outside of the purview (which in our simplified algorithm is the same as the mechanism). This marginalization process is accomplished by prescribing a label to each row in the TPM to designate which state

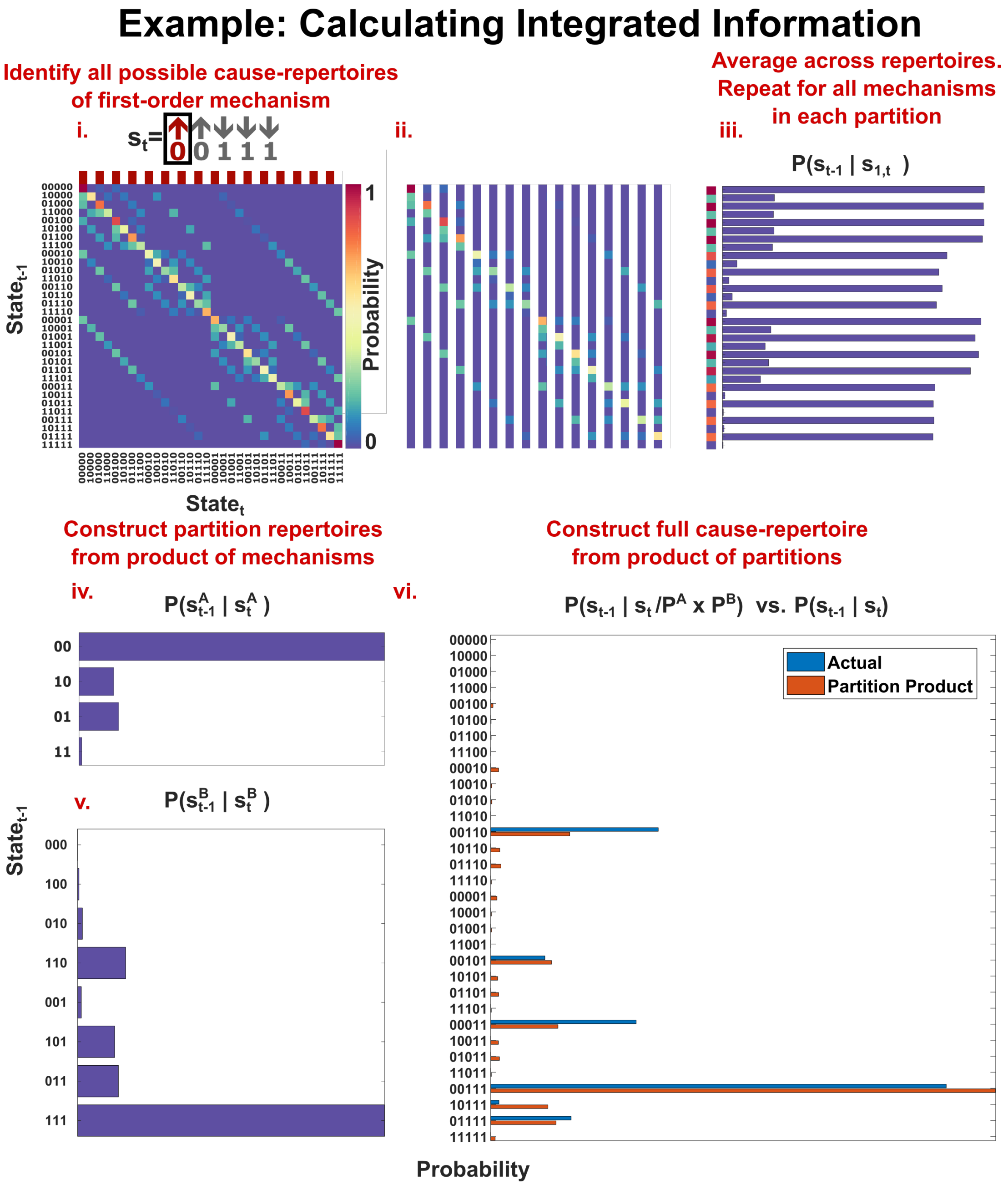

Figure 1.6: **i:** All the possible cause-repertoires corresponding to the first-order mechanism $\vec{s}_t = \uparrow$ are highlighted in red out of all the cause-repertoires in the TPM. **ii:** Since these cause-repertoires are equivalent from the perspective of the mechanism which only constrains a single element, they are all averaged as shown in **iii**. **iv-v:** the repertoires of the partitions are constructed from the repertoires of the first-order mechanisms. **vi:** The full cause-repertoire of the entire system is constructed as the product of the repertoires of the partitions, coloured orange. This is compared to the original cause-repertoire in blue.

it corresponds to from the perspective of the partition. For partition A which is of size 2, there are $2^2 = 4$ possible states. Thus, each of the $2^5 = 32$ states in the full TPM correspond to one of the 4 states in the partition where all states corresponding to the same partitioned state have their probabilities marginalized. At this stage of the calculations, each partition should have its own cause-repertoire generated with respect to the state of that partition. Figure 1.6-**iv-v** illustrates the cause-repertoires corresponding to the partitions in our worked example.

This process of calculating the marginal probabilities conditioned on the mechanisms of the state of the partitioned system can only ever *reduce* the total information we have about the full system. To see this, one only needs to reconstruct the probability distribution of the full system by calculating the product of the probabilities of the two partitions. If the two partitions are in fact causally disconnected, for example if $J_{ij} = 0$ for the nodes across the two partitions, then no information will be lost. If there *are* interactions between the nodes across the partitions, and if these interactions are causally informative, then the act of partitioning will destroy some of the information contained within the probability distributions. Since in this worked example the network we are using is fully connected, partitioning the system will always destroys some information about the past probabilities. Comparing the original cause repertoire of the full system with the one reconstructed from the partitions demonstrates this loss of information as shown in Figure 1.6-**vi**. The differences between these two probability distributions indicates how much integrated information arises from the interaction of the elements across from the partitions. As always, this distance measure between the two probability distributions is only ever defined as integrated information if the partition which gives rise to such loss of information is the MIP.

The metric by which IIT measures this loss of information has changed over the course of its different versions. In older versions the KL-divergence (Equation 1.10) was used to compare the two probability distributions, but more recently in IIT 3.0 the Earth Movers Distance has been employed for reasons outside the scope of this thesis (Oizumi et al., 2014).

$$D_{KL}\left(P \mid\mid Q\right) = -\sum_i P(i) \log \frac{Q(i)}{P(i)} \tag{1.10}$$

Our simplified algorithm employs the simpler KL-divergence to measure the information lost by partitioning the system. Calculating the KL-divergence of this particular example under this particular partition yields a loss of information of $D_{KL}\big(p(s_{t-1} \mid s_t) \mid\mid p(s_{t-1} \mid s_t / P^A \times P^B)\big) = 0.1447$ bits, where the unit of bits comes from the fact that we are using a base of 2 for the logarithms. To calculate integrated information, the partition that yields the minimum information loss must be found, this is the minimum information partition or MIP. The MIP is the partition that maximally reduces the system and represents the most natural way to partition the system into two. Other partitions that are not the MIP can yield losses of information that are high because one has cut across strong causally connected elements. This can lead one to believe that a system is highly integrated when in reality there may exist a partition that can completely reduce the system to (semi-) independent.

An intuitive example would be if one naively looks at the brains of two individuals and treats it as one unified system. From this perspective, the most natural division/-factorization, the MIP, would be the partition that separates the brain of the first individual from the second person since these two systems are not strongly causally connected relative to the strong internal connections in each individual brain. However, if one never finds this partition, one might believe that the two-brain system generates high integrated information simply because the most irreducible partition was not found. Therefore, finding the MIP is crucial in order to maximally reduce the system, and only when the system is maximally reduced can one look to see if it has behaviours that are above and beyond the sum of its components. In the end, this is the property that IIT seeks to measure, the emergence of behaviours that cannot be explained by the sum of the parts. In other words, if a system's behaviours, characterized by its cause/effect probability distributions, can be explained by reducing the system to smaller parts, then the larger system as whole does not intrinsically exist. On the other hand, if all attempts at reducing the system to its parts yield losses of information, then the conclusion is that the interaction of the parts yields behaviours that the sum of the independent components cannot account for, giving rise to integrated information and therefore consciousness.

At the moment of writing this thesis, a brute-force method is employed to find the MIP such that all possible bi-partitions are checked to assess which particular partition yields the minimum information lost. In the intuitive example given above, one can imagine that a smarter algorithm could make better guesses at what the MIP might be instead of simply trying all combinations. Indeed, in the pyPhi library there are some simple methods employed to discount certain combinations of partitions based on the connectivity matrix of the system, however this only takes into account elements that are completely disconnected and is not sensitive to weakly interacting components. In fully connected systems, one needs a method that can distinguish which elements form natural clusters of mutual causal connectivity. Tegmark (2016) suggests methods that threshold the connectivity matrix of interacting systems to better guess the MIP, however these ideas have not been explored rigorously and will be necessary in any serious future work on IIT. Interestingly enough, this factorization process that IIT deems so necessary and fundamental seem analogous to some of the re-scaling methods employed in renormalization group flow. In IIT the factorization is considered at the level of the causal structure whereas with renormalization methods we are normally dealing with transformations of the Hamiltonian to model rescaling of interaction terms. It seems likely that these ideas are in some shape cousins to each other as further suggested by the coarse-graining and black-boxing methods employed to find emergent causal structures (Hoel et al., 2016; Marshall et al., 2016).

1.2.4 Summary

At this point in time, there is no working definition of 'consciousness' that is universally agreed upon by the different domains of science. Each domain interested in the phenomenon has carved out its own definition specific to the tools and understanding that that domain is capable of. While many of these definitions overlap, some contradict and so it is difficult to have a conversation across (or even within) disciplines about what 'consciousness' is, how one should measure it, or how it arises. Integrated Information Theory has presented a relatively formal definition of consciousness starting from phenomenological axioms which in principle should be self-evident truths that all conscious beings should agree upon. The theory formalizes the concept of **intrinsic information** which is notably distinct from the more commonly used Shannon information. Furthermore the theory also formalizes the concept of **integration** which measures how/if the whole is larger than the sum of its parts. By taking the intrinsic nature of information (and simultaneously 'meaning') seriously, a mathematical formulation of consciousness is defined where the major claim is that integrated information *is* consciousness. Integrated information is a complicated function of the causal structure of the system, where causality is defined by assessing all possible perturbations of a system in all possible states across all possible partitions. While the theory has opened many doors in analyzing the integration of systems by way of the suggested algorithms of IIT, there remain many limitations due to the intractability of computing all such perturbations and partitions. This thesis uses the definitions and subsequent algorithms formalized by IIT on very small Ising systems where comprehensive perturbations and partition computations can be done in reasonable time. Integrated information is then calculated for these systems across different model parameters where certain critical parameters of the model compare favourably to empirical brain measurements.

Chapter 2

Modeling the Brain

To model something is to make a set of predictions about its past or future when given some kind if initial information. It is not always obvious which observables in the brain we would like to predict and it is also not obvious what kind of information could even yield such predictions. To model a system as complex as the brain, one has to construct careful questions and combine them with meticulous experimental observations to find the underlying relationships. In this sense, modeling the brain is a task unto itself and requires one to confront some philosophical questions regarding the scale at which one ought to model and the nature of causality. In this chapter we explore some of the questions one faces when attempting to model an intractably complex system such as the brain and ultimately decide to use a generalized Ising model. The model takes information from experimental brain tractography images to map out the underlying networks and then simply simulates Ising transitions and outputs a timeseries. Ultimately, a correlation network is generated from the timeseries which is then compared to empirical brain fMRI images. The control parameter of the model, the temperature is swept to find the temperature at which the model fits best with empirical observations.

2.1 Thermodynamics of the Brain

2.1.1 Reductionism

One of the fundamental assumptions built into most physicists' minds is that the universe must be reducible to its parts. In other words, if you know exactly the microscopic motions of the universe, you could in principle derive how macroscopic order can emerge (i.e. stars, galaxies, planets, life). Reductionism is at the heart of physics and science and without it it is hard to imagine what a self-consistent Theory of Everything would look like. However, there is an explanatory gap between the concept of reductionism and emergence that has yet to be understood in a satisfying way. The nature of this ignorance arises from the difficulty to define causality where many modern measures of 'causality' are mostly measuring statistical correlations or are heuristic approximations (Albantakis et al., 2017; Hoel, 2017; Pearl, 2003). Though it is not in the scope of this thesis to discuss the exact nature/definition of causality, it is important to emphasize the strangeness of emergence as a concept when juxtaposed to the assumption that we live in a reductionist world. Integrated Information Theory offers some hints and suggestions how emergence can be naturally measured or predicted from the reductionist perspective.

2.1.2 Traversing Different Scales

The bridge between the microscopic and macroscopic, or from Newton's laws to thermodynamics, was built slowly over the course of the 19th century with Ludwig Boltzmann largely considered the father of the statistical mechanics movement and ultimately of many branches of modern statistical physics (Sklar, 2015). The macroscopic world is generally accessible to us through our human-scale interactions and macroscopic relationships were estimated heuristically through scientific experiment. The microscopic world is inferred by us through our knowledge of classical mechanics and its relationships were found through mathematical exploration, creativity and derivation. From the scientific process of building a causal bridge between these different scales of reality statistical mechanics was born. I have been intentionally vague about what the macro-observables are. It seems reasonable to assume that a massive number of macro-variables can be generated from even a small number of micro-variables. Why then do we only care about some macro-variables and not others?

2.1.3 The Emergence of Macro-variables and Order Parameters

Perhaps our choice of macro-variable is simply contingent on how much informative power it gives us and evolution naturally picks out the macro-variables sufficient and/or necessary for survival. The complexity of natural language and the words we invent are a testament to the utility and necessity for a large number of macro-descriptions. Abstract concepts like love, happiness, contempt, hunger are as real to us as the microscopic patterns that they are physically composed of. The experience of these macro-variables from our subject point of view is entangled with the notion of emergence; microscopic objects interacting with each other can create macro-objects that are qualitatively different than their constituent parts. This is what is colloquially described as when "the whole is greater than the sum of its parts" and is at the heart of the emergence of complexity and consciousness. These ideas are formalized in IIT and are discussed in the next section. The ideas of IIT have also recently been extended in the context of emergence and rescaling of causal structures (Hoel, 2017), where Figure 2.1 demonstrates an example of the outcome of such a process.

While this emergence does not contradict the reductionist perspective, it demonstrates the concept that it is not always useful or productive to look at the world microscopically. It is often useful, if not necessary, to coarse-grain our observations in order to get anything done. For example if one is trying to design a heat engine, it is

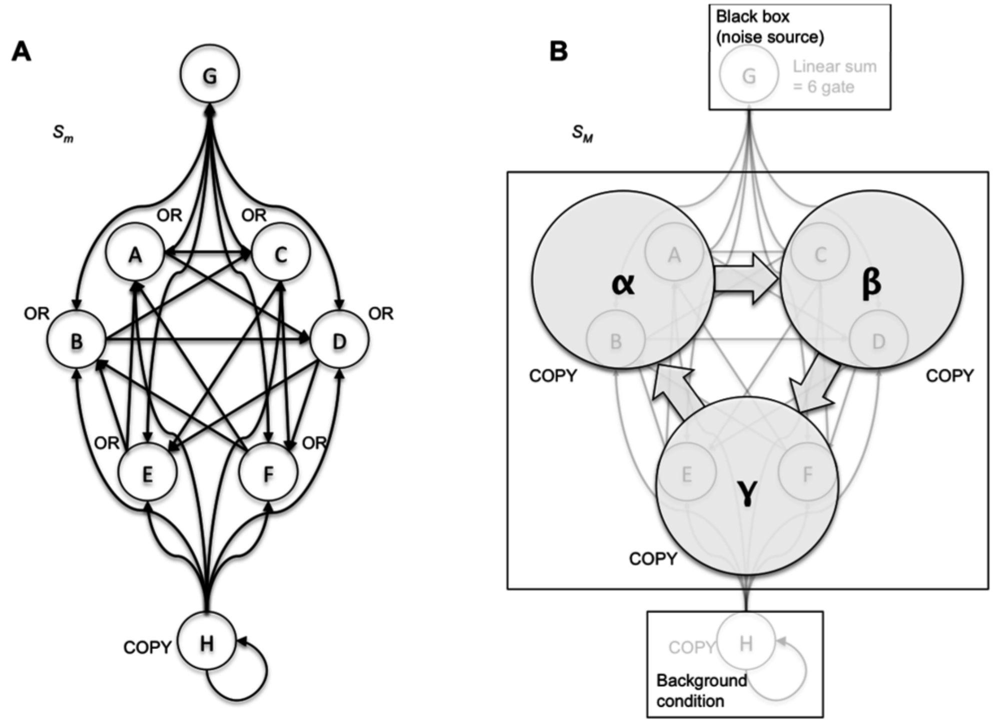

Figure 2.1: Figure is modified from (Hoel, 2017, Fig. 4) as a demonstration of coarse-graining and black-boxing methods that aim to rescale causal structures. Here, the microscopic systems on the left can be causally represented by the macroscopic system on the right.

not necessary to know the exact position and momentum of each particle. Instead, we tend to use macroscopic variables like pressure, temperature, density, volume, energy or work to describe the system and for this thermodynamic context these variables do pretty well in allowing one to make predictions and understanding the system under a variety of real-world conditions. With machine learning, neural networks as one example can be trained to pick out the most natural or useful features of some arbitrary data set. In the realm of physics, recent work by Carrasquilla and Melko (2017) has shown how machine learning can be utilized to teach a neural network to detect order parameters in a simulation.

In fact, the process of machine-learning and particularly unsupervised learning is tantamount to discovering the macro-variables of a system that minimize the uncertainty (or maximize the likelihood) of classifying the state of a system. While the exact strategy that a machine learning algorithm employs to describe a system can vary case by case, the overall strategy of coarse-graining a data-set is something that we do naturally not only in our scientific endeavors but also in our day to day lives. An appreciation of the parallels between the strategies employed by statistical physics, the inverse Ising problem, and machine learning can help us better understand how our coarse-grained perspective of the universe emerges from the microscopic processes that underlie it (Nguyen et al., 2017; Shalizi and Moore, 2003).

2.1.4 The Brain as a Thermodynamic System

Now what happens if we try to understand the brain in the context laid out above? The brain, a complex interconnected system composed of a very large number of similar but importantly diverse set of neurons, is very much like its own state of matter (Tegmark, 2014). Moreover, the microscopic dynamics of the neuronal elements of the brain are not completely mysterious either and have been relatively well studied for some decades, for example with the Hodgkin—Huxley model. The next logical step is to better understand *systems* of these models (Hansel et al., 1993; Hodgkin and Huxley, 1952). So the brain as a complex system lends itself quite readily to the idea that we can treat the brain as a thermodynamic system. While this may indeed be a good analogy, let us begin by considering the ways in which the brain is different to, say, the ideal gas that is commonly used in thermodynamic examples.

For starters, the brain has a much smaller number of ‘particles’. For a box of gas we might have 10^{23} particles of gas whereas the brain ‘only’ has 10^{10} neurons, so the scale of the problem is a few orders of magnitude different. Since most problems in thermodynamics tend to be in the limit as the number of particles goes to infinity, one must be cognizant of the truncated scale of our physical system. Is the size of the system large enough so that it no longer suffers from finite-size effects or is the finite size of the system a fundamental property of its quality? Perhaps its finite size extends its critical point to a critical regime (or Griffiths phase) or perhaps such a critical regime is instead a property of the modular hierarchical organization of brain matter (Moretti and Muñoz, 2013; Muñoz et al., 2010; Rubinov et al., 2011; Wang and Zhou, 2012). If a system like the brain does indeed have some kind of limit as the number of its neuronal components approaches infinity, then it will be important to know at what rate it approaches this limit in order to describe any finite-size effects the brain system might have.

The next main qualitative difference between the classical ensembles in thermodynamics and the neuronal counterpart in the brain is the nature in which the system interacts with itself. Generally in the classical ensembles in statistical mechanics one is dealing with particles that at any one point in time interact only with nearest neighbours where the organization of the matter is either in a well-defined lattice or some kind of homogeneous gas. For good reason, these kinds of systems are simpler to solve and are good models for a wide range of natural phenomena. From these idealizations/simplifications, there is a strong source of symmetry to take advantage of in order to allow one to calculate probabilities and ensemble averages. The brain on the other hand is much less symmetric than these idealizations and is capable of longer range and importantly non-linear interactions. Across human individuals one can easily find important differences in brain architecture and in fact this is a major technical challenge in neuroscience where the brains of different individuals must be normalized to some standard space in order to lend themselves to comparison and analysis (Brett et al., 2002). That being said, the brain still has symmetric properties, the most obvious of which might be the symmetry between the left and right hemispheres. However, even here there seems to be symmetry breaking with regards to the division of functional modules in the brain. For example, the Broca's area of the brain, a region link to language processing (Kennison, 2013) found in the dominant hemisphere (usually left) (Cantalupo and Hopkins, 2001; Van Essen et al., 2011), does not redundantly exist on the opposite hemisphere. While redundancy most definitely exists in the structure of the brain, this symmetry is judiciously broken for certain brain functionality and architecture. An active field of research is involved with accurately imaging and analyzing the structural connectivity of the brain where current research indicates that the brain is organized like a hierarchically modular small-world network (Kaiser, 2011; Meunier et al., 2009, 2010; Sporns, 2010; Sporns et al., 2004). While there may be major sources of symmetry to take advantage of when analyzing the brain, the task is clearly much more nuanced than the idealizations in the classical ensembles where translational and rotational symmetry are usually employed to simplify calculations. This increase in topographic complexity is further compounded by the non-linear and long-range neuronal activation functions.

Another vital detail of the thermodynamic perspective of the brain is the fact that the brain is an open system with matter and energy capable (and necessary) of being exchanged. These kinds of open systems tend not to fit the simpler paradigms of equilibrium thermodynamics and are in need of their own specific treatment. One particular emerging paradigm to understand the organization of such open systems is that of self-organized criticality (SOC) (Bak and Chen, 1991). However, the scale of description from the SOC perspective may not be sufficient for the control of those systems and is likely in need of a finer-grained understanding of these complex, non-linear open systems.

2.1.5 The Need for a Non-equilibrium Theory

The problem of understanding non-equilibrium systems is at the heart of many natural phenomena and indeed life itself. From the energy poured onto Earth from the Sun, the heat and density differentials and winds and ocean currents emerge. Rivers cut valleys into mountains, glaciers scrape lakes as they creep, bacteria metabolizes its free energy and life, complexity, and consciousness are born. While we have made a great many leaps in understanding our natural world through the thermodynamics and statistical mechanics of equilibrium systems, there is yet a new chapter to write on non-equilibrium systems. Perhaps in the same way that thermodynamics revolutionized our understanding of the macroscopic dynamics of the world and the nature of their emergence, a non-equilibrium theory may allow us to understand the nature of diversity and the juxtaposition of chaos and order which seems to permeate all complex phenomena.

Indeed, the role of physics in understanding the brain will be a crucial one in the decades to come. In the same way that statistical mechanics was the bridge that connected microscopic physics to macroscopic observables, a new theory must emerge to build the bridge between the neuronal micro-scale to the macro-scale of consciousness. In fact if one takes seriously the task of deriving the emergence of macroscopic variables in the brain one must make some kind of judicious choice in their macro-variable. From the subjective phenomenological perspective our thoughts and experiences and qualia are the most natural macroscopic observables, however these observables are only subjectively accessible and do not lend themselves easily to the rigors of the science method. Instead we might be forced to set the bar a bit lower and settle for mesoscopic observables that can come from brain imaging. These mesoscopic variables will ultimately be treated like an order parameter and used to describe the state of the system in some way. For example, analogous to the case of the Kuramoto model (Acebrón et al., 2005; Kuramoto and Araki, 1975) the global synchrony of the brain could act as an order parameter if this order parameter was useful for describing the condition of the brain. In this regard, the Integrated Information Theory of consciousness describes a new measure that aims to do exactly this.

2.2 The Ising Model

From the middle of the 20th century onwards, the Ising model has been the work horse of the statistical mechanics world. Due in large part to its initial simplicity it has been retro-fitted to model a plethora of phenomenon. In this project, the Ising model is used extensively in computer simulation models of the brain. This section will give a brief overview of the history of the Ising model, its ubiquity in the scientific world, its contributions to interdisciplinary science, and its contributions to the philosophy of science.

2.2.1 A Brief Overview of the Ising Model

The Ising model was originally introduced by Wilhelm Lenz (1920) and given to his PhD student Ernst Ising to be solved (Brush, 1967). The goal was to discover phase transitions in a simple mathematical model of ferromagnetism. The original version of the model was a 1 dimensional network of nodes which represented atomic spins. Each node is connected to its neighbours. Each node could be in one of two possible states, spin up or spin down. The Hamiltonian of the system (Equation 2.1) is defined by the interaction terms between the spins with coupling J_{ij} between nodes i and j and the applied field h_i acting on the nodes i. In the classical version of the model, the interactions between spins are restricted to nearest-neighbour interactions which is denoted in the Hamiltonian sum as $\langle ij \rangle$. In one dimension there are only ever 2 nearest-neighbours, in the two dimensional square lattice there are 4 and so forth.

$$H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j - \sum_i h_i s_i \tag{2.1}$$

At the end of his PhD in 1924, Ising had solved the 1 dimensional case of the model and had found that no phase transitions exist in the model (Ising, 1925). He erroneously extrapolated his results to higher dimensions claiming that no phase transitions exist above the 1 dimensional case also. It would take another 20 years until Onsager rectified this mistake and solved the 2D case of the model, discovering that the model does in fact exhibit a phase transition. This was a very powerful discovery for the world of solid state physics and statistical mechanics. It demonstrated that a seemingly simple mathematical model that could abstractly represent the physical world could exhibit a rich and diverse set of behaviours that where far from trivial to derive (Brush, 1967). The innocuous Hamiltonian of the model has shown to be capable of demonstrating behaviours that pierce into the heart of the physics of co-operation, many-bodied systems, and complexity in general. Since Onsager's results, attempts to solve the Ising model for higher dimensions or for finite systems have not yet succeeded. For larger than 4 dimensions, a mean field theory of the Ising model is capable of describing the system, but the 3 dimensional Ising model remains an enigma that is thought to be analytically intractable (Taroni, 2015).

This project numerically simulates the Ising model using Monte Carlo Metropolis methods. The transition probability for a spin to flip is given by the Boltzmann factor:

$$p = e^{\frac{-\triangle E}{k_B T}} \text{ if } \triangle E < 0 \tag{2.2}$$

where $\triangle E$ is the change in energy if the spin were to flip, k_B is the Boltzmann factor which we can set to unity, and T is the temperature of a connected heat bath which we can use as our control parameter. Each iteration a random node is picked, a random number between 0 and 1 is rolled and compared to the transition probability. If the random number is less than the probability of a transition then the spin will flip. If $\triangle E \geq 0$ then $p = 1$. The system is thermalized by running the simulation for hundreds of iterations (500). Statistical measures are accumulated by observing the system for thousands of iterations (2000). These methods were tested in the undergraduate thesis precursor to this project Khajehabdollahi (2015).

2.2.2 Criticality & Universality

The utility of the Ising model to abstractly represent a multitude of different systems hints at a deep fundamental relationship between the phenomena that it can model. The nodes of the model, originally representing atomic spins, where eventually generalized to model molecular species, lattice gases, social behaviour, political affiliation, economic modeling (Bourgine and Nadal, 2013; Zhou and Sornette, 2007) (see references within), and more recently, neurological modeling (Fraiman et al., 2009; Haimovici et al., 2013; Marinazzo et al., 2014). The model is essentially the simplest (maximum entropy) way to represent some kind of causal network (Mora and Bialek, 2011) so naturally it has diffused and mutated into all reaches of scientific endeavor. To what extent does the abstraction and generalization of causal systems help in understanding these systems? What about predicting them? Is there even a distinction? The answers to these questions seem to lie somewhere murky, near the concepts of universality, criticality, and phase transitions.

Phase Transitions

A phase transition is a complex concept to nail down and define. The concept of a phase transition is a mix of phenomenology and mathematics. A system is said to undergoes a phase transition when certain defining properties of it change discontinuously or drastically. For example, when water freezes, it obtains a completely new property it did not have before, rigidity. As (relatively) macroscopic beings, we observe the macroscopic phenomena of phase transitions quite commonly. When water melts or freezes, when the roads are filled with cars during rush hour and then empty again at night, when stock markets crash, mass migrations occur, systems collapse of entire societies, all these phenomena exhibit the characteristic nature that seemingly small quantitative changes result in massive qualitative changes. This characteristic of our complex world has not gone unnoticed. Karl Marx for example made this observation at the socioeconomic scale: “quantitative changes turn into qualitative ones” (Anderson et al., 1972).

A phase transition can roughly be thought as the transition between different qualities of organization. More generally, one side of a phase transition, its sub-critical regime, can be thought of as an ordered state, and the other side, the super-critical regime, can be thought of as a disordered state. The phase transition itself lies at the critical point somewhere in between (Binney et al., 1992; Stanley, 1971). Generally, there exists an entire taxonomy of the possible modes of order and the phase transitions that lead to them. (This phenomenon can be observed in a plethora of systems, for example in hydrodynamic systems which readily exhibit a rich variety of phase transitions, for example in flow-phase diagram of Taylor-Couette flow. Small quantitative changes in the Reynolds numbers of the apparatus give yield to a rich variety of flow structures (Andereck et al., 1986; Grossmann et al., 2016)). As the symmetries in the system are broken towards the sub-critical regime, more and more complex forms of matter and organization may emerge. Conversely, in the direction of the super-critical regime, quantitative differences in the system which, in the sub-critical regime would separate into different ordered states, would coalesce and merge into symmetry. This concept is not foreign to cosmologists and particle physicists that expect a grand unification regime in the early stages of the universe, where the currently distinct forces are expected to become unified (Chaisson, 2001; Layzer, 1991).

Modular Hierarchy of Knowledge

To this end, one can look to the modular and hierarchical organization of the sciences to further witness the emergence of new order and quality from fluctuations in quantity. Why is it that chemistry is so qualitatively different from physics to the point that the tools and rules used to understand chemistry are qualitatively different than those used to understand physics? Is chemistry not the logical conclusion of physics? As is biology the natural extension of chemistry, followed by psychology, sociology, and economics and politics. Why is our understanding of the world so compartmentalized when science tells us that it should all be inter-related? Surely if one has knowledge of the microscopic, the macroscopic conclusions should follow, albeit with some work. Statistical mechanics at the turn of the century was a resounding success in confirming that macroscopic phenomena can, in principle, be deduced from the microscopic. Recently, a new paradigm is taking hold to try to explain the hierarchical organization of, not just science, but the entire universe (Morowitz, 2004; Smith and Morowitz, 2016). These ideas revolve around the notion that our universe, from the big bang, to the formation of its stars, galaxies, planetary systems, Earth, life, and consciousness has been a cascade of phase transitions, each bifurcating our universe into a new realm of complexity with new modalities of description. To quote Edward Robert Harrison, "Hydrogen is a light, odorless gas, which, given enough time, turns into people."

Criticality

There is more to the phase transition than the order (or disorder) on the other side. The transition has shown to be an interesting regime in and of itself and is where many disparate systems converge to universal behaviours across essentially all orders of magnitude (Stanley, 1999; Watkins et al., 2016). At this critical point, the microscopic features of the system become less important. It enters a behavioural regime that is less contingent on the microscopic causal nature of the system but rather something more invariant. This is observed in systems whose microscopic structures are different but will still converge to the same class of systems at their critical points. Another defining property of criticality is its self-similar nature. Critical systems have no specific characteristic length. Like a fractal, there are structures at all scales. Unlike the ideal fractal, in nature these structures usually have some sort of cut-off and are not perfectly self-similar (breaking more symmetries). While nature might not be ideally fractal as a result of these cut-offs and broken symmetries, it can still be self-similar for a broad regime. The convergences of these broad phenomena to similar modes of organization can be seen when a microscopic system is rescaled into a macroscopic set of variables.

Renormalization Group Flow

Renormalization group flow explains this phenomenon by making a distinction between relevant, irrelevant, and marginal variables. In general, all the variables needed to describe a microscopic system are not required to describe a macroscopic one. Relevant observables increase in magnitude with the scale of the system, irrelevant observables decrease with scale, and some observables may change non-monotonically (Wilson, 1975). Renormalization group flow demonstrated how one can traverse the scales of a system and observe the flow of the observables of interest. This powerful technique was able to explain how universal behaviours can emerge by demonstrating how most microscopic observables become irrelevant leaving just a few variables to describe the macroscopic system. The scaling relations of these variables are defined succinctly by a set of critical exponents. Close to the critical point, variables like the order parameter (an abstract variable that represents some form of order in the system, like Magnetization, or Integrated Information), the susceptibility of the system to some driving force, the specific heat, etc., are described simply by their distance to the critical point and a scaling exponent. The convergence of the critical exponents of a variety of disparate systems demonstrates this property of universality.

While universality as a phenomenon gives hope that science can keep complex, many-body problems tractable, there is still a diversity of universality classes to understand as well. Not all critical systems converge to the same critical exponents and this requires a taxonomy of the variety of universality classes. However, if such a catalog can be made then there may be hope that complex systems such as the neuronal dynamics of the brain or global financial dynamics can be understood in the paradigm of some universality class. While this paradigm may still not allow precise predictions of micro-variables, it may in principle allow us to make accurate forecasts of the macro-variables, the phases, and impending transitions these systems may face.

One particular vein of this frontier is the process of renormalization on networks (Gandhi; Newman and Watts, 1999; Rozenfeld et al., 2010). Combined with Monte Carlo renormalization group methods (Pawley et al., 1984; Swendsen, 1979) these techniques may prove to be not just useful, but necessary to tackle some of the intractable problems and scales of complex systems like the ecosystems, the brain, society, etc.

Properties of Critical Systems

Critical systems essentially lie in a Goldilocks zone between two (or more) regimes dominated by different forces (Crutchfield, 2012; Stanley, 1971). In the Ising model this is characterized by the integrating forces of the interactions that spins have with their neighbours opposing the random perturbations of the heat bath it lies in contact with. As the temperature approaches zero (sub-critical), the interaction forces become so strong relative to the fluctuations that the system falls into a (meta-)stable minimum energy state without much probability of fluctuating. Conversely, as the temperature approaches infinity, the random fluctuations become so dominant that any interaction forces between elements in the model are completely drowned out by the noise induced by the heat bath. Between these two extremes there is a sweet spot where the integrative nature of the interaction forces and the segregating nature of the heat bath are balanced in such a way to give rise to patterns and organization that neither extreme is capable of. The Ising model at criticality has a scale-invariance in the size of the structures that it is composed of. There is no preferential scale of the model at criticality and this concept is fundamental to the definition of criticality in general. Critical systems tend to also have very strong, long-range correlations characterized by a diverging correlation length. If interaction forces are too high, the system is too constraint to fluctuate and allow for strong perturbations in time. Conversely, if the temperature is too high the system is too random for any perturbative information to travel far either in space or time. However, critical systems are highly susceptible to perturbative forces as their perturbations tends to persist in both time and space as integration ensures that information can travel far and segregations ensures that the perturbations can actually occur in the first place. The maximization of susceptibility gives rise to a high dynamic range, information capacity, and information fidelity (due to error-correction) (Beggs, 2008; Hesse and Gross, 2014; Shew and Plenz, 2013). The strong, long-range correlations that persist in critical systems give rise to redundancy in the way information propagates throughout the network. Thus, if a propagating perturbation is randomly mutated by the heat bath, essentially adding error to the information being propagated, the redundantly correlated network can correct such errors as other information channels can average out any errors that accumulate. In critical dynamical systems the divergence of the system's susceptibility that manifests itself temporally defines the concept of critical slowing down. In ecological systems that are undergoing critical slowing down, perturbations of the system will not dissipate on time scales corresponding to the characteristic cycles that exist in that system and may ultimately force an ecological system to collapse or transition into a new phase (Scheffer et al., 2015). Therefore it is crucial to develop our understanding of critical systems in order to be cognizant of the dynamic phases our natural environment may reside in, be it in neurological, financial, or ecological systems.

Detecting Criticality with Power Laws

One hallmark of critical systems, and indeed of many complex systems in general, are 'heavy-tailed' distributions for a number of variables of the system (Watkins et al., 2016). As opposed to something like the normal distribution, heavy-tailed distributions tend to have much larger probabilities for 'rarer' events. For example if one was to naively use a normal distribution to predict the frequency that different size earthquakes would occur, one would underestimate the probabilities associated with very large or very small events. There exist many different families of heavy-tailed distributions and distinguishing between them is often a tricky but necessary task for understanding the causal origins (Buzsáki and Mizuseki, 2014). Power law distributions are observed in a wide variety of phenomena from the famous Gutenberg-Richter law that relates the total number of occurrences of an earthquake given a minimum intensity (Gutenberg and Richter, 1954), to Zipf's law for natural language that states the frequency of a word is inversely related to the frequency rank of that word (Zipf, 1935), or the Pareto distribution which was originally used to describe wealth distribution in society (Pareto, 1964). These observations extend to a variety of different systems such as the population distribution of cities (Reed and Jorgensen, 2004), area damaged by forest fires (Schoenberg et al., 2003), stock market fluctuations (Coronel-Brizio and Hernandez-Montoya, 2005), or starquakes and plasma instabilities in astro/plasma-physics (Bak, 1996). The ubiquity of power law distributions arises from the concepts discussed earlier involving scale-invariance. In scale-invariant systems events of all sizes can occur as the system has no characteristic size where this property manifests itself as power law distributions in a set of defining observables for that system (Bak, 2013).

While it is erroneous to assume that all power law distributions are a result of critical phenomena, it is quite natural for critical phenomena which by definition have scale-invariant properties to have observables obeying power law distributions. This non-mutual relationship has for decades been a source of contention as many researchers simply tend to look for power laws in order to make the claim that a process is indeed critical (Beggs and Timme, 2012; Watkins et al., 2016), and while this may usually be true and an efficient technique to quickly assess whether or not a system has the capacity to be critical, there can be false positives so one must tread carefully in these regards.

2.2.3 Self-organized Criticality in Nature

The ubiquity of power laws and critical systems poised at the cusp of different phases is a worthwhile mystery in and of itself. From the outset, to create a critical system a control parameter of some kind generally needs to be fine-tuned to a critical point until the system is constraint to be in a critical state. In the Ising model the temperature T acts as this control parameter (though this parameter can be generalized/extended by combining the temperature and interaction weights J_{ij} into a new parameter K). How is it that so many systems tend to be critical (Bourgine and Nadal, 2013; Hesse and Gross, 2014; Scheffer et al., 2015; Schwab et al., 2014; Tagliazucchi et al., 2012; Taroni, 2015; Zhou and Sornette, 2007)? How do these systems get tuned to criticality and why do they stay there? Bak and Chen (1991) coined the term self-organized criticality to encapsulate the idea that many systems become critical by self-organizing themselves in an evolutionary sense to arrive at such a critical point. Using simple sandpile, earthquake or domino models as illustrative examples, this succinct perspective on the organization of the universe has made reverberations in the understanding of complex, many-body systems. The idea hopes to explain the dynamics of these large interacting systems all the way from the genetic scale up to ecological, geological, and astrophysical systems. Much like the self-similar systems it was meant to describe, the concept of self-organized criticality has created a catastrophic avalanche of ideas, good ones, bad ones and controversial ones, which in all likelihood will correlate and reverberate well into the future of humanity and the sciences as it diffuses into all disciplines and regimes of existence.

2.2.4 The Critical Brain

With the realization that essentially all complex systems of energetically interacting elements seem to self-organize to criticality, the observation that the brain also exhibits critical properties might then be no surprise (Beggs, 2008; Beggs and Plenz, 2003; Brochini et al., 2016; Chialvo, 2004; de Arcangelis and Herrmann, 2010; Expert et al., 2011; Hesse and Gross, 2014; Moretti and Muñoz, 2013; Tagliazucchi et al., 2012; Timme et al., 2016). Critical brains seem to be the ideal candidate for a learning (Carrasquilla and Melko, 2017), adaptive (Hidalgo et al., 2014), dynamic system that is capable of a wide range of behaviours to perturbations in its environment. The observations of power laws and critical behaviors in a variety of measurements of the brain in conjunction with what is already known about critical systems hints at a deep underlying relationship between the physics/mathematics of phase transitions and computation and evolution.

2.2.5 The Generalized Ising Model of the Brain

In this project, the classical 2 dimensional Ising model is generalized to fully connected graphs. The connectivity matrix $\mathbf{J}$ composed of elements J_{ij} defines the interactions between nodes i and j. This matrix, or equivalently, this graph, is largely what defines the type of Ising model we are interested in. Modifications of the connectivity can transform the classical Ising model of ferromagnetism into, for example, a model on social segregation (Bourgine and Nadal, 2013; Zhou and Sornette, 2007). In order to contextualize the model to the brain, this project uses connectivity matrices that are inspired by and have been imaged using diffusion tensor imaging (DTI) of the human brain in the Human Connectome Project (Andersson and Sotiropoulos, 2015, 2016; Andersson et al., 2003; Fischl, 2012; Glasser and Van Essen, 2011; Glasser et al., 2013; Jenkinson et al., 2002, 2012; Van Essen et al., 2011).

Apart from this distinction, the Generalized Ising Model of the Brain is not very different from the classical version. It extends the concept of nearest neighbour interactions to a global level where every element is connected to every other element with weight J_{ij}. These weights, in the case of the human connectome, are acquired through Diffusion Tensor Imaging (DTI) of the brain which roughly maps the white matter tracts in the brain. Each weight J_{ij} therefore represents the number of tracts that connect region i to region j. In this project, the raw voxel-space images of the full brain human connectome are parcellated into 84 labeled regions (see Appendix A) using FSL, Freesurfer and MRTrix. For the simulations in this project pertaining to Integrated Information Theory a further reduction scheme is utilized in order to ensure that IIT algorithms will halt in reasonable time. This reduction scheme, explained in more detail in chapter 2 simply separates the brain into 9 different resting state networks (RSNS) representing each module independently using just 5 nodes. Each of these networks represents, crudely, one of the RSNs of the brain (Auditory, Default Mode, Executive Control L/R, Salience, Sensorimotor, and Visual Lateral/-Medial/Occiptal). As computational power increases and IIT algorithms improve, future work may extend such methods to larger networks that more accurately represent the active networks in the brain.

2.2.6 Summary

The development of the Ising model started off as a mathematical exercies to model the phenomenon of phase transitions. While Ernst Ising's work was ultimately unable to model phase transitions in 1 dimension, his contribution triggered an avalanche of investigative research on the model that resulted in a new found appreciation for the complexity that could arise from such simplicity. Though the Ising model can be thought of as one of the most well-studied models in physics, it continues to this day to shine new light on the most complex systems in our universe. It has pushed the boundaries of physics, recruited the help of mathematicians, piqued the interest of sociologists, economists, and now computational neuroscientists, and tied it all together with the ubiquitous concepts of universality and criticality. What was originally a model aimed to discover the microscopic origins of magnetism has carved out a path to discover the nature of emergence and in doing so has triggered a chain of events that has been entangling and integrating the entire tree of human knowledge.

2.3 Thesis Statement and Contribution

In order to explore the utility of the Integrated Information Theory of Consciousness and to gauge its ability to describe consciousness this thesis connects IIT to empirical brain dynamics by proxy of the Generalized Ising Model of the Brain. As it stands, IIT's main difficulties lie in the fact that the computations in the theory remain intractable for large systems and therefore it is hard to apply and therefore test the theory to empirical data from the brain. Our novel approach to bring together the simple Ising model of the critical brain and analyze it with the tools and machinery that arise with IIT is, to our knowledge, the first such attempt to apply IIT's tools on critical systems resembling the brain. The journey of this research project has illuminated the notion that the emergence of complexity and criticality are ubiquitous/universal phenomenon and that the concepts in the physics of statistical mechanics, complexity, and consciousness are all inter-related such that understanding any of them in any deep way requires the integration of all these branches of knowledge. This project adds further support to the hypothesis that self-organization towards criticality is an evolutionary attractor and the emergence of conscious systems are an inevitable conclusion of matter entropically self-organizing in a complex world.

Bibliography

Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., and Spigler, R. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137, 2005.

Albantakis, L., Marshall, W., Hoel, E., and Tononi, G. What caused what? An irreducible account of actual causation. *arXiv preprint arXiv:1708.06716*, 2017.

Alkire, M. T., Hudetz, A. G., and Tononi, G. Consciousness and anesthesia. *Science*, 322(5903):876–880, 2008.

Andereck, C. D., Liu, S., and Swinney, H. L. Flow regimes in a circular Couette system with independently rotating cylinders. *Journal of Fluid Mechanics*, 164: 155–183, 1986.

Anderson, P. W. et al. More is different. *Science*, 177(4047):393–396, 1972.

Andersson, J. L. and Sotiropoulos, S. N. Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussian processes. *Neuroimage*, 122:166–176, 2015.

Andersson, J. L. and Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*, 125:1063–1078, 2016.

Andersson, J. L., Skare, S., and Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003.

Bak, P. Earthquakes, Starquakes, and Solar Flares. In *How Nature Works*, pages 85–104. Springer, 1996.

Bak, P. *How nature works: the science of self-organized criticality.* Springer Science & Business Media, 2013.

Bak, P. and Chen, K. Self-organized criticality. *Scientific American*, 264(1):46–53, 1991.

Beggs, J. M. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 366(1864):329–343, 2008.

Beggs, J. M. and Plenz, D. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003.

Beggs, J. M. and Timme, N. Being critical of criticality in the brain. *Frontiers in physiology*, 3, 2012.

Binney, J. J., Dowrick, N. J., Fisher, A. J., and Newman, M. *The theory of critical phenomena: an introduction to the renormalization group.* Oxford University Press, Inc., 1992.

Bourgine, P. and Nadal, J.-P. *Cognitive economics: an interdisciplinary approach.* Springer Science & Business Media, 2013.

Brett, M., Johnsrude, I. S., and Owen, A. M. The problem of functional localization in the human brain. *Nature reviews neuroscience*, 3(3):243–249, 2002.

Brochini, L., de Andrade Costa, A., Abadi, M., Roque, A. C., Stolfi, J., and Kinouchi, O. Phase transitions and self-organized criticality in networks of stochastic spiking neurons. *Scientific reports*, 6:35831, 2016.

Brush, S. G. History of the Lenz-Ising model. *Reviews of modern physics*, 39(4):883, 1967.

Buzsáki, G. and Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, 15(4):264–278, 2014.

Cantalupo, C. and Hopkins, W. D. Asymmetric Broca's area in great apes. *Nature*, 414(6863):505–505, 2001.

Carrasquilla, J. and Melko, R. G. Machine learning phases of matter. *Nature Physics*, 2017.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine*, 5(198):198ra105–198ra105, 2013.

Chaisson, E. Cosmic Evolution: The Rise of Complexity in Nature. Cambridge; London: Harvard University Press. 2001.

Chialvo, D. R. Critical brain networks. *Physica A: Statistical Mechanics and its Applications*, 340(4):756–765, 2004.

Coronel-Brizio, H. and Hernandez-Montoya, A. On fitting the Pareto–Levy distribution to stock market index data: Selecting a suitable cutoff value. *Physica A: Statistical Mechanics and its Applications*, 354:437–449, 2005.

Crutchfield, J. P. Between order and chaos. *Nature Physics*, 8(1):17–24, 2012.

de Arcangelis, L. and Herrmann, H. J. Learning as a phenomenon occurring in a critical state. *Proceedings of the National Academy of Sciences*, 107(9):3977–3981, 2010.

Expert, P., Lambiotte, R., Chialvo, D. R., Christensen, K., Jensen, H. J., Sharp, D. J., and Turkheimer, F. Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *Journal of The Royal Society Interface*, 8(57):472–479, 2011.

Fischl, B. FreeSurfer. *Neuroimage*, 62(2):774–781, 2012.

Fraiman, D., Balenzuela, P., Foss, J., and Chialvo, D. R. Ising-like dynamics in large-scale functional brain networks. *Physical Review E*, 79(6):061922, 2009.

Gandhi, S. Renormalization group on complex networks.

Glasser, M. F. and Van Essen, D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of Neuroscience*, 31(32):11597–11616, 2011.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, 80:105–124, 2013.

Goldenfeld, N. and Woese, C. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.*, 2(1):375–399, 2011.

Grossmann, S., Lohse, D., and Sun, C. High–Reynolds number Taylor-Couette turbulence. *Annual review of fluid mechanics*, 48:53–80, 2016.

Gutenberg, B. and Richter, C. F. Frequency and energy of earthquakes. *Seismicity of the Earth and Associated Phenomena*, pages 17–19, 1954.

Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical review letters*, 110(17):178101, 2013.

Hansel, D., Mato, G., and Meunier, C. Phase dynamics for weakly coupled Hodgkin-Huxley neurons. *EPL (Europhysics Letters)*, 23(5):367, 1993.

Harari, Y. N. and Perkins, D. *Sapiens: A brief history of humankind.* Harvill Secker London, 2014.

Hesse, J. and Gross, T. Self-organized criticality as a fundamental property of neural systems. *Frontiers in systems neuroscience*, 8, 2014.

Hidalgo, J., Grilli, J., Suweis, S., Muñoz, M. A., Banavar, J. R., and Maritan, A. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, 2014.

Hodgkin, A. L. and Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

Hoel, E. P. When the map is better than the territory. *Entropy*, 19(5):188, 2017.

Hoel, E. P., Albantakis, L., Marshall, W., and Tononi, G. Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neuroscience of Consciousness*, 2016(1):niw012, 2016.

Hopfield, J. Physics, computation, and why biology looks so different. *Journal of Theoretical Biology*, 171(1):53–60, 1994.

Ising, E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. Fsl. *Neuroimage*, 62(2):782–790, 2012.

Kaiser, M. A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage*, 57(3):892–907, 2011.

Kennison, S. M. *Introduction to language development.* Sage Publications, 2013.

Khajehabdollahi, S. Consciousness, Integrated Information Theory, and the Ising Model: an Investigation of the Critical Behavior of the Brain. Undergraduate thesis, University of Western Ontario, Ontario, Canada, 2015.

Kitazono, J., Kanai, R., and Oizumi, M. Efficient Algorithms for Searching the

Minimum Information Partition in Integrated Information Theory. *ArXiv e-prints*, December 2017.

Kuramoto, Y. and Araki, H. Lecture notes in physics, international symposium on mathematical problems in theoretical physics. 1975.

Layzer, D. *Cosmogenesis: the Growth of Order in the Universe*. Oxford University Press on Demand, 1991.

Marinazzo, D., Pellicoro, M., Wu, G., Angelini, L., Cortés, J. M., and Stramaglia, S. Information transfer and criticality in the ising model on the human connectome. *PloS one*, 9(4):e93616, 2014.

Marshall, W., Albantakis, L., and Tononi, G. Black-boxing and cause-effect power. *arXiv preprint arXiv:1608.03461*, 2016.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., and Bullmore, E. T. Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, 3, 2009.

Meunier, D., Lambiotte, R., and Bullmore, E. T. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4, 2010.

Mora, T. and Bialek, W. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

Moretti, P. and Muñoz, M. A. Griffiths phases and the stretching of criticality in brain networks. *arXiv preprint arXiv:1308.6661*, 2013.

Morowitz, H. J. *The emergence of everything: How the world became complex*. Oxford University Press, USA, 2004.

Muñoz, M. A., Juhász, R., Castellano, C., and Ódor, G. Griffiths phases on complex networks. *Physical review letters*, 105(12):128701, 2010.

Newman, M. E. and Watts, D. J. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.

Nguyen, H. C., Zecchina, R., and Berg, J. Inverse statistical problems: from the inverse Ising problem to data science. *arXiv preprint arXiv:1702.01522*, 2017.

Oizumi, M., Albantakis, L., and Tononi, G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol*, 10 (5):e1003588, 2014.

Pareto, V. *Cours d'économie politique*, volume 1. Librairie Droz, 1964.

Pawley, G., Swendsen, R., Wallace, D., and Wilson, K. Monte Carlo renormalization-group calculations of critical behavior in the simple-cubic Ising model. *Physical Review B*, 29(7):4030, 1984.

Pearl, J. Causality: models, reasoning and inference. *Econometric Theory*, 19(675-685):46, 2003.

Reed, W. J. and Jorgensen, M. The double Pareto-lognormal distributiona new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753, 2004.

Rozenfeld, H. D., Song, C., and Makse, H. A. Small-world to fractal transition in complex networks: a renormalization group approach. *Physical review letters*, 104 (2):025701, 2010.

Rubinov, M., Sporns, O., Thivierge, J.-P., and Breakspear, M. Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS computational biology*, 7(6):e1002038, 2011.

Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., Rosanova, M., Casali, A. G., Brichant, J.-F., Boveroux, P., et al. Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Current Biology*, 25(23):3099–3105, 2015.

Scheffer, M., Carpenter, S. R., Dakos, V., and van Nes, E. H. Generic indicators of ecological resilience: Inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics*, 46:145–167, 2015.

Schindler, S. Global Mind Change.

Schoenberg, F. P., Peng, R., and Woods, J. On the distribution of wildfire sizes. *Environmetrics*, 14(6):583–592, 2003.

Schwab, D. J., Nemenman, I., and Mehta, P. Zipfs law and criticality in multivariate data without fine-tuning. *Physical review letters*, 113(6):068102, 2014.

Shalizi, C. R. and Moore, C. What is a macrostate? Subjective observations and objective dynamics. *arXiv preprint cond-mat/0303625*, 2003.

Shannon, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

Shew, W. L. and Plenz, D. The functional benefits of criticality in the cortex. *The neuroscientist*, 19(1):88–100, 2013.

Sklar, L. Philosophy of Statistical Mechanics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2015 edition, 2015.

Smith, E. and Morowitz, H. J. *The origin and nature of life on earth: the emergence of the fourth geosphere*. Cambridge University Press, 2016.

Sporns, O. *Networks of the Brain*. MIT press, 2010.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. Organization, development and function of complex brain networks. *Trends in cognitive sciences*, 8(9):418–425, 2004.

Stanley, H. E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Reviews of modern physics*, 71(2):S358, 1999.

Stanley, H. Introduction to Phase Transitions and Critical Phenomena Oxford Univ, 1971.

Swendsen, R. H. Monte Carlo renormalization group. *Physical Review Letters*, 42 (14):859, 1979.

Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Frontiers in physiology*, 3, 2012.

Taroni, A. Statistical physics: 90 years of the Ising model. *Nature Physics*, 11(12): 997–997, 2015.

Tegmark, M. Consciousness is a state of matter, like a solid or gas. *New Scientist*, 222(2964):28–31, 2014.

Tegmark, M. Improved measures of integrated information. *PLoS computational biology*, 12(11):e1005123, 2016.

Timme, N. M., Marshall, N. J., Bennett, N., Ripp, M., Lautzenhiser, E., and Beggs, J. M. Criticality maximizes complexity in neural tissue. *Frontiers in Physiology*, 7, 2016.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., and Coalson, T. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex*, 22(10):2241–2262, 2011.

Wang, S.-J. and Zhou, C. Hierarchical modular structure enhances the robustness of self-organized criticality in neural networks. *New Journal of Physics*, 14(2): 023005, 2012.

Watkins, N. W., Pruessner, G., Chapman, S. C., Crosby, N. B., and Jensen, H. J. 25 years of self-organized criticality: concepts and controversies. *Space Science Reviews*, 198(1-4):3–44, 2016.

Wilson, K. G. The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47(4):773, 1975.

Zhou, W.-X. and Sornette, D. Self-organizing Ising model of financial markets. *The European Physical Journal B*, 55(2):175–181, 2007.

Zipf, G. K. The psychology of language. *NY Houghton-Mifflin*, 1935.

Chapter 3

The Emergence of Integrated Information, Complexity, and Consciousness at Criticality

A growing body of evidence in the past few decades has emerged suggesting that many disparate natural and particularly biological phenomena reside in a critical regime of dynamics on the cusp between order and disorder (Beggs, 2008; Beggs and Plenz, 2003; Brochini et al., 2016; Crutchfield, 2012; de Arcangelis and Herrmann, 2010; Expert et al., 2011; Hesse and Gross, 2014; Moretti and Muñoz, 2013; Tagliazucchi et al., 2012; Timme et al., 2016). This seemingly ubiquitous phenomena has sparked a renaissance of new ideas attempting to understand the self-organizing nature of our world (Bak and Chen, 1991). More specifically, it has been shown that the Ising model at criticality models the statistics of brain dynamics quite well (Deco et al., 2012; Fraiman et al., 2009b; Haimovici et al., 2013; Marinazzo et al., 2014), which combined with evidence of critical variables in brain dynamics has led to the emergence of the critical brain hypothesis (Beggs, 2008; Hesse and Gross, 2014). Systems tuned to criticality, self-organized or otherwise, exhibit a number of useful informational properties that allow for the efficient distribution of and susceptibility to information (Beggs, 2008; Marinazzo et al., 2014; Shew and Plenz, 2013; Timme et al., 2016). These ideas have been further developed to suggest more broadly that critical systems are evolutionary advantageous and stable attractors for systems living in complex environments as they are more effective at reacting to their environment and ensuring their continued survival (Goldenfeld and Woese, 2011; Hidalgo et al., 2014; Mora and Bialek, 2011). In this paper we attempt to understand an emerging theory of consciousness, the integrated information theory of consciousness (IIT), by modeling a toy-brain using a generalized version of the Ising model (Oizumi et al., 2014). Integrated information, or Φ (big Phi), is calculated for the model as a function of the temperature T and for different connectivity networks. We find that the susceptibility of Φ maximizes at criticality, a property deemed important for systems that are immersed in complex environments (Hidalgo et al., 2014). These results further support the critical brain hypothesis suggesting that conscious systems likely self-organize/evolve towards criticality in order to maximize the repertoire of environments that they can survive. These results also reconcile one particular criticism of IIT that claims that 'intuitively unconscious' simple systems are capable of generating high Φ, and therefore paradoxically experience consciousness contradicting the phenomenological motivations of IIT (Aaronson, 2014). Our results highlight the point that what intuitively separates 'life' systems from 'inanimate-yet-sentient' systems is that life-like systems are animated, dynamic, and susceptible to their surrounding. While the low-temperature limit of the Ising model is capable of generating high Φ, only at criticality is Φ maximally susceptible. If evolution is in fact strongly attracted to criticality then consciousness also likely falls under the purview of this very general phenomenon.

The integrated information theory of consciousness (IIT) is a top-down, phenomenological approach to defining consciousness (Oizumi et al., 2014). Starting from phenomenological axioms the theory constructs mathematical postulates that create a workspace for scientists to test and explore this particular definition of consciousness. Unfortunately, many calculations in the theory prove to be intractable, scaling super-exponentially with respect to the size of the system of interest. If one wants to analyze the brain with the perspective of IIT, some sort of bridge needs to be built to link IIT and brain dynamics. In this paper, the generalized Ising model acts as this bridge by proxy. The Ising model acts as a proxy to brain dynamics by allowing function to be simulated from structure. In one set of simulations, the model is first fitted to empirical functional connectivity (FC) maps of the human brain starting with a structural connectivity (SC) map inspired from diffusion brain imaging. In the second, a large number of randomly generated non-sparse SCs are used to generate broader statistical results. IIT calculations are embedded within the simulations where measurements of Φ are calculated on a state-by-state basis. By using this simple model as a proxy we make the utility of IIT more accessible allowing for the exploration of the properties and predictions of the theory.

The main measure in IIT is integrated information (Φ), big Phi. Though other measures exist (Sarasso et al., 2015) which try to capture some form of integration or complexity, this paper will use Φ as its main metric. For a wholesome overview of the mathematical taxonomy of the possible variations in defining integrated information, see (Tegmark, 2016).

To measure integrated information one needs to have access to the transition probabilities of the system. Naturally this is information we are not always privy to when it comes to complex phenomena like brain dynamics. This problem is circumvented by using a sufficiently simple model where the transition rates can be readily calculated, in this case the generalized Ising model. The 2D Ising model which was famously found to exhibit a phase transition at a critical temperature has been shown to also exhibit similar statistical properties to that of the brain which is also thought to be critical (Beggs, 2008; Beggs and Plenz, 2003; Brochini et al., 2016; Chialvo, 2004, 2010; de Arcangelis and Herrmann, 2010; Expert et al., 2011; Fraiman et al., 2009a; Haimovici et al., 2013; Hesse and Gross, 2014; Moretti and Muñoz, 2013; Onsager, 1944; Tagliazucchi et al., 2012; Timme et al., 2016). We then generalize the 2D Ising model such that its interactions are not confined to only nearest-neighbors and instead can use any SC that is given as input.

Reduced SC matrices (see methods) from diffusion tensor images (DTI) of the brain are taken as input for the Ising model's interaction couplings which are then simulated with the Metropolis algorithm (Hastings, 1970). 159 randomly generated non-sparse connectivity matrices are also generated and simulated within the model. From these results 2 different important temperatures (our fitting parameter in the model) are searched for; 1. T_c, the critical temperature which maximizes the magnetic susceptibility and corresponds to the transition point of the model from an ordered (magnetized) phase to a disordered (non-magnetized) phase, and 2. T_{min}, the temperature that minimizes the distance between our simulated FC and empirical FC.

Our results corroborate previous results that the critical temperature of the model is the point which fits best with the empirical statistics and furthermore suggests that integrated information as an order parameter also undergoes a phase transition near the critical point when analyzing the fluctuations of Φ as a function of temperature. Furthermore, these results fit into a larger paradigm that seeks to understand the nature of evolution and the adaptive advantage of critical systems.

The brain-like networks are then compared to the random simulations. We find that while the random networks and brain networks can be similar in many ways, the brain networks generally demonstrate improvements over the random networks to predict the empirical connectivity, and in the case of the Default Mode Network, did so significantly. Furthermore, a rich variety of qualitative behaviours were demonstrated by each network. For example some networks generated their maximal integrated information at criticality while others did not. These results were initially surprising as our hypothesis going into this project was that integrated information would be maximized exclusively near criticality. While initially we were able to confirm this hypothesis using a simpler algorithm for ϕ (small phi) written in MATLAB not presented in this paper, the published algorithm pyPhi for calculating Φ (Big Phi) was not able to recreate these results with consistency. It remains a task for future works to explore more rigorously the relationship between criticality and integrated information, however this paper illustrates provisional results indicating that Φ seems to undergo its own transition point near criticality. These results contextualized within the paradigm of the physics/mathematics of emergence and complexity hint that consciousness, criticality, and complexity are deeply intertwined concepts that may strongly overlap with the physics of phase transitions and universality.

3.1 Results

3.1.1 Integrated Information and Criticality

159 Ising simulations are generated using N = 5 nodes, fully-connected networks with random weights. The magnetic susceptibility χ, the variance of integrated information (which we consider generally to be the susceptibility of Φ) and integrated information for each simulation is calculated as a function of the fitting parameter T.

$$\begin{aligned} \chi &= \frac{\langle M^2 \rangle - \langle M \rangle^2}{T} \\ \sigma^2(\Phi) &= \langle \Phi^2 \rangle - \langle \Phi \rangle^2 \end{aligned} \tag{3.1}$$

Averaging these properties across all random simulations shows a strong relationship between the susceptibility of the system and the integrated information it generates (Figure 3.1). Near the onset of criticality, which can generally be approximated by the peak of the susceptibility curve (Severino et al., 2016), integrated information, much like the magnetization in the Ising model, also seems to undergo a phase transition which is seen as a peak in the variance of Φ. The Ising model at criticality has already been shown to model the functional connectivity of the brain and our results show that this also coincides with the regime where the fluctuations of integrated information is maximized which suggests a transition point for integrated information as an order parameter (Fraiman et al., 2009a; Haimovici et al., 2013). Of the 159 of random networks simulated, only 6 demonstrated the ability to maximize Φ at criticality while the rest had the general tendency to decrease Φ as a function of temperature, though not necessarily monotonically.

9 brain networks of size N = 5 are also simulated in the Ising model. These networks are coarse representations of 9 major resting state networks in the brain (see Methods for details). Like the random networks, simulating these networks in the Ising model gives the susceptibility and integrated information as a function of temperature. In contrast to the 6/159 of random networks that had Φ maximize at criticality, 3 of the 9 brain networks (Auditory, Default Mode, Visual Medial) demonstrated that their Φ maximizes near criticality. While it's difficult to say that these provisional results are significantly characteristic of brain networks, it does show that certain networks have $\Phi(T)$ profiles that are capable of maximizing their integrated information at criticality, which combined with maximal susceptibility may be evolutionary advantageous.

In Figure 3.2 we plot the results from these 9 simulations where the values are normalized for comparison. Integrated information is seen to be capable of a number of diverse forms across the different simulations. Notably, all the networks seem to have two branches in their Φ curve as the temperature approaches 0. For as of yet undiscovered reasons, the integrated information generated by the model seems to be capable of spontaneous symmetry breaking where the different minimum energy

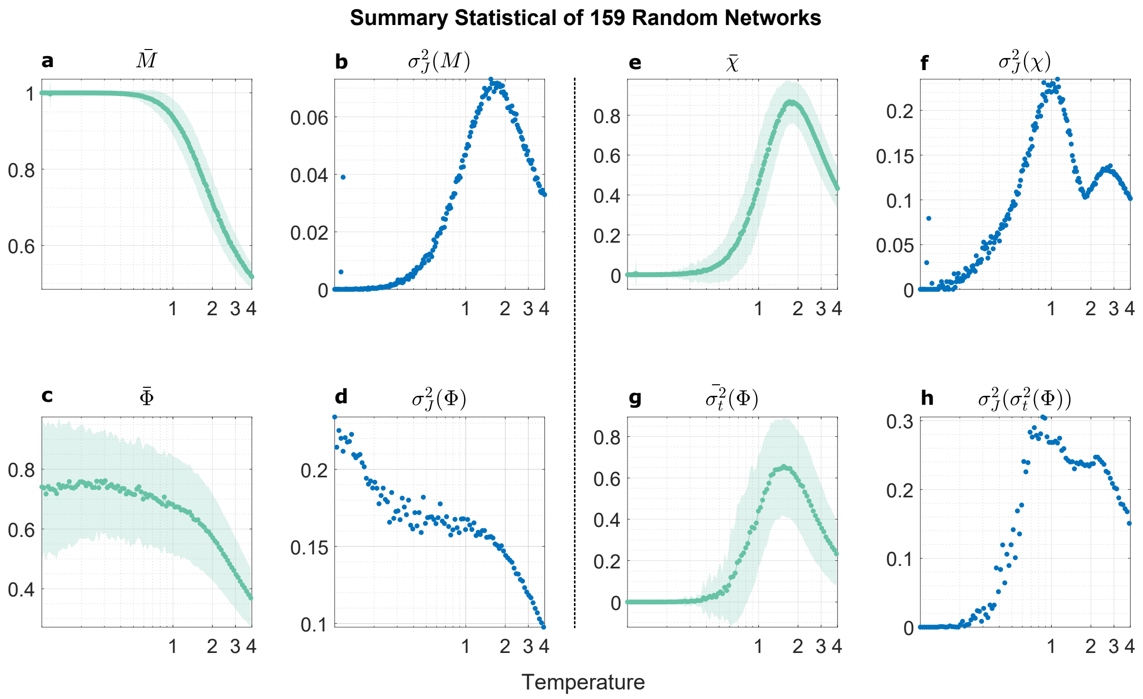

Figure 3.1: The summary statistics for the two order parameters, Magnetization M and Φ (panels **a, c**) across all the 159 random network simulations are shown. The variance of Φ, $\sigma_t^2(\Phi)$ (panel **g**) is interpreted as a susceptibility of Φ and is compared to the magnetic susceptibility χ (panel **e**). These susceptibilities peak at the same critical temperature indicating the phase transition of the integrated information as an order parameter. The variance of both order parameters and their susceptibilities across different connectivities are also compared. These are the σ_J^2 plots (panels **b, d, f, h**). These plots demonstrate the susceptibility of these variables with respect to changes in the connectivity. The variation σ_J^2 in the susceptibilities drop to a local minimum near criticality with local maxima on the outset.

states corresponding to the opposite magnetizations of ±1 are each unique in their capability of generating Φ. In Figure 3.1 some summary statistics for the random networks are shown. Magnetization M and its corresponding susceptibility χ, are plotted in the top row, first and third columns from the left respectively. Φ and its variance $\sigma_t^2(\Phi)$ are plotted in the second row. The variance of these variables across the different connectivities are plotted along the second and fourth columns from the left respectively: $\sigma_J^2(M), \sigma_J^2(\chi), \sigma_J^2(\Phi), \sigma_J^2(\sigma^2(\Phi))$. The variances summarize the tendency for these variables to fluctuate within simulation and across simulations, quantifying their susceptibility to environmental fluctuations and internal connectiv-

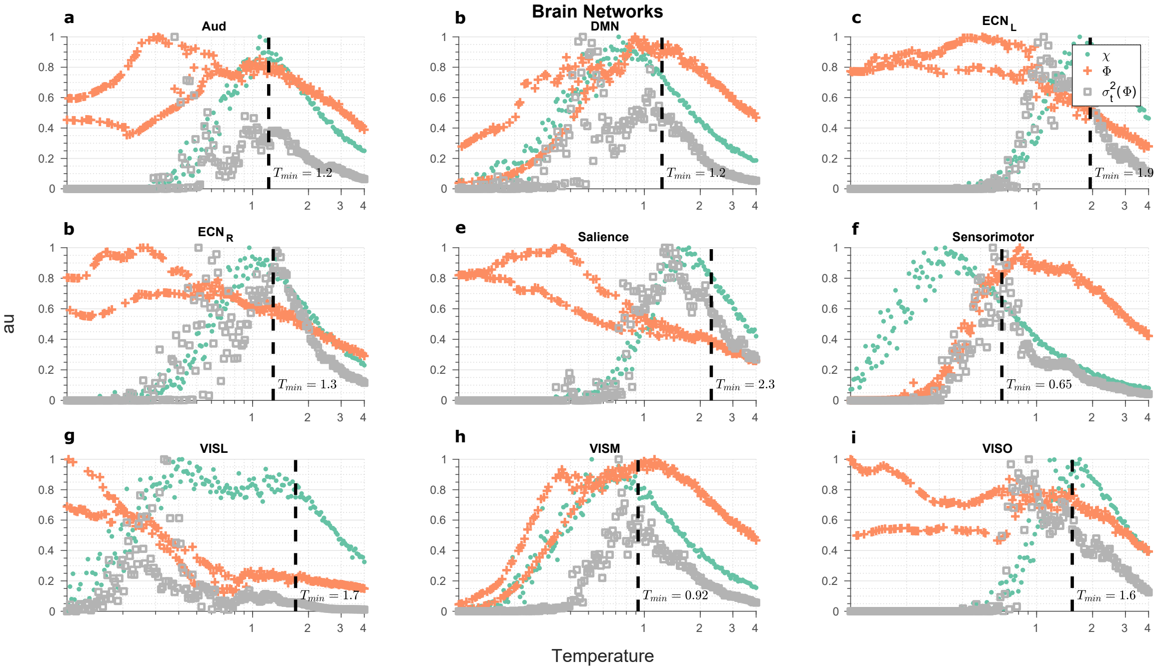

Figure 3.2: The 9 networks representative of the resting state networks of the brain are simulated and the Φ, χ, and $\sigma_t^2(\Phi)$ are plotted as a function of the temperature T. The minimum temperature T_{min} where the model fit best with the empirical FCs are marked with the dotted line and labeled.

ity fluctuations. We note that at the critical temperature, denoted roughly by the peaks of χ, the 'susceptibility' of Φ, $\sigma_t^2(\Phi)$ also peaks. When looking at Φ across different simulations, $\sigma_J^2(\Phi)$, we observe that there seems to be two transition points. One transition point at low temperatures leading into a plateau region followed by a second transition close to the classical critical point where the variations in Φ begin to fall off. These results illustrate the regions where changes in the structural connectivity of the model have the most influence on the generation of integrated information. While the magnetization of the model near criticality is maximally sensitive to changes in the structural connectivity, integrated information instead has a broad plateau region of uniform sensitivity. This result is useful in assessing how structural changes in a system can lead to functional changes which are capable of generating integrated information or consciousness.

3.1.2 Model Fitting

To be able to access the tools offered by IIT the Ising model was chosen as a proxy. The intractable nature of the calculations involved with IIT force us away from large datasets and/or simulations. So any analysis involving IIT must be for small systems (in our case, N = 5 nodes). Furthermore, the theory requires one to have a complete knowledge of the transition rates of the system for all possible configurations (though there have been work-arounds introduced under a Gaussian assumption Oizumi et al. (2016); Tegmark (2016)). We can further contextualize our analysis of IIT on neurological systems by choosing an appropriate model. As a neurologically motivated choice, the Ising model at criticality has demonstrated in a wide variety of applications to emulate the statistics of the brain Chialvo (2004, 2010); Fraiman et al. (2009a); Haimovici et al. (2013). Furthermore, the Ising model simulated on the Metropolis algorithm can be viewed as a Markov chain Teif (2007) and therefore its transition rates can be readily calculated. For these reasons the Ising model was justified to be the proxy for exploring the utility of IIT while simultaneously maintaining a neurological motivation.

To assess how well our simulations from the 5-node brain networks are capable of fitting the empirical FCs, the Euclidean distance between the fisher-transformed ($z \equiv \text{arctanh}(r)$) simulated FCs and empirical FCs are calculated as a function of temperature. The empirical FCs are generated from the average over 69 resting-state fMRI FCs from healthy control subjects (see Methods). The temperature T_{min} corresponding to the point of minimum distance $d_{min} = d(T_{min})$ marks the value of T where the model best fits the data. For larger simulations not presented in this paper (for example N = 84) T_{min} is very close to T_c, however for our N = 5 simulations of the brain networks the variability can be much larger. To assess the significance of these fluctuations, randomly generated networks of the same size are simulated to generate a null distribution of minimum distances (see Methods). Comparing the minimum distances from the random networks to the results from the brain networks gives the significance for each of the 9 brain networks ability to model its respective network (Figure 3.3).

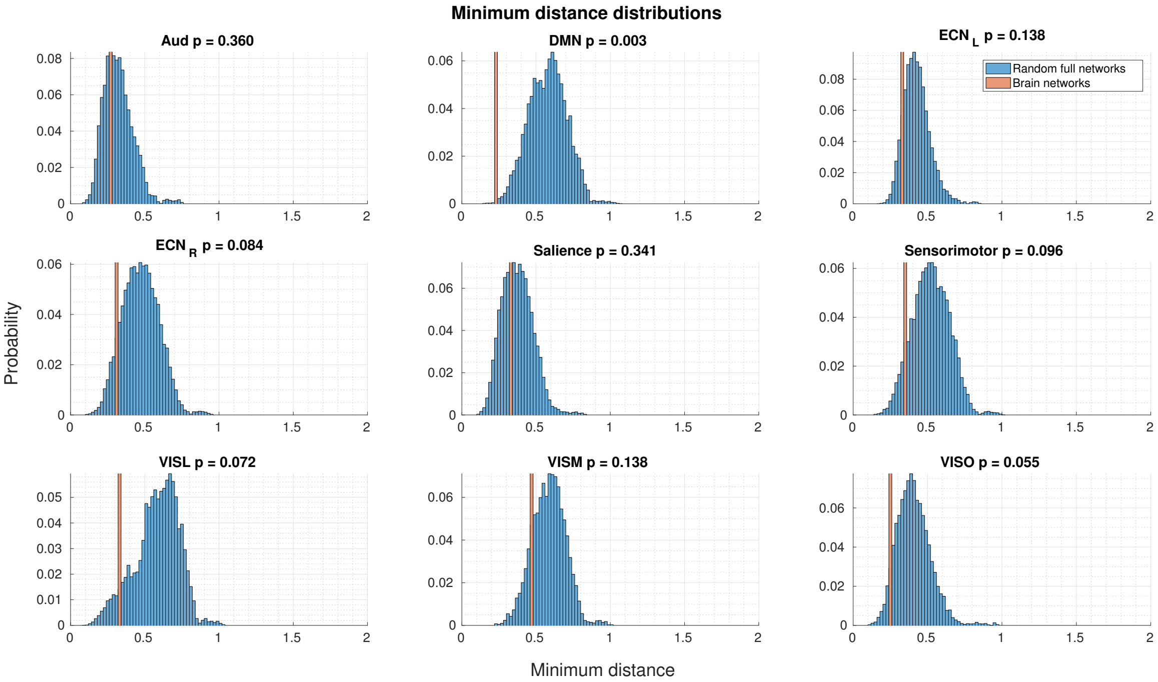

Figure 3.3: To assess the capacity of the 5-node Ising brain networks to predict the empirical FCs of human brains the minimum euclidean distances between the simulations and FCs are binned. The random networks act as a null-distribution to assess the significance of our results. All simulations were on the left of the null-distribution, so none in general did worst than random chance. Most however were not significantly better than random with the sole exception of the DMN, which the random networks were not readily able to predict. This is likely due to the particularly sparse DMN network.

Not surprisingly, we find that a few of the brain networks did not have a significantly smaller distance to the empirical FCs than the random networks indicating that these particular 5-node brain networks did not predict brain functional connectivity any better than the distribution of random networks. For example, the Auditory network, Salience network, Visual Medial network, and the Executive Control Left network had p-values of $p_{AUD} = 0.36$, $p_{SAL} = 0.34$, $p_{VIS_M} = 0.14$, $p_{ECN_L} = 0.14$ which were among the least predictive networks. The rest of the networks in our analysis all had $p < 0.1$. Looking at the magnitudes of the minimum distances it is clear that some of the networks were in general better represented by the Ising model than others. These results summarize the capability of the coarse 5-node brain-like networks utilized in this paper to predict the empirical functional connectivity of the networks in the human brain. With an awareness that these networks are essentially cartoon representations of the brain networks we can try to interpret these results to generalize how integrated information can behave in Ising-like systems.

3.2 Discussion

To investigate the properties of this new measure of integrated information introduced by IIT we have in this study employed the relatively simple Ising model to act as a proxy to the real brain. The Ising model is generalized to use any graph as its connectivity where in this study we have looked at 159 random networks of 5 nodes as well as 9 networks representing the human brain RSNs also composed of just 5 nodes. The results from the Ising model analyzed with IIT show that integrated information tends to be maximally susceptible at the critical temperature. The statistics of the 159 random networks summarize these results across variations of fully connected connectivity matrices to show that while there exists a rich variety of $\Phi(T)$ curves, on average the 'susceptibility' of $\Phi(T)$, $(\sigma^2(\Phi(T)))$, behaves quite similarly to the magnetic susceptibility that is normally the marker for the second order phase transition of the classical 2D Ising model. These results indicate that integrated information as an order parameter likely has its own class of phases which a system can transition to and from. To generate a taxonomy of the possible phases that integrated information could exhibit would require a much more thorough exploration of the possible structural connectivities and dynamical rules that a system could obey. This project confined itself to the Ising model on fully connected graphs obeying the Metropolis algorithm. In the future as more efficient algorithms for calculating Φ emerge (or as a compromise accurate correlates of Φ) combined with Monte Carlo and network renormalization group methods (Gandhi; Newman and Watts, 1999; Pawley et al., 1984; Rozenfeld et al., 2010; Swendsen, 1979; Wilson, 1975) the exploration of larger networks of different classes (e.g. sparse, modular hierarchical, small-world, fractal) could lead to the identification of a rich taxonomy of phases of integrated information.

The exploration of integrated information in the context of critical systems undergoing phase transitions motivates a few new questions in regards to the relationship between evolution, complexity, and consciousness. In the work done by Albantakis et al. (2014); Joshi et al. (2013) on complexity and the evolution of neural models and integrated information, it was shown that fitness can correlate strongly with Φ when the system is constrained in size/resources. While it is not always true that a system will evolve to generate high Φ under more liberal constraints (infinite resources), it does seem to be that there may be some evolutionary advantage for having high Φ. Since Φ essentially measures the emergence of higher-order concepts within a system, intuitively it may not be surprising that systems that are capable of generating higher-order concepts will be capable of representing a more diverse set of states than systems that cannot. Therefore for resource-limited systems, having an efficient means to represent internal and external states may automatically give rise to high Φ or consciousness.

It is fair to think of integrated information as a type of complexity measure as it aims to measure how mechanisms in a system interact and constrain each other in emergent and irreducible ways. The theory aims to measure emergent properties of a system that cannot be explained by independent (or semi-independent) components of that system. The measure is sensitive to not just information, which in general can be maximized by deterministic systems with unique pasts and futures, but also to the distribution and integration of information which in general can be maximized by strongly coupled systems. To have a system that is both strongly coupled and informative requires a balance between segregating forces that act to differentiate the system into diverse states as well as integrating forces that create new forms of information that could not otherwise arise from the components. In a system like the Ising model, it is expected that these exact properties emerge near the critical temperature at the onset of its phase transition.

By definition, critical systems have diverging correlation lengths and critical slowing-down (integration in space and time), and simultaneously exhibit distinct and segregated structures at all scales (scale-invariance). They are generally found in regimes of systems undergoing some kind of transition between different phases (e.g. magnetized vs. non-magnetized in the Ising model, synchrony vs. asynchrony in the Kuramuto model (Acebrón et al., 2005; Cumin and Unsworth, 2007; Hansel et al., 1993; Kuramoto, 2012; Kuramoto and Araki, 1975)) . In contrast to sub-critical regimes which can become completely uniform due to their strong coupling (high integration, low differentiation) and super-critical regimes which can become completely noise driven (low integration, high differentiation), critical systems sit in the sweet spot to generate non-negligible Φ that is maximally susceptible to the perturbations of its environment and its own state. Our results indicate that while sub-critical regimes are quite capable of generating Φ, the variations in Φ in this regime are negligible. Only near the critical point does Φ have both large values and large fluctuations indicating that the critical point of the system is maximally receptive to its internal (or external) states.

Timme et al. (2016) showed that in neural tissues and in a cortical branching model (which is not too different from the classical Ising model) that neural complexity is maximized at criticality. Bak (1996); Bak and Paczuski (1995) even define the origins of complexity as the ”tendency of large dynamical systems to organize themselves into a critical state”. The novelty of this study is that by using IIT we argue that consciousness arises at criticality and that the brain self-organizes into states that maximize both its magnitude of consciousness and its susceptibility to internal and external states.

3.3 Methods

3.3.1 Empirical Networks

A set of sixty-nine healthy subjects, between 22 to 35 years old, were studied during wakefulness. Informed consent to participate in the study was obtained from every subject. The Ethics Committee of the Washington University and the University of Minnesota approved the study. Structural and functional data were acquired at the Washington University - University of Minnesota Consortium of the Human Connectome Project (WU-Minn HCP). Details about the data acquisition and preprocessing can be found here (Andersson and Sotiropoulos, 2015, 2016; Andersson et al., 2003; Fischl, 2012; Glasser and Van Essen, 2011; Glasser et al., 2013; Jenkinson et al., 2002, 2012; Van Essen et al., 2011). The raw voxel-space images of the full brain human connectome is parcellated into 84 labeled regions using FSL, Freesurfer and MRTrix. Due to the sparse nature of the SC map, a new transformed SC is constructed from the inverse of the minimum distances of the original SC. The inverse of the minimum distances removes the sparsity from the SC. This process is necessary due to the small size of our simulations. SCs that are too sparse may give disconnected networks when the 5-node sub-networks are extracted which will then behave trivially in Ising simulations. The inverse minimum distances are used to circumvent this problem. For larger simulations this step can be skipped since the SC will not have disconnected nodes. The inverse-minimum-distance SC maps are then normalized such that their largest weight is unity. This process is visualized in Figure 3.4.

9 sub-networks are then extracted from the SC map where each sub-network

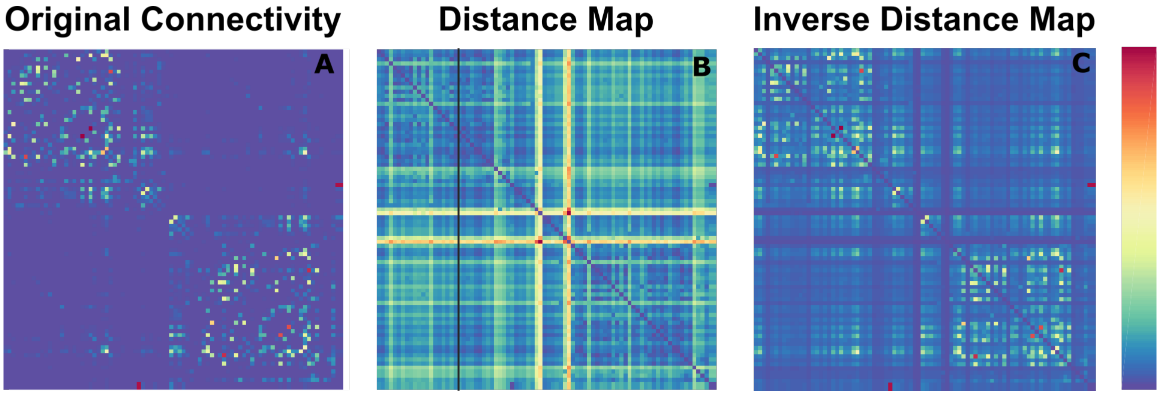

Figure 3.4: **A:** The connectivity matrix averaged across the 69 subjects imaged in the HCP. **B:** The shortest paths to each neighbour give us the distance matrix. **C:** The inverse of the distance matrix gives back a matrix that is less sparse than our original map by considering once removed connections. This helps stabilize results by avoiding almost disconnected edges.

is modeled after resting state networks in the brain. The nine RSNs were identified from an independent set of nineteen healthy controls Demertzi et al. (2014); Ribeiro de Paula et al. (2017) and an average z-map template was created for each RSN. For each z-map template, the top five most representative ROIs were chosen to represent each of the nine networks (see Appendix). The ROIs were chosen such that hemispheric redundancies were removed. Future studies that include the symmetric redundancies can be of interest as well, though in this particular study this was unfeasible due to the constraints imposed by IIT on the network size. Using five representative ROIs for each of the nine RSNs, an SC sub-network is extracted from the larger whole-brain SC map. Similarly, FC sub-networks are also extracted from the whole-brain FC map. This process leaves us with 9 SCs and FCs maps, one for each RSN. The SC maps are then used to define the connectivity of the Ising model, and the FC maps are used to fit the model to empirical results and find T_{min}. This process for each connectivity matrix is summarized in Figure 3.5.

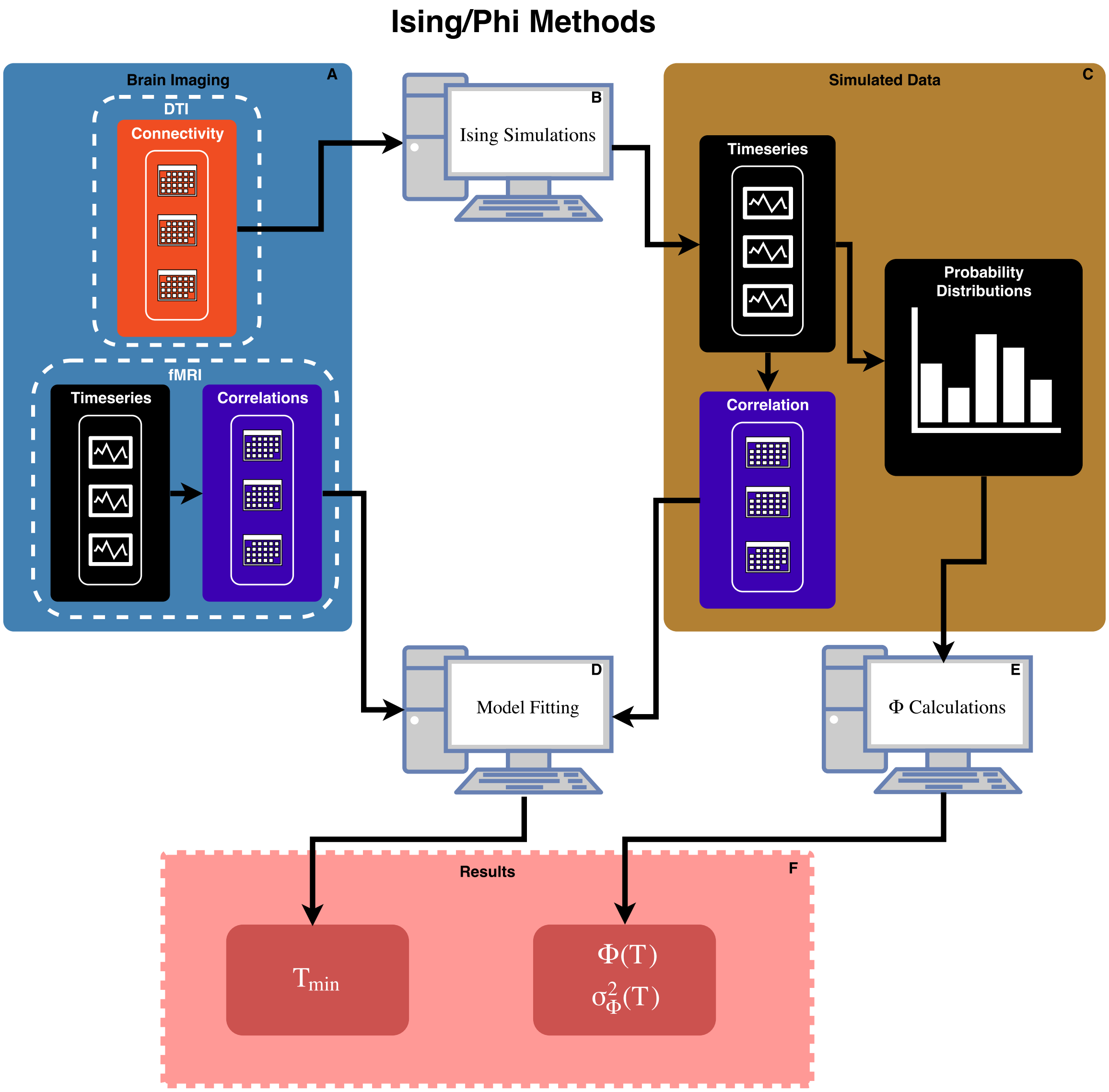

Figure 3.5: **A:** Empirical DTI-weighted brain tractography and fMRI images taken from the HCP are used to simulate and fit the Ising model of the brain respectively. **B:** The connectivity matrices obtained from the tractography are the only inputs used to model the brain in the Ising model. **C:** From the Ising simulations, timeseries, correlation matrices and state/transition probabilities are generated. **D:** The empirical correlation distribution is compared with the simulated correlation distribution by measuring the euclidean-distance of the fisher-transformed correlation coefficients to find the temperature which minimizes their distance. This minimum temperature, T_{min} fits the Ising model to empirical results. T_{min} tends to be around the critical temperature T_c. **E:** Using the state/transition probabilities, integrated information (Φ) is calculated. **F:** The results of this project are summarized by functions $\Phi(T)$ and its susceptibility, $\sigma^2_\Phi(T)$ and where T_{min} highlights the regions where the model fits best.

3.3.2 Random Networks

One major limitation of the methods used in this study comes from the fact that the simulated networks contain only 5 nodes, which we take to represent an entire functional network in the brain. To assess how well the empirically extracted brain networks represent the real brain, we compare their results with randomly generated networks. 159 fully connected networks with random weights uniformly sampled between 0 and 1 are generated. The networks are then normalized such that their strongest weight is always unity. Under the null-hypothesis that the empirically driven SCs and the randomly generated SCs are identical, the results obtained from the random networks are used as a null-distribution to assess the significance of the results obtained from the brain networks.

3.3.3 Phi

Integrated Information (Φ) is calculated in the 5-node Ising model for 2000 iterations after the model reaches a steady-state which is assumed to be achieved after 500 iterations. The transition probability matrix (TPM) for the entire system of 5 nodes is calculated. Φ is calculated using the pyPhi toolbox (Oizumi et al., 2014) for each state of the simulation across all its iterations. These measurements are gathered into ensemble averages of temperature bins, measuring $\Phi(T)$.

Bibliography

Aaronson, S. Why i am not an integrated information theorist (or, The unconscious expander). *Shtetl Optim. Blog Scott Aaronson. http://www. scottaaronson. com/blog*, 2014.

Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., and Spigler, R. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137, 2005.

Albantakis, L., Hintze, A., Koch, C., Adami, C., and Tononi, G. Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Comput Biol*, 10(12):e1003966, 2014.

Andersson, J. L. and Sotiropoulos, S. N. Non-parametric representation and prediction of single-and multi-shell diffusion-weighted MRI data using Gaussian processes. *Neuroimage*, 122:166–176, 2015.

Andersson, J. L. and Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*, 125:1063–1078, 2016.

Andersson, J. L., Skare, S., and Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003.

Bak, P. Complexity and Criticality. In *How nature works*, pages 1–32. Springer, 1996.

Bak, P. and Chen, K. Self-organized criticality. *Scientific American*, 264(1):46–53, 1991.

Bak, P. and Paczuski, M. Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences of the United States of America*, 92(15):6689, 1995.

Beggs, J. M. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 366(1864):329–343, 2008.

Beggs, J. M. and Plenz, D. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003.

Brochini, L., de Andrade Costa, A., Abadi, M., Roque, A. C., Stolfi, J., and Kinouchi, O. Phase transitions and self-organized criticality in networks of stochastic spiking neurons. *Scientific reports*, 6:35831, 2016.

Chialvo, D. R. Critical brain networks. *Physica A: Statistical Mechanics and its Applications*, 340(4):756–765, 2004.

Chialvo, D. R. Emergent complex neural dynamics. *Nature physics*, 6(10):744–750, 2010.

Crutchfield, J. P. Between order and chaos. *Nature Physics*, 8(1):17–24, 2012.

Cumin, D. and Unsworth, C. Generalising the Kuramoto model for the study of neuronal synchronisation in the brain. *Physica D: Nonlinear Phenomena*, 226(2): 181–196, 2007.

de Arcangelis, L. and Herrmann, H. J. Learning as a phenomenon occurring in a critical state. *Proceedings of the National Academy of Sciences*, 107(9):3977–3981, 2010.

Deco, G., Senden, M., and Jirsa, V. How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. *Frontiers in computational neuroscience*, 6, 2012.

Demertzi, A., Gomez, F., Crone, J. S., Vanhaudenhuyse, A., Tshibanda, L., Noirhomme, Q., Thonnard, M., Charland-Verville, V., Kirsch, M., Laureys, S., et al. Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex*, 52:35–46, 2014.

Expert, P., Lambiotte, R., Chialvo, D. R., Christensen, K., Jensen, H. J., Sharp, D. J., and Turkheimer, F. Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *Journal of The Royal Society Interface*, 8(57):472–479, 2011.

Fischl, B. FreeSurfer. *Neuroimage*, 62(2):774–781, 2012.

Fraiman, D., Balenzuela, P., Foss, J., and Chialvo, D. R. Ising-like dynamics in large-scale functional brain networks. *Phys. Rev. E*, 79:061922, Jun 2009a. doi: 10.1103/PhysRevE.79.061922. URL `https://link.aps.org/doi/10.1103/PhysRevE.79.061922`.

Fraiman, D., Balenzuela, P., Foss, J., and Chialvo, D. R. Ising-like dynamics in large-scale functional brain networks. *Physical Review E*, 79(6):061922, 2009b.

Gandhi, S. Renormalization group on complex networks.

Glasser, M. F. and Van Essen, D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of Neuroscience*, 31(32):11597–11616, 2011.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, 80:105–124, 2013.

Goldenfeld, N. and Woese, C. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.*, 2(1):375–399, 2011.

Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical review letters*, 110(17):178101, 2013.

Hansel, D., Mato, G., and Meunier, C. Phase dynamics for weakly coupled Hodgkin-Huxley neurons. *EPL (Europhysics Letters)*, 23(5):367, 1993.

Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

Hesse, J. and Gross, T. Self-organized criticality as a fundamental property of neural systems. *Frontiers in systems neuroscience*, 8, 2014.

Hidalgo, J., Grilli, J., Suweis, S., Muñoz, M. A., Banavar, J. R., and Maritan, A.

Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, 2014.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. Fsl. *Neuroimage*, 62(2):782–790, 2012.

Joshi, N. J., Tononi, G., and Koch, C. The minimal complexity of adapting agents increases with fitness. *PLoS Comput Biol*, 9(7):e1003111, 2013.

Kuramoto, Y. *Chemical oscillations, waves, and turbulence*, volume 19. Springer Science & Business Media, 2012.

Kuramoto, Y. and Araki, H. Lecture notes in physics, international symposium on mathematical problems in theoretical physics. 1975.

Marinazzo, D., Pellicoro, M., Wu, G., Angelini, L., Cortés, J. M., and Stramaglia, S. Information transfer and criticality in the ising model on the human connectome. *PloS one*, 9(4):e93616, 2014.

Mora, T. and Bialek, W. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

Moretti, P. and Muñoz, M. A. Griffiths phases and the stretching of criticality in brain networks. *arXiv preprint arXiv:1308.6661*, 2013.

Newman, M. E. and Watts, D. J. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.

Oizumi, M., Albantakis, L., and Tononi, G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol*, 10 (5):e1003588, 2014.

Oizumi, M., Amari, S.-i., Yanagawa, T., Fujii, N., and Tsuchiya, N. Measuring integrated information from the decoding perspective. *PLoS Comput Biol*, 12(1): e1004654, 2016.

Onsager, L. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Phys. Rev.*, 65:117–149, Feb 1944. doi: 10.1103/PhysRev.65.117. URL `https://link.aps.org/doi/10.1103/PhysRev.65.117`.

Pawley, G., Swendsen, R., Wallace, D., and Wilson, K. Monte Carlo renormalization-group calculations of critical behavior in the simple-cubic Ising model. *Physical Review B*, 29(7):4030, 1984.

Ribeiro de Paula, D., Ziegler, E., Abeyasinghe, P. M., Das, T. K., Cavaliere, C., Aiello, M., Heine, L., Perri, C., Demertzi, A., Noirhomme, Q., et al. A method for independent component graph analysis of resting-state fMRI. *Brain and behavior*, 7(3), 2017.

Rozenfeld, H. D., Song, C., and Makse, H. A. Small-world to fractal transition in complex networks: a renormalization group approach. *Physical review letters*, 104 (2):025701, 2010.

Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., Rosanova, M., Casali, A. G., Brichant, J.-F., Boveroux, P., et al. Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Current Biology*, 25(23):3099–3105, 2015.

Severino, F. P. U., Ban, J., Song, Q., Tang, M., Bianconi, G., Cheng, G., and Torre, V. The role of dimensionality in neuronal network dynamics. *Scientific Reports*, 6, 2016.

Shew, W. L. and Plenz, D. The functional benefits of criticality in the cortex. *The neuroscientist*, 19(1):88–100, 2013.

Swendsen, R. H. Monte Carlo renormalization group. *Physical Review Letters*, 42 (14):859, 1979.

Tagliazucchi, E., Balenzuela, P., Fraiman, D., and Chialvo, D. R. Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Frontiers in physiology*, 3, 2012.

Tegmark, M. Improved measures of integrated information. *PLoS computational biology*, 12(11):e1005123, 2016.

Teif, V. B. General transfer matrix formalism to calculate DNA–protein–drug binding in gene regulation: application to OR operator of phage λ. *Nucleic acids research*, 35(11):e80, 2007.

Timme, N. M., Marshall, N. J., Bennett, N., Ripp, M., Lautzenhiser, E., and Beggs, J. M. Criticality maximizes complexity in neural tissue. *Frontiers in Physiology*, 7, 2016.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., and Coalson, T. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex*, 22(10):2241–2262, 2011.

Wilson, K. G. The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47(4):773, 1975.

Chapter 4

Conclusion

The ethereality and allure of consciousness can perhaps be tamed by philosophy, probabilities, and Boltzmann distributions and this project has been an attempt to bring two contrasting perspectives together to help understand what it means to be conscious, or at least integrated.

Integrated information theory summons a definition of consciousness following a set of phenomenological axioms. It claims that the integrated information Φ generated by a system is a measure and description of, consciousness (or at least the human variety). Integrated Information is a function of the causal structure of the system and calculations are made by marginalizing over partitioned probabilities whose computational complexity scales super-exponentially (with some variance depending on which algorithm you choose).

The thermodynamic generalized Ising model simulates a system of nodes that oscillate and resemble what it might be to be a network of neurons. A population of these kinds of systems of neurons are analyzed. By varying the control parameter temperature, T, we push the systems from an ordered state through a phase transition into a disordered state. We observe that our measurement resembling a generalized susceptibility of integrated information maximizes near criticality. Maximally susceptible systems of the critical variety offer computational and adaptive advantages among other things. A system at criticality operates with its integrated information being maximally susceptible to ’stimulus’ in the form of concepts it can ’feel’.

Critical (self-organizing) systems are ubiquitous in the pockets of complexity in our universe and it seems that integrated information, or consciousness, is yet another example of a system riding the tide of criticality. Is consciousness just another branch off the phase diagram of our universe?

4.0.1 Future Work

One of the largest reservoirs of entropy when dealing with Ising networks is the different motifs of graphs you can make with real valued weights for edges. However, the pruning of a network is also a concept that can also be appreciated entropically via evolutionary algorithms. Understanding how the brain develops and grows into this learning machine is equivalent to understanding a magnificent growth/pruning process. Evolutionary thermodynamic systems, driven non-equilibrium systems, and self-organizing critical systems are all fascinating vectors to explore to better understand the statistics of evolution and development and can be easily explored with simple simulations of little Boltzmann machines.

The extension of these integrated Ising networks into larger interacting communities also seems to be a promising idea. Seeing how the networks would adapt to their neighbours over time would certainly be interesting as well. Contextualizing these communities/simulations in the framework of statistical mechanics and criticality can help package the complexity of these worlds and help understand the emergence of complexity. For example by identifying power laws, scaling exponents can help summarize the system statistics and help contain the complexity of critical systems.

Chapter 5

Appendix

5.1 Resting State Networks

9 resting state networks are defined with their associated region labels using the 84 node Freesurfer parcellation. For each RSN, the top 5 most representative regions were picked. In many cases there were redundancies between the left and right hemispheres and in such cases only one regions in one of the hemispheres were chosen. The only exception to this case is in the Visual Medial network where the right precuneus region is strongly a part of this network.

Auditory	
1.	L-transversetemporal
2.	L-insula
3.	L-superiortemporal
4.	L-supramarginal
5.	L-postcentral

Default Mode	
1.	L-isthmuscingulate
2.	L-rostralanteriorcingulate
3.	L-parahippocampal
4.	L-precuneus
5.	L-inferiorparietal

Executive Control Left	
1.	L-inferiorparietal
2.	L-parsorbitalis
3.	L-parsopercularis
4.	L-caudalmiddlefrontal
5.	L-parstriangularis

Executive Control Right	
1.	R-parsorbitalis
2.	R-parsopercularis
3.	R-inferiorparietal
4.	R-rostralmiddlefrontal
5.	R-caudalmiddlefrontal

Salience	
1.	R-parsopercularis
2.	R-insula
3.	R-caudalanteriorcingulate
4.	R-parstriangularis
5.	R-rostralmiddlefrontal

Sensorimotor	
1.	L-paracentral
2.	L-postcentral
3.	L-posteriorcingulate
4.	L-precentral
5.	L-transversetemporal

Visual Lateral	
1.	R-lateraloccipital
2.	R-fusiform
3.	R-parahippocampal
4.	R-lingual
5.	R-superiorparietal

Visual Medial	
1.	L-cuneus
2.	L-pericalcarine
3.	L-lingual
4.	L-isthmuscingulate
5.	R-precuneus

Visual Occipital	
1.	L-pericalcarine
2.	L-lateraloccipital
3.	L-lingual
4.	L-cuneus
5.	L-fusiform

5.2 Code

A collection of the scripts used to simulate the Ising model and calculating Φ is available online: `https://github.com/heysoos/Ising_Phi`.

The 'atom' of this code is the Monte Carlo simulation of the (generalized) Ising model. This atom is upgraded into a 'molecule' by introduction of the functions necessary to calculate integrated information. In the **Parallel** folder are run scripts to mass produce these simulations in parallel. This letst us play with population scale statistics that work to smooth out pretty nicely in face of the small-scale Ising models. A series of load files run the scripts necessary to make the figures in the paper.

There are MATLAB functions that calculate Φ but also python scripts to measure Φ using the pyPhi library: `https://github.com/wmayner/pyphi`. The figures shown in this thesis were based on the results from the latter method.

Curriculum Vitae

Name:	Sina Khajehabdollahi	
Post-Secondary Education:	Master of Physics, Western University, London, Canada	2016-present
	Honours Specialization in Astrophysics, Western University	2011-2015
Related Work Experience:	Teaching Assistant, Western University	2016-present
	Research Assistant, Western University	2014-present
Awards and Collaborations:	Long-term visitor at ELSI: Origins Network *Earth-Life Science Institute, Tokyo Institute of Technology*	Winter 2018
	Travel Award from OIST *ISSA Summer School 2017, Okinawa Institute of Science and Technology*	Spring 2017
	Ranked top 2% in the COMAP: MCM *Representing Western University internationally at the Consortium for Mathematics and its Applications: Mathematical Contest in Modeling*	Winter/Spring 2015