

# Black-box Brain Experiments, Causal Mathematical Logic, and the Thermodynamics of Intelligence

**Sergio Pissanetzky**

*School of Science, University of Houston, Clear Lake, Texas, USA*

SERGIO@SCICONTROLS.COM

**Felix Lanzalaco**

*Department of Mathematics and Statistics, Open University, Walton Hall, Milton Keynes, UK*

FL729@MY.OPEN.AC.UK

**Editor:** Randal Koene, Diana Deca

## Abstract

Awareness of the possible existence of a yet-unknown principle of Physics that explains cognition and intelligence does exist in several projects of emulation, simulation, and replication of the human brain currently under way. Brain simulation projects define their success partly in terms of the emergence of non-explicitly programmed biophysical signals such as self-oscillation and spreading cortical waves. We propose that a recently discovered theory of Physics known as Causal Mathematical Logic (CML) that links intelligence with causality and entropy and explains intelligent behavior from first principles, is the missing link. We further propose the theory as a roadway to understanding more complex biophysical signals, and to explain the set of intelligence principles.

The new theory applies to information considered as an entity by itself. The theory proposes that any device that processes information and exhibits intelligence must satisfy certain theoretical conditions irrespective of the substrate where it is being processed. The substrate can be the human brain, a part of it, a worm's brain, a motor protein that self-locomotes in response to its environment, a computer.

Here, we propose to extend the causal theory to systems in Neuroscience, because of its ability to model complex systems without heuristic approximations, and to predict emerging signals of intelligence directly from the models. The theory predicts the existence of a large number of observables (or "signals"), all of which emerge and can be directly and mathematically calculated from non-explicitly programmed detailed causal models. This approach is aiming for a universal and predictive language for Neuroscience and AGI based on causality and entropy, detailed enough to describe the finest structures and signals of the brain, yet general enough to accommodate the versatility and wholeness of intelligence.

Experiments are focused on a black-box as one of the devices described above of which both the input and the output are precisely known, but not the internal implementation. The same input is separately supplied to a causal virtual machine, and the calculated output is compared with the measured output. The virtual machine, described in a previous paper, is a computer implementation of CML, fixed for all experiments and unrelated to the device in the black box. If the two outputs are equivalent, then the experiment has quantitatively succeeded and conclusions can be drawn regarding details of the internal implementation of the device. Several small black-box experiments were successfully performed and demonstrated the emergence of non-explicitly programmed cognitive function in each case.

**Keywords:** AGI, brain emulation, causal mathematical logic, self-programming, causal entropic forces

## 1. Introduction

AGI research attempts to confront the difficult issues of "human-level intelligence" by stressing on the versatility and wholeness of intelligence, rather than on the parts of which intelligence appears to be composed. AGI also proposes that the engineering practice should be carried out according to an outline of a system comparable to the human mind. Several current projects on brain mapping, emulation, simulation, or replication, seem to be headed in a very similar direction.



Here, we propose to extend similar views to Neuroscience. We propose a universal language for Neuroscience and AGI based on causality, detailed enough to describe the finest structures of the brain yet general enough to accommodate the versatility and wholeness of intelligence. We also propose a new fundamental theory based directly and exclusively on the fundamental principle of causality, of which that language is the expression, and we further propose that engineering practice in AGI and Neuroscience should be carried out in the context of this theory.

We differ in part from the stated vision of AGI. Our vision is that very simple machines can be built, with a basic, repetitive structure, that are very different, much simpler, and much faster than the brain and yet can learn by training to a point where they exhibit the brain's full functionality. We treat the brain, or suitable parts of it, as an implementation of the causal theory, with the caveats that many different implementations may be possible and that they do not necessarily reflect all the capabilities of a theory. Yet, we take the brain as our primary example and source of inspiration, because it is the only known physical system that is intelligent and we can observe. In a sense, the brain defines intelligence and provides a point of reference for our work.

One of us (SP) has proposed a virtual machine that can be implemented on hardware and has those features (see Pissanetzky, 2012f, Sec 4.1). Here we review several computational experiments carried out on a personal computer implementation of the machine. The experiments put the theory, as well as its scope, to a test, and all of them demonstrate the emergence of intelligent behavior within obvious limitations.

Consequent with the wholeness of intelligence, and the universal scope of the proposed language and theory, topics found in this paper, as well as the examples selected for the experiments, touch on AGI subjects of interest not only to Neuroscience but also to other disciplines. This work is, in essence, strongly multi-disciplinary.

To introduce our work, we must first explain the nature of information and how we follow its flow through any device where it is being processed, such as the brain, or a computer. We must then discuss causality, state that information in nature is always causal, introduce the representation of causal information by ordered (cause, effect) pairs in fundamental Mathematics, and introduce Causal Mathematical Logic (CML) as the corresponding quantitative formalization. This paper is divided into three parts.

Part I attempts to establish Causal Mathematical Logic (CML) as a new stable and self-consistent theory of Physics, grounded on the greatest experimental base ever assembled and adequately represented by the fundamental principles of Physics that apply to all things physical. Argued in Part I is that CML can represent the dynamics of a physical system in as much detail, or in as much generality as is possible or necessary, and seamlessly refine or expand the granularity as more is learned about the system, in sharp contrast with traditional statistical and differential approaches. The main purpose of Part I is to introduce the causal theory in the full extent of its scope, which is universal for all physical systems, rather than being just a theory of AGI or one of the brain. It is also the purpose of Part I to place the specific argument of this paper within the general picture. Part I is necessarily abstract. Some readers, depending on their disciplines, may prefer to read Part II first. Part II contains many references to Part I and may serve as an introduction to it for a second reading.

Part II attempts to overcome the abstractness of Part I and expand the reader's vision in a practical direction. Its purpose is to constrain the wide scope of the causal theory to practical problems and place the reader into a practical mindset. This should help to better understand its place in today's technology and engineering, and its role in solving problems that can't be solved without it. Part II tries to establish CML as a theory of the brain, under the working hypothesis that the brain is physical and intelligence can and should be explained by Physics. The argument in Part II proposes that the brain, in its capacity as an organ in an organism, is best understood as a causal analog system, specialized in processing causal information, and fundamentally different from any Turing computer. That intelligence is not intended in the design of the brain but emerges as a side effect of resource preservation. That any digital or analog computer, or even an appropriately designed nanomaterial, that stores information in a causal format and adequately simulates the analog behaviors of neurons, but is not explicitly programmed to simulate the brain's corresponding cognitive response, will nevertheless cause that correct cognitive response to emerge. And that any physico-mathematical description of the brain must be carried in the context of a universal theory that emphasizes the

versatility and wholeness of intelligence, rather than other theories where simplifying assumptions are made, such as statistical distributions for statistical methods, or assumptions of smoothness and differentiability for differential methods, or the encoding of causality in a string that requires a human to deal with the encoding protocol.

Part III is directly focused on a particular target: the black-box brain experiments. It addresses immediate actions needed to jump-start a CML project and expand it to a point where CML is sufficiently developed for application in brain research. The notion of black-box brain experiments is defined, the first brain experiment where the action functional was discovered is described, and several other experiments that are either successfully completed or under way are discussed. Current limitations, and the steps necessary to overcome them, are near the focus of the discussion. A preliminary project plan is included, and an application for a patent is recommended.

This paper is of a multi-disciplinary nature. The reader will find some sections interesting and others that may not be, depending on that reader's particular interests and background. Another reader will prefer a different selection of sections.

The new theory described in this work is different from other existing theories, and the conclusions and predictions are also different and must not be judged in the context or by the tenets of other theories. They must be judged only by way of making predictions and comparing them with experimental observations. The theory itself is judged by Lee Smolin's three rules: a theory, in order to be a scientific theory, must be confirmable, falsifiable, and based on the simplest hypothesis that explains the phenomena of interest. The third rule is also known as Occam's razor. This theory is confirmable because experiments that confirm it are possible, and some have already been completed. It is falsifiable: one single experiment that is proved to contradict the theory will falsify it or some part of it. And the hypotheses are the simplest, as we explain below.

## PART I

Part I attempts to establish Causal Mathematical Logic (CML) as a new stable and self-consistent theory of Physics, grounded on the greatest experimental base ever assembled and adequately represented by the fundamental principles of Physics, and applies to all things physical. Also argued in Part I is that CML can represent the dynamics of a physical system in as much detail and with as much generality as is possible or necessary, and seamlessly refine the detail as more is learned about the system, in sharp contrast with other traditional statistical and differential approaches.

## 2. Information is Physical

The physical nature of information – the fact that information itself is a physical entity with properties of its own, such as entropy and energy – has been known for a long time. More recently, Landauer has made this concept explicit (Landauer, 1996, 1999), arguing that structure has a deep existence of its own and the laws of Physics are algorithms executable in the real universe. The recent direct experimental measurement of the energy of one single bit of information (Berut, Arakelyan, and et al., 2012) has added weight to Landauer's views. And information certainly has an economic value, as IT companies know.

But we expect more of a physical entity. We say that a system is physical when it can be observed and its properties measured. We expect the physical system to interact with other systems, and to have states and experience transitions between them, and invariants so it can be observed. We then build a mathematical model of the system and try to find a law that explains the transitions and conforms with (or falsifies) the principles of Physics. The causal theory confirms Landauer's vision. In the causal theory, structure and algorithms do have an existence of their own. Information can be observed, for example in a complex system, and its properties, the structures, are invariant and can be measured. Information has two extreme states, causal set (high entropy, fragmented and disorganized) and hierarchy (low entropy and fully organized), and a plethora of intermediate states that correspond to local minima of the action. It can interact with the environment and exchange energy with it, and when it does that a transition happens and its state

changes. Furthermore, there is a mathematical model, the causal set. And there is a law that explains the transformations, CML, and this law conforms to the principles of causality and symmetry. It would not be possible to attribute all these facts to the complex system, or even to complexity in general, because they are valid for all causal systems, except that black holes are currently under controversy. And there is even a quantum for information: the cause-effect pair.

The discovery of the causal theory, followed with the prediction in 2011 of a major feature of the brain, that dendritic trees are optimally short (Pissanetzky, 2011a, § 3.8), directly from the properties of information, and its subsequent independent experimental confirmation in 2012 with a wide experimental base including humans and other species (Cuntz, Mathy, and Häusser, June 2012), have shed additional light on the thesis and brought it to a point where it can be used for brain research.

In this work we treat information as a physical entity with physical properties of its own and try to follow its path through the brain as it follows causal links originating in the sensory organs and the brain's memory, is processed in the connectome, and exits the brain via motor nerves and other efferent nerves. Our focus remains on the information, not on the brain anatomy or physiology, but we apply the general physical and mathematical properties of information to derive requirements that the anatomy and physiology must satisfy if information is to be processed the way it is. The requirements place effective constraints on our understanding of the brain, add to neuroscientific knowledge in general, and can be directly compared with experimentally observed features and help to explain intelligent behavior in humans and animals.

We do all this in the light of this new theory known as Causal Mathematical Logic (CML), or, in older texts, as the Theory of Detailed Dynamics. CML is a general theory of Physics, not just AGI or the brain, directly grounded on the fundamental principles of Physics and nothing else. CML is distinguished from statistical or differential theories that disregard details in the information and may be considered as approximations. CML is also distinguished from the Theory of Computation, but can be emulated on a computer. CML is a discrete theory. It describes information in full detail, just as it is captured from the environment by a sensor or sensory organ, and carries that detail throughout all mathematical calculations without any approximations or simplifying assumptions. Hence, it is not surprising that CML is more general than other theories and reaches detailed conclusions that other theories cannot.

There are strong reasons to believe that CML is a theory of the brain. CML applies to causal information and proposes the solution for the famous *binding* problem – the *association* of elements of information. Based on that solution, CML proposes mathematical explanations for many phenomena commonly considered as intelligence and associated with the brain. These phenomena include the formation of *patterns*, the origin of multiple *preset goals*, the nature of *adaptive behaviors*, the process of *learning*, the mathematics of the *conscious* and the *unconscious*, the nature and effects of *causal entropic forces* on information and behavior, and many others (Pissanetzky, 2009, 2011a, 2010, 2012f).

We believe that any attempt to explain intelligence without first understanding the principles of causal information will necessarily revert and lead back to the need to understand that information first. It is just as impossible to understand the brain, or to understand AGI, without first understanding information, as it would be to understand a lubrication system without understanding oil, or a steam engine without understanding steam. Or understanding a computer's hardware without first understanding the transistor. Hence, the attempt to explain intelligence would be much easier if carried in the correct order.

In the causal theory, causal information problems are solved in *causal space*, not in spacetime as is done in Physics. Causal space is possibly very similar or the same as *information space*, proposed by Manfred Eigen (at the time of this writing, the book had not yet been released). However, we believe that “causal space” is a more appropriate name and will stick to it at least for now. In causal space, there is only one dynamical system – information, only one dynamical representation – the causal set, and only one dynamical equation – Causal Mathematical Logic (see Pissanetzky, 2012f, and references thereof), which gives origin to our familiar notions of meaning, semantics, and understanding, including our understanding of the laws of Physics, life, and thought. Causal space and spacetime map easily from one another, although maps have not yet been fully explored for some types of physical phenomena. The example “Euler Equations” in the supplementary material (Pissanetzky, 2012d) illustrates one use of the causal-space/spacetime map in Classical Physics. The notion of causal space is unifying for all sciences.

### 3. Observables and Observations

For the purpose of this theory, it is necessary to refine the meaning of the term *observable*. In Physics, an observable is usually a number that can be measured or calculated in terms of a system's state, and that remains invariant (does not change) during the dynamical evolution of the system. This definition is too narrow for AGI purposes. It is too narrow even in Physics, where the role of the observer in an observation is subject to frequent controversy, revealing the fact that a scientist's sentient perception plays a role in the observation because perception is what provides science with its observational data. That meaning is exceedingly narrow for a work such as our's, where the purpose is precisely to explore sentience, perception, cognition, intelligence, emotions, and other phenomena such as emergence and self-organization that can be "observed" and described but are not easy to measure, and to eventually recognize them in the human brain or elsewhere, emulate them on a machine, and make them measurable.

An observable is the result of an observation, performed by a human with or without the help of instruments. We have just used the word "result" quite casually, because that's how it is commonly used. But how is this "result" obtained? What exactly is an observation? What is the difference between an observable and the observation of that observable? Douglas Hofstadter has summarized this question on page 633 of *Metamagical Themas*:

*The major question of AI is this: "What in the world is going on to enable you to convert 100,000,000 retinal dots into one single word 'mother' in one tenth of a second? Perception is where it's at."*

This problem is known as the *binding problem*. It was started by Bertrand Russell a century ago. A more modern name is the *mind-body problem*. The retina, and the mouth that pronounces 'mother' are the body. The 100 million dots are the observation, 'mother' is the observable. The causal link between the 100 million dots and the myriad signals that the brain sends to muscles in the lungs, vocal chords, tongue, lips, face, to cause them to pronounce 'mother', are the mind. How does the mind work? How does perception work?

In Physics, observables must satisfy certain transformation laws that relate observations made under different conditions or by different observers. For example, tensors are used in classical Physics because they are invariant under rotations of the frame of reference. The generalized observables we are proposing in this Section must also satisfy such laws. The observable 'mother' in Hofstadter's example must remain the same even if the 100 million dots change due to light conditions, distance or angle of observation, body or eye movement, or the presence of other objects, as long as enough information is captured to make the recognition. This is a critical requirement for an observable, and it is met by CML, provided, once again, that enough geometric information is captured to make the assessment. In Physics, the invariance of an observable must be proved by a closed mathematical argument. In the causal theory, it is proved by CML. The group-theoretical argument used for the proof of the Central Theorem is the closed mathematical argument.

We assign names to observables. All names we use in speech designate observables. Anything we know and we can recognize is an observable and we assign a name to it. 'Mother' in Hofstadter's example is the name of an observable. But the meaning is not in the name. The meaning is in the observable, or, more precisely, in the causality that supports the observable. Shared observables with shared names are what makes communication possible. Function  $\mathcal{E}$  introduced in Pissanetzky (2012f) and further discussed in this paper formalizes a unique, one-on-one bijective correspondence between any given body of causal information and the corresponding observables. Function  $\mathcal{E}$  is deterministic. If two persons possess exactly the same information they will derive exactly the same observables. If their information is slightly different, then their observables will be different too. Because of extreme sensitivity of function  $\mathcal{E}$  to initial conditions (butterfly effect), the observables can be very different even if the information is only slightly different.

The additional degree of freedom introduced by name assignment gives rise to natural languages. The differences in observables caused by differences in information give rise to conflict. Shared observables make communication fast and efficient. Why are natural languages so difficult to translate into one another? The number of observables is infinite numerable. But languages are finite. Each language selects a finite subset of observables and assigns names to them, but the subsets selected by different languages are different.

Translation between two languages with slightly different subsets of observables is relatively easy, but it may be very difficult if the subsets of observables are very different. In either case, detailed translation can be achieved only by detailed reference to the underlying causality.

But communication between people and machines is difficult, and slow. The Turing test has not been met, and technology is nowhere close to meeting it. People/machine communication is difficult because machines do not implement function  $\mathcal{E}$  and can not calculate observables from information and preserve their meaning the way humans do. The Turing test will be met only when function  $\mathcal{E}$  is implemented on machines. Then, machines and man will be able to communicate efficiently.

But where is the mind? Neuroscientist David Eagleman (Eagleman, 2011) proposes that the mind is unconscious. Nineteen's century Physicist and Physiologist Hermann von Helmholtz, in his studies of vision, discovered that a form of inference was necessary to explain image recognition. He named it "Unconscious Inference" because we are not aware of its existence. We concur with both. In addition to being unconscious, we propose that the mind is of a thermodynamic nature, not algorithmic, that the function of the mind is to segregate the algorithms or behaviors that our consciousness needs from known but fragmented causality, and that function  $\mathcal{E}$  formalizes the function of the mind.

Contemporary mathematical philosopher Carey Carlson views observation as a sentient occasion of human perception. He argues that science delivers only the bare causal pattern of events, but that among these events are sentient occasions of human perception, which provide science with its observational data. He continues to argue that the events of human perception are routinely excluded from the causal patterns delivered by science for the simple reason that they are considered as mental events that cannot be explained, and that a coherent view of the world will be obtained only when those sentient events required for the causal patterns are included. We concur.

Carlson also proposes that Russell and Whitehead solved the mind-body problem when they recognized the incompleteness of science and the need to include events of sentient perception in the causal patterns. But that solution is a qualitative one. Here we seek a quantitative solution, one that can be mathematically described and implemented on a computer, as needed for a brain emulation or an AGI machine. This, Russell and Whitehead did not do. Attempts at solving the mind-body problem quantitatively continued during the better part of the 20th century. The various lines of thought that were attempted produced negative results, and were mostly abandoned after Wittgenstein's work. In Sections 1, 2 and 3 of his Turing Centennial paper, evolutionary biologist Stuart Kauffman (Kauffman, 2011), describes this research in extensive detail. But the problem remained unsolved, and has recently acquired critical proportions because it stands in the way of critical technologies such as AGI, the Semantic Web, the Internet of Things, natural language, image, voice and smell recognition, and many others known collectively as the GUAPs, or Great Unsolved Automation Problems of Software Engineering (Pissanetzky, 2011b).

We carry Carlson's first argument further, beyond science and into everyday life involving intelligence and adaptive behavior, and propose that an explanation for intelligence and adaptive behavior will not be obtained until the sentient events that connect observations with observables are quantitatively reduced to causal patterns and included into our science. We propose that **causal association** of elements of knowledge binds them together and provides the causal link between observations and observables we are seeking. And we propose that causal association is formalized by the theory of CML discussed in this paper and other previous publications, and that, consequently CML is the quantitative solution to the binding problem and the mind-body problem, as well as the GUAPs.

But CML is much more than that. CML is a new fundamental theory of Physics (Pissanetzky, 2012b, 2013) of generality far beyond that of AGI, which all things physical must obey. In particular, CML was proposed by one of us (SP) as the explanation for phenomena of emergence and self-organization (Pissanetzky, 2011a) observed in all kinds of physical systems that *do not* require the participation of any intelligent or adaptive being, and can indeed go as far as to explain the very origin of intelligence from the fundamental properties of matter. This consideration has motivated SP to propose that AGI and brain function can, and must be explained from and only from the fundamental properties of matter (Pissanetzky, 2011a, 2012f). A strong confirmation of this point of view was provided by a team of neuroscientists (Cuntz, Mathy, and Häusser, June 2012) who in 2012 independently discovered that dendritic trees in the human and animal

brain must satisfy a  $2/3$  power law that makes them optimally short. SP had predicted in 2011 that dendritic trees had to be optimally short in order for CML to work in the brain, irrespective of what caused them to be optimally short, and proposed that they were optimally short because evolution optimized space and biological resources in the brain, and that CML followed as a side effect. That prediction was made against the non-optimally short  $4/3$  power law accepted in Neuroscience at that time, and was confirmed later when Cuntz et al. announced their new optimally short  $2/3$  power law. Dendritic trees are brain, CML is mind, the  $2/3$  law is required for the mind to work, and it describes how the brain has implemented the mind.

In the area of Cosmology, the presence of “intelligence signals” at the high entropy end of the spectrum has recently been confirmed by Wissner-Gross and Freer (April 2013), based on an interpretation of some schematic dynamics as intelligent behavior. However, the conclusions are buried under excessive and inadequate formalism that hinders their explanatory power and hides their profound physical meaning. There is no mention of mathematical properties of causal set models that can explain that dynamics. Comparable results are also reported by Gardner and Conlon (May 2013), again under inadequate formalism and without mention of fundamentals. But in the context of CML, the entropy/intelligence connection has been known for quite some time, and both ends of the entropy spectrum had been studied before. See for example Table I in (Pissanetzky, 2011a), where odd-numbered rows correspond to least-action, and even-numbered rows to most-action. Causal entropic forces were explored too, as well as their effect on trajectories as a whole, (see for example Pissanetzky, 2012f, Fig 3), where they have been called “pulls.” We propose that exploration of causal entropic forces should continue in both directions, not just entropy maximization, and that they should be explored in full detail and not under assumptions of differentiability or statistical distributions.

## 4. The Theory

The causal theory discussed in this paper proposes to represent causal information by means of ordered cause-effect pairs. This is in stark contrast with the use of strings of symbols to represent information in other theories such as the theory of computation and the theory of information.

*In the causal theory causality is recognized as the supreme principle of nature, and the theory is derived directly from the fundamental principle of causality, that effects follow their causes.*

We notice that senses in animals and sensors in machines always capture causal pairs, where the cause is in the environment (dots of light in the Hofstadter’s example) and the effect is in the system (spikes in the optical nerves). This process, when viewed as an *interaction* where information is acquired from the environment - the cause - and affects the internal state of the system - the effect - is the essence of the notion of *learning*. Hence we capture the pairs right there, at the sensory organs or the sensors, and we follow the causal chains they give rise to inside the system, be it the brain or in a machine.

### 4.1 The Mathematics of Causal Theory

We begin by introducing the Mathematics of causal theory. Then, in the next sections, we address the profound physical meaning of the mathematical entities and their role in Neuroscience.

#### CAUSAL SETS

In Mathematics and Physics, *causal sets* are used to formalize causality. A causal set  $\Sigma$  is a set with a partial order, defined as follows:

$$\Sigma = (S, \omega), \quad (1)$$

where  $S$  is a finite set and  $\omega$  is a partial order for set  $S$ . A *set* is a collection of elements with no order, no value, and no meaning. Note also that  $S$  is a finite set, which makes causal theory a discrete theory. As we will see below, this is not a limitation and does not preclude the theory from handling continuous functions or smooth manifolds and differential equations.

### PARTIAL ORDER

A *partial order* is a collection of ordered pairs of elements of  $S$  related by the  $\prec$  (precedence) relation. In our case, the relation is causality, and the ordered pairs are (cause, effect) pairs, where cause  $\prec$  effect. There is no one-one implication in this definition. A cause can have many effects, and the cause precedes each one of the effects. Conversely, an effect can have many causes, and each of them precedes the effect. Let  $a, b, c \in S$  be distinct elements of  $S$ . Then the partial order  $\omega$  is required to be:

- irreflexive:  $a \not\prec a$ ;
- acyclic: if  $a \prec b$ , then  $b \not\prec a$ ; and
- transitive: if  $a \prec b$  and  $b \prec c$ , then  $a \prec c$ .

### TOTAL ORDERS

Consider now a *search process* that visits all causal pairs and elements of  $S$  and assigns sequential natural numbers to the elements as they are visited. All causal pairs must be visited exactly once, an element can be numbered only when all its predecessors have been numbered (or if it has no predecessors), and a (cause, effect) pair can be visited only when 'cause' has been numbered. The search produces a *total order* on  $S$  that is compatible with the given partial order  $\omega$ . For any given  $\omega$ , there can be many different searches and many different total orders. Let  $t$  be a total order on  $S$  and let  $T(\Sigma) = \{t(\Sigma)\}$  be the set of all total orders on  $S$  found by the searches. Note that set  $T$  is a property of  $\Sigma$  and nothing else.

### CAUSAL METRIC

Causal theory is a theory of Physics, and a theory of Physics needs a metric. In Mathematics, a *metric* is a function that defines a distance between elements of a set. In our case, this amounts to assigning a value to each one of the ordered (cause, effect) pairs in  $\omega$ . There are many possible metrics, and it is critical to choose the right one for the causal theory to work. One way to determine the correct metric is the principle of parsimony, known as Occam's razor, or as Lee Smolin's third condition for competing hypothesis. When the metric is decided by this principle, a formal axiomatic presentation of the causal theory is obtained. However, the correct metric was actually experimentally discovered, as discussed in Section 9.2. The axiomatic presentation is general and mathematically strong, but lacks physical meaning. The experimental presentation is less general and weaker in a mathematical sense, but is endowed with a profound physical meaning. The fact that the two metrics are the same is very significant. In this mathematical section, we follow the axiomatic presentation.

Let  $t \in T$  be a total order on  $S$ , and let  $t(e)$  be the number assigned by  $t$  to some element  $e \in S$ . We omit the  $\Sigma$  dependency to simplify notation. Let also  $(a, b) \in \omega$  be an ordered pair in  $\omega$ , where  $a, b \in S$  are distinct elements of  $S$ . The *measure* of pair  $(a, b)$  in order  $t$  is defined as the distance from  $a$  to  $b$  in order  $t$ :

$$m((a, b), t) = t(b) - t(a). \quad (2)$$

Finally, the metric for the causal theory is defined by assigning a measure to total order  $t$  itself, defined as the sum of the measures of all ordered pairs in  $t$ :

$$\mathcal{M}(t) = \sum_{(a,b) \in \omega} m((a, b), t). \quad (3)$$

This metric is introduced in the causal theory by axiom. The metric assigns a value to each total order in  $T$ . As order  $t$  conforms with  $\omega$ , measure  $m$  is always a positive integer, and so also is measure  $\mathcal{M}(t)$ . Hence,  $\mathcal{M}(t)$  is a positive definite, monotone metric, and only the empty set has a measure of 0.

### THE METRIC PARTITION

Consider now again the set  $T$  of all the  $\omega$ -conforming total orders of  $S$ . The metric  $\mathcal{M}(t)$  assigns a value, not necessarily different, to each total order  $t \in T$ . By doing so, the metric induces a *partition* of set  $T$



into disjoint subsets, where each subset contains one or more total orders with the same measure. We refer to this particular partition as the *metric partition*, and to the subsets in the metric partition as *metric subsets* or *metric slots*. The metric partition is “special” because it is a natural property of the given causal set  $\Sigma$  and of nothing else. In the sequel, we will discuss properties of sets of total orders that apply to any subset of  $T$  or to  $T$  itself. If we were to apply these properties to “any” subset, we would have to specify how “any” is determined. Doing so would introduce a spurious element into a theory that is otherwise genuine and fundamental because it follows from the principle of causality and the metric axiom alone. We will not do this. We will work with the metric partition and the metric subsets only. We are not saying that spurious partitions can not exist, we are just not considering them here.

The number of total orders contained in a certain metric subset of the partition is called the *population* of that subset. The following theorem applies to every subset of total orders, but is specialized here for metric subsets:

*To every metric subset  $T^* \subseteq T$  in the metric partition there corresponds a unique permutation groupoid and a unique group-theoretical block system.*

A *block system* is a partition of the original set  $S$  into subsets known as *blocks*, which have the property of being invariant under all the total orders in  $T^*$ . In other words, the *same* partition into compact blocks applies to each of the total orders in  $T^*$ . The blocks in the block system are the invariants we are seeking. The theorem is proved by construction, as follows.

#### PERMUTATIONS, GROUPOIDS, AND BLOCK SYSTEMS

A *group* is a set with a binary function to itself. A *groupoid* is a set with a partial binary function to itself. Let  $T^* = \{t_1, t_2, \dots\}$ ,  $T^* \subseteq T$  be one of the metric subsets in the metric partition.  $T^*$  is the set of total orders for the groupoid predicted by the theorem. To complete the groupoid, we need to construct the partial function that acts on set  $T^*$ . This function is defined by a set of permutations of the original set  $S$ .

A *permutation* of a set  $S$  is a bijective map  $S \rightarrow S$  that maps  $S$  to itself. Permutations are usually represented in two-line notation, where the top line lists the elements of  $S$  in some arbitrary order, and the bottom line lists the same elements of  $S$ , but in the corresponding order. Hence, a permutation of  $S$  can be specified as  $(p, q)$ , where  $p, q$  are total orders on  $S$ . Permutation  $(p, p)$  is called the *identity permutation*. The identity permutation must exist in the partial function.

To construct the partial function, arbitrarily select one of the total orders in  $T^*$ , say  $t_1$ , as a reference order. Then, for each  $t_i \in T^*$ , define permutation  $\pi_i = (t_1/t_i)$ , and let  $\Pi = \{\pi\}$  be the set of these permutations. Set  $\Pi$  always includes the identity permutation  $(t_1/t_1)$ , as required by the partial function. Set  $\Pi$  defines the partial function over set  $T^*$  we were seeking. The groupoid is now defined as the following pair:

$$G = (T^*, \Pi). \quad (4)$$

As every groupoid, groupoid  $G$  has a unique group-theoretical block system. A *block system* is a partition of  $S$  that remains *invariant* under all the permutations in  $\Pi$ .

#### HIERARCHIES OF BLOCK SYSTEMS

In the block systems, each block is composed of ordered pairs, taken from the original partial order  $\omega$ . What happens with the ordered (cause, effect) pairs when the block system is constructed? When both cause and effect are in the same block, the pair is *encapsulated* into that block. It becomes hidden, “invisible” from outside the block. This notion corresponds, exactly, to the notion of encapsulation in object-oriented programming. When cause and effect are in two different blocks, then the ordered pair is *induced* in the block system itself, and the two blocks become one the cause and the other the effect. Hence, the block system itself can be considered as a new, smaller causal set, where the blocks are the elements and the induced pairs the partial order. The entire procedure just discussed can be repeated for this new causal set, and a new, smaller block system is obtained. The process can be continued until exhaustion, and the result is a *hierarchy of block*

*systems*. The hierarchy is determined by, and only by, the original causal set  $\Sigma$  under the axiomatic metric. The hierarchy is a *mathematical fractal*.

#### THE CENTRAL THEOREM

The Central Theorem of causal theory summarizes this entire Mathematics section as the definition of a function  $\mathcal{E}$  that maps the collection  $\mathcal{C}$  of all causal sets to the collection  $\mathcal{H}$  of all finite fractal hierarchies:

$$\mathcal{E} : \mathcal{C} \rightarrow \mathcal{H}. \quad (5)$$

Function  $\mathcal{E}$  was introduced in Pissanetzky (2012f). Hence, a unique correspondence exists from each causal set in  $\mathcal{C}$  and each fractal hierarchy in  $\mathcal{H}$ . To each causal set there corresponds a hierarchy. The notion of *observable* was generalized in Section 3. Based on this generalized definition, the Central Theorem of the Theory of Causality can be stated as follows:

*A correspondence exists from the collection of all causal sets and the collection of all observables. The correspondence is formalized by function  $\mathcal{E}$ .*

There follows a corollary. The size of set  $\mathcal{C}$ , or the total number of causal sets, is infinite countable. The size of  $\mathcal{H}$ , or total number of finite fractal hierarchies, is also infinite countable. This means that a one-one correspondence must exist, that this correspondence is effected by function  $\mathcal{E}$ , and most importantly, that nothing is left outside the central theorem. In other words, because there are no hierarchies outside of  $\mathcal{H}$  and no causal sets outside of  $\mathcal{C}$ , any causal set or any hierarchy that is found in nature is already accounted for by function  $\mathcal{E}$ . If a hierarchy is found, there is a corresponding causal set. If a causal set is found, there is a corresponding hierarchy. This corollary allows procedures such as the conversion of software into causal format. It guarantees that the conversion is possible, which is a powerful conclusion.

Many hierarchies predicted by the theorem are trivial, meaning that the only blocks are singletons, and are the elements of the causal set themselves. This means that the elements are themselves invariant observables. The proportion of non-trivial to trivial hierarchies, or the total number of non-trivial hierarchies, are important questions, but the answers are not known. The Central Theorem is the constitutive theorem of the Causal Theory.

## 4.2 Causal Dynamics and Causal Logic

The notions of interaction and transfer of information by way of ordered causal pairs has led one of us (SP) to consider causal pairs as dynamical entities (Pissanetzky, 2009), thus giving them a physical meaning beyond Mathematics. In this new context, information, hence energy, *flows* from the causes to their effects and is what makes the effects come into existence and become available to serve as causes for other effects. We refer to this dynamical process as the *execution* of a causal pair. Hence, a collection of causal pairs is an *algorithm*, and we also say that the algorithm executes when all pairs have executed and all elements of  $S$  have been visited or *initialized* by the flow of information. The notion that a causal set when considered as a dynamical entity is an algorithm answers the fundamental question about the origin of algorithms (Pissanetzky, 2012f, Sections 2.4 and 2.5). The algorithmic dynamics corresponds to the mathematical search process discussed in Section 4.1 of this paper.

The order in which the elements of  $S$  are visited or initialized by execution is described by a list of the elements of  $S$  in that order, i.e. by a legal *total order* of the elements of  $S$  – one that does not violate the partial order. This total order is known as a *trajectory*, and the dynamical space in which trajectories live is called the *causal space*. The causal space corresponds, exactly, to the mathematical set  $T$  of total orders in Section 4.1.

There may be many different trajectories in causal space, corresponding to many different total orders of set  $S$ . Any of the trajectories can be followed by the causal execution, giving rise to an indeterminacy, or *uncertainty*, which corresponds to the physical notion of *entropy*. Furthermore, the equivalence between causality and time, both of which are directional and both of which are related to the direction of entropy, has

led SP to introduce the concept of *action* for causal pairs. In Physics, there are various definitions of action, but action is always associated with “traveling energy,” energy that travels during a certain interval of time (not over a distance!), which is very much consistent with our definition, as given below in more detail.

Causal dynamics is not the same as causal logic. Causal logic is also a dynamic process, but a different one. It happens *to* the algorithm, but *is not* the algorithm, and is of a thermodynamic nature, not algorithmic. Causal logic selects trajectories according to their uncertainty, or entropy, re-organizes the algorithm, and finds a structure for it composed of *invariant observables* (Section 3). The invariant observables are the mathematical blocks of Section 4.1. Because the observables are invariant under transformations, such as translation or rotations, they are also *certain*, no longer uncertain as the original unstructured algorithm used to be. And this certainty, originated from a purely thermodynamic process, is what we call *semantics*. It represents meaning, as proposed in Pissanetzky (2012f) and previous publications. The elements of the original causal set are initially devoid of value and meaning, but the process of causal logic creates structure and is the base for semantics. And because algorithms represent behaviors, and causal logic makes the behaviors certain and invariant, these structured behaviors are intelligent behaviors.

More recently, Manfred Eigen (Eigen, 2013) proposed ideas some of which are very similar to SP’s. Eigen examines the physics of matter and the properties of material systems that lead to entropy and the picture of quantitative information initially devoid of value and meaning. He proposes *information space*, possibly very similar or the same as our causal space (the book had not been released at the time of this writing), the dynamics of which is the basis for semantics and leads to an exploration of life and thought as a new problem of Physics. These ideas, including the place of causal logic as the source of semantics, have been introduced by SP time ago. Eigen’s apparently independent ideas provide strong support to causal theory and certainly help to establish life and thought as problems of Physics. As SP has also proposed long ago.

The fact that causal logic acts on the algorithm has been interpreted as an entropic “pull” on the trajectories themselves (Pissanetzky, 2012f), the effect of which is to select trajectories and create intelligent behaviors. The net effect is that the thermodynamic force of an entropic nature acts on the trajectories and creates intelligent behavior. More recently, the same interpretation has been re-discovered and again associated with intelligent behavior, this time in a Cosmological environment (Wissner-Gross and Freer, April 2013), and renamed as Causal Entropic Forces, all of this apparently independently and under the cover of a complex formalism that hides the essential simplicity and nature of these forces.

The preceding overview of the causal theory may appear to be quite abstract. But in fact it maps very directly, we’d say surprisingly directly to familiar systems such as the brain or a computer, and actually to all dynamical systems. This map is what allows the Central Theorem of causal logic to be directly applied to any of those systems, and important conclusions, of the kind that cannot be arrived at by any other means, to be drawn immediately. An article explaining how the abstract causal space maps directly to dynamical systems traditionally understood in terms of state space or phase space, is found in Pissanetzky (2012h).

### 4.3 The Physics and Neuroscience of Causal Theory

The causal mathematical entities introduced in Section 4.1 are abstract and fundamental, but they are profoundly meaningful for both Physics and Neuroscience. In this Section, we establish a correspondence between the mathematical entities and more traditional concepts used in science. There appears to be an impressive one-one correspondence.

Causal sets are used in causal theory to represent information. Information is considered here as a dynamical system by itself, with properties such as energy, and uncertainty and entropy. Ordered causal pairs are considered as dynamic entities describing the flow of information, i.e. energy, from cause to effect. The cause-effect relationship itself is considered as time, hence giving rise to the notion of action as energy that travels from event to event during a certain period of time. Causal pairs form chains in many different combinations, causing a *combinatorial explosion*. The elements of the set correspond to variables in the dynamical system. The search process that visits the elements of the set in a certain order describes the dynamics of the system. Intermediate steps in the search where some elements have been visited and others

have not, correspond to states in state space. The chains of causal pairs correspond to the trajectories of a dynamic system in state space and each pair corresponds to a transition from state to state. Hence, a causal set alone provides a rich mathematical model of a dynamical system in causal space, that directly corresponds to a more traditional model of the same system in state space or phase space. In this causal model, we discuss the trajectories first.

#### DYNAMICS, STATES, AND TRAJECTORIES

Let  $T$  be the set of legal total orders of  $S$  under partial order  $\omega$ . Set  $T$  is very important because it does several things. Set  $T$  lists all possible trajectories of execution of the causal set in causal space. Alternatively, and using a term familiar to software developers, it can also be said that  $T$  lists all possible *refactorings* of the equivalent algorithm.

#### SYMMETRY AND CONSERVATION

Set  $T$  also describes the *symmetry* of causal set  $\Sigma$ . In Physics, we say that a physical system has a *symmetry of the action* when it can be described in more than one way in the mathematical model. The total orders of  $S$  listed in  $T$  are, precisely, all the ways that causal set  $\Sigma$  can be described. Hence, we have an important conclusion:

*Every causal set with more than one legal total order has a symmetry of the action.*

There follows that every causal physical system has a symmetry of the action, although the symmetry may be trivial in some cases, such as a total order or no order. Set  $T$  formalizes the uncertainty/entropy value of the trajectories, giving rise to the laws of Thermodynamics. Set  $T$  also gives rise to a groupoid, and groupoids have invariant observables. The existence of set  $T$  proves the second fundamental principle of Physics, the principle of symmetry, that states that any system with a symmetry of the action also has conserved quantities. Hence:

*Every causal set with more than one legal total order has conserved quantities, determined as hierarchies of invariant blocks.*

A finite set  $S$  with  $n$  elements has no order, and can be represented mathematically in  $n!$  different ways, the  $n!$  permutations of the  $n$  elements of that set. In this case, there are no non-trivial invariant blocks, as no two elements can stick together in all permutations. All block systems are trivial and the invariant quantities are the elements of the set themselves.

When a partial order  $\omega$  is added to  $S$ , then the partial order partially *breaks* the original symmetry, leaving a lower symmetry. The lower symmetry can give rise to both trivial and non-trivial invariant blocks. The proportion of non-trivial to trivial is not known, and is an important question that will have to be theoretically addressed. However, judging by the abundance of invariant observables in life and thought, the proportion must favor non-trivial invariants very strongly. There must be a very large number of them, but whether that number is infinite is not known. This conclusion can be weighted in relation with a generally accepted “law” of Physics: the discovery of one non-trivial and hence physically meaningful invariant equals one Nobel prize.

#### BINDING AND MASSIVE PARALLELISM

The fact that metric  $\mathcal{M}(t)$  in Eq. (3), where  $t$  is a total order of  $S$ , is positive definite, is critical, because it solves the famous *binding problem*. Metric  $\mathcal{M}(t)$  is the sum of many terms, each of which is a positive number. But  $\mathcal{M}(t)$  is also the action functional. When the principle of least-action is applied and action is minimized, the only way to minimize  $\mathcal{M}(t)$  is to minimize each term in Eq. (3) *independently* of the others. What does this mean for the brain? It explains the proverbial “massive parallelism” of the brain, that has never been fully explained. It means that each neuron can *locally* minimize its own contribution to the action by making its own connections optimally short, using only local information available to the neuron,

without any external direction, and by “pushing” its neighbors as needed to optimize its local morphology. It means that all neurons work together, simultaneously, guided by Thermodynamic rules alone and not explicitly programmed to create any particular behavior, much less an intelligent or adaptive behavior. It means that intelligence is a *side effect* of Thermodynamics, guided by forces known as *causal entropic forces* that originate in causality and entropy. These forces can act in both directions. They have been called “pulls” or “net pulls” (see Pissanetzky, 2012f, §4.2 and 4.3). They have been more recently independently confirmed in Cosmology (Wissner-Gross and Freer, April 2013).

The same considerations can be applied to a computer, and the same results can be obtained, including massive parallelism and intelligent behavior. SP has proposed a *causal virtual machine* (Pissanetzky, 2012f, § 4.1), where the same principles are applied and the same results are obtained, including massive parallelism and intelligent behavior. “Neurons” are represented as simple processors, and “push each other” is described as “commutation” between adjacent neurons.

The distribution of populations in the metric slots plays a very important role in the causal theory: non-trivial invariants exist preferably in slots with small populations. Hence, each block is an algorithm, or behavior. These blocks, we propose, appear in our everyday life. The term “invariant” means that they do not change, they remain the same in some sense in our complex environment. We recognize them, we assign names to them, they have *meaning* for us, although the original causal set  $\Sigma$  that represents our knowledge of the world had no meaning at all. They define a universal *semantics* determined by causality and entropy alone, not by “us.” We assign names to them. They are the words of our languages, they correspond to neural cliques in the brain (Lin, Osan, and Tsien, 2006), they are Fuster’s cognits (Fuster, 2005), Jeff Hawkins’ invariant representations (Hawkins, 2004), the conserved quantities that underpin all natural sciences. Blocks are the adaptive behaviors and the thoughts and ideas that help us to survive. The self-organized structures observed in many dynamical systems are blocks. The brain waves used as life signals because they are observable and we can recognize them are blocks. Many other types of blocks are also observable and recognizable and can be used as life signals. Such as ‘mother’ in the Hofstadter example. Or so we propose.

We propose all that as part of our working hypothesis. Because invariant blocks solve the uncertainty problem. Least-action removes entropy, and removes uncertainty, leaving the blocks that are certain and are needed to make predictions. Block systems are well understood in group theory, and easy to compute from a given groupoid. These considerations explain the origin and abundance of fractal structures in nature.

The hierarchy represents *compression* of knowledge all the way to the top of the hierarchy, and makes memory associative. Shared hierarchies allow us to communicate. ‘Mother’ in the Hofstadter example is the top of a hierarchy. Hierarchies tend to be compact, but they may also be distributed as networks all over the brain.

## LIFE SIGNALS

In Neuroscience, *life signals* such as electromagnetic brain waves and oscillations are used to detect brain function. Computer simulations/emulations of the brain that have not been explicitly programmed to reproduce these signals are expected to reproduce them anyway. Waves and oscillations are used because they are *externally observable* and hence serve to judge the quality of the simulations. Here we propose that waves and oscillations are just one example of an enormous variety of life signals, the blocks in the block systems, such as Hofstadter’s ‘mother’. These blocks are observables, but generally not amenable to easy external detection.

As discussed above, blocks are algorithms that control the external phenomena that we observe. At the same time, we expect those algorithms to be somehow generated by a simulation *where they have not been explicitly programmed*. In Computer Science, this is known as *self-programming*. And self-programming has been explicitly covered in Pissanetzky (2012f), where it has been argued that algorithms originate precisely from the same ordered (cause, effect) pairs that we are discussing here and that we receive via our senses, or the computer receives via its input, or a robot via its sensors. Coincidence? No. This is a very strong confirmation that we are on the right path.

Black-box experiments are central to addressing the issue of external observation. In the causal theory, observables - that is, algorithms, including waves and oscillations, are generated by Thermodynamics, not by computer program. The causal virtual machine we propose generates its algorithms thermodynamically. In a computer simulation of the virtual machine, there is no program, nothing needs to be programmed that is related to the generation of the observables. There is of course a program for input and output, visualization, and for minimizing the action functional. But the life signals emerge as a side effect when the functional is minimized, and not by program. Black-box experiments bridge the gap between internal life signals and externally observable life signals.

We hereby propose that a concerted research program is needed, based in black-box brain experiments, specifically intended to bridge the gap between internal life signals and external observables.

#### PRACTICAL USE OF THE THEORY

The Mathematics of the theory seems to be complicated, but its application to practical problems is really very simple. The first step is to construct the causal set model of the system of interest from input. The input can be a statement describing the system, a stream of data coming from sensors, or any other source. Many possible sources have been discussed in Section 3 of Pissanetzky (2012f). If the system has a memory, then any information stored in the memory has to be considered as part of the input. The total input consists of all information that's available to CML, irrespective of its source. CML's function  $\mathcal{E}$  takes as input a snapshot of the total input and returns the hierarchies corresponding to that snapshot. CML is ongoing. It functions as a movie camera. It constantly composes new snapshots and produces a movie of the evolution of the structures. This is how we see objects that move or break or change shape as they move.

There is no need to describe the system in full detail. Quite on the contrary, it is recommended not to. It is better to use some coarse information to start the model and gain some experience. Later, additional information can be entered to refine the model. After all, making changes in CML is very easy.

Causal modeling is very different from software development. In software development, the need to provide *all* information is emphasized, the reason being that a usually large team of developers, analysts, and managers is awaiting the information to begin a long and complicated process where "objects" are designed by the human analysts, architectures are proposed by designers, and software developers convert them into code, which is then tested and adjusted for missing features through intense interaction with subject-matter engineers. Any changes of the initial information are considered a major problem, and placed under change management to assess the additional cost, giving rise to a very costly iteration.

Nothing of this happens in causal modeling, because the objects, architectures and code are automatically created by the causal machine. The output consists of ready-to-use functionally specific algorithms. "Functionally specific" means focused on a specific task. There is nothing to adjust at that level. Given the causal set the process follows as a consequence of the causality. The iteration takes place at the level of input, and any new information or changes are simply added to the model for the purpose of refining it. The important part is that the input must be expressed in causal set format, no matter what.

The task of refining a causal model is very much like teaching a child. First, a teacher provides some knowledge to the child. The knowledge usually comes as an algorithm, i.e. in causal form. Then she asks the child to solve some problem where that knowledge is necessary. If the child fails, the teacher may detect what is missing and teach the child some more. The child can also do it on his own, by asking questions or playing with toys. Given the means, the child can acquire the missing information on his own. All this is possible because the child has the capacity to integrate additional information automatically, without the need for a software developer.

With a causal machine, it works the same. The machine is trained with some initial information. The machine can ask questions, see for example Section V.F "Supervised Learning" in Pissanetzky (2009). If the machine has sensors, then it can acquire additional causal information directly from the environment. If the machine has actuators, then it can play with toys, and "see" what happens. One day, causal machines will be able to learn from each other, just like we humans do. There is one single reason why all this is possible with

a causal machine but not with an ordinary computer: the machine has the capability to bind and integrate information from different sources and create the focused algorithms of adaptive behavior.

In some cases, for example when the human force consists entirely of software developers untrained in causal modeling, or when a computer model of the system already exists and needs to be integrated with other programs and causal models, it is possible to convert software directly to causal set format. This possibility is discussed in more detail in Section 10.1, where a graphics user interface (GUI) is proposed for general causal modeling work and visualization.

Several examples and case studies in causal modeling have been published, see for example Pissanetzky (2012a,g,d) in the supplementary material. In Pissanetzky (2012c), a study involving hundreds of small causal system has been published.

## 5. Causal Information

For the purpose of confronting the AGI problem we have adopted the working hypothesis that intelligence is of this world and must be explained in the context of a theory of Physics grounded on the supreme principle of nature, the fundamental principle of causality. We also proposed that other fundamental principles, symmetry and least-action, can be derived from the principle of causality (Pissanetzky, 2013), and in Section 2 we argued that information should be considered as a physical entity.

### THE THIRD MEMORY

We have insisted on our views that the brain is the only intelligent system we know and we can observe, and that in some sense the brain is not only an example of intelligence but also a measure for it. In particular, the brain's memory has been recognized as necessary for prediction and intelligence and the subject for studies and imitation. However, These studies traditionally focus on just two types of memory, the species-wide DNA, and the individual-specific learned memory of the connectome. We propose that this context is too narrow for intelligence. We propose this on the grounds that intelligence is wider than a species – who would say that a dog is not intelligent? – and even wider than life. In fact, phenomena of intelligence share many features in common with phenomena of self-organization and emergence, observed in inanimate systems and studied in Complexity Science. We propose that a Third Memory is necessary to account for such generality, that is universal and not inherited in DNA or stored in connectomes, and that this third memory is matter itself. In fact, matter acts as a memory inherited by all things material, where all the properties of matter are stored, and used to determine its behaviors – the phenomena of nature. The causal theory incorporates the third memory explicitly at the precise point where the principle of causality is incorporated. At this time, the causal theory appears to be the only theory to recognize matter as playing a role in intelligence.

### INFORMATION IS DISCRETE AND CAUSAL

Information in nature is discrete and finite, and is best represented as a collection of ordered cause-effect pairs. In Mathematics and Physics, such a collection of ordered pairs is known as a causal set, defined as a finite partially ordered set that satisfies some additional conditions. Our work is entirely described in terms of causal sets. We use causal sets because they are essential for understanding information, they provide a universal representation and communications protocol for nature, and they are naturally recursive. When we learn something we acquire information that is naturally causal. Being causal means that effects follow their causes (Gell-Mann, 1994). It does not mean that all causes are necessarily known, it only means that they exist. Information arriving in our brains via afferent nerves or biochemical signals is causal. Every process that happens in the brain and affects information in any way is causal. Even neurons that spike randomly are causal because random only means that the cause is not known, or that we don't care to know, not that it does not exist. Signals sent from the brain to motor nerves and other efferent nerves are causal. The universe is causal, and all information that flows in it is causal. The origins of the universe have been traced all the way back to the Big Bang. Many physicists believe that causality and time are one and the same. We know for a fact that time is the same everywhere in the universe, except in black holes, and we know for a fact that

causality is a fundamental principle, also everywhere but in black holes. The two facts are one and the same. The universe is causal, and so also are information and brains.

#### CAUSAL INFORMATION AND THE TURING MACHINE

A causal set, defined in Section 4.1 as the pair  $\Sigma = (S, \omega)$ , is a Turing machine. There exists a one-one correspondence between the two. Causality represented by the partial order  $\omega$  is the algorithm, the transition function of the Turing machine. Set  $S$  is the data that  $\omega$  operates upon, the string written on the tape of the Turing machine. There is an important difference, though: unlike the totally ordered Turing string, set  $S$  has no order. This is because  $\omega$  refers to data by name, not by position. The causal set is like a programming language, where names and not memory addresses are used to identify data. The causal set is also a virtual machine, one that is universal and can be mathematically manipulated, unlike most other virtual machines. The two ingredients, the Turing machine/causal set equivalence, and the fundamental principle of causality represented by the causal set, may be the two ingredients needed for a proof of the universality of the Turing machine, but this proof is not attempted in the paper.

When Turing introduced the use of a string to represent data, he separated reality into two parts, the string with the data, and the causality that controls them, and left the causality part to be taken care of by human developers. Later, the von Neumann stored-program architecture put data and instructions in the same memory, but the human developers remained in charge of the instructions. This was good for efficient computation, but left the humans in charge and in the loop. Shannon’s information theory followed suit. Data is stored in a string, humans deal with causality. A shared communications protocol is assumed to have been agreed upon by the humans, for example a TV set comes with a controller that tells it what to do with the incoming stream of data. The TV signal alone can not produce a picture. The shared protocol is good for efficient transmission, but humans were again left in the loop and the machine as their slave.

But the brain does not work like a Turing machine. There is no “human in the loop” to take care of causality and write a transition function, and an intelligent machine with a human telling it how to be intelligent would not be intelligent. It is necessary to remove humans from the loop. Completely, not some of them.

#### CAUSAL INFORMATION AND THE BRAIN

Information is not processed in the brain the same way it is processed in a computer. There are three different processes going on in the brain. The three processes are concurrent and collocated. The first is the formation of the connectome from highly entropic information that arrives in the brain in the form of unrelated causal fragments from senses and other afferent nerves. We do not discuss this process here.

The second process, is the formation of short dendrites. It doesn’t matter whether they are made short or somehow become short. The fact is that they are short. This is a mechanical process, but it has a thermodynamic effect on the information stored in the connectome: it minimizes the action in it by removing energy. With the energy, also go the entropy and uncertainty, and the net result is that both the connectome and the information it contains get self-organized. The connectome self-organizes physically, forming hierarchies of neural cliques (Lin, Osan, and Tsien, 2006). The information self-organizes logically, forming hierarchies of structures. This is, we propose, how CML works in the brain. Critical to add credibility to our proposal is to verify that the dendrites are optimally short. They are, indeed (Cuntz, Mathy, and Häusser, June 2012).

The third process is the execution of the algorithm. The causal set *is* the algorithm, there is no need to “invent” one. The causal information arriving from the senses in fragments is the algorithm, and CML does not create algorithms. CML associates like fragments together by forming the block systems. This association or *binding* happens because the mechanical process of shrinking the dendrites has removed entropy and uncertainty from the information, and has finally generated the *invariants* needed for semantics, i.e. for us to be able to comprehend and manipulate the information. These invariant algorithms are the behaviors we call *adaptive*. They execute when the neurons spike, and the chain of spikes sends signals to the muscles. It is all automated, no humans in the loop.



The critical question in this Section is of course: Can we do the same on a Turing computer? If we have removed the human from the loop, but the Turing machine requires a man-made transition function, what do we do about it? Easy: CML is the transition function, as described by  $\omega$  in the causal set, and  $S$  is the Turing string. And CML is not man-made. This is a *non-explicitly programmed* Turing computer, or virtual machine, as described in Section 4.3 of Pissanetzky (2012f), where the transition function is mathematical, and the data is directly acquired from senses, or sensors in the case of the machine. It can also be viewed as a ZISC (zero instruction set architecture), as invented by Guy Paillet and used for pattern matching. Like the CML virtual machine, the ZISC architecture allows a constant execution time regardless of the number of processors connected in parallel.

In Computer Engineering, the process of binding is known as *refactoring*. Refactoring, and Object-Oriented design, are done by human developers. Attempts at automating these tasks were many, but were not successful, and as of today only humans can do it. Not a surprise, after reading what precedes.

## ON GOALS

In Engineering, when a goal is prescribed, a causal path is frequently manually assembled from existing fragments of causal paths, trying to find a complete path that reaches the prescribed goal. If a gap is found, as is frequently the case, then the goal changes and much of the engineering effort is concentrated into closing the gap by searching for the missing causal fragment. This search is not just a computational search, it may involve research and experiments. When the causal path is complete, then the goal is achievable, and the path is the algorithm that allows the prescribed goal to be reached. The algorithm may be a computational algorithm, but it also could be an industrial process.

By contrast, in Science Fiction, we fictionally assume that the gaps have been bridged and concentrate our attention on the consequences.

But in the causal theory, goals are viewed very differently. If a goal is a state that can be reached from some initial state by following an existing causal path, then all states found in the causal space are goals, except the initial states, the ones that have no known predecessors. Hence, the notion of goal-based agents is not necessary. Any state in the causal space can be designated as “the goal”, and decisions necessary to reach it can be found by goal-directed reasoning going backwards from the designated goal. If, however, the desired goal is not in the causal space, then the goal changes to that of expanding the causal space by learning more information. Which may involve research, observation, and experiments to determine the missing causality. Once the causal path is complete, then the original goal is recovered.

The theory of causality describes information as a multiple preset goal system. Or, to say it differently, as a single goal system, defined as that of minimizing the action, that leads to the creation of multiple goals as all possible final states of the system’s evolution. This feature is described in more in detail in Pissanetzky (2012h). The notion of goal, as used in AI or Engineering, does not exist in the causal theory.

## PART II

Part II attempts to establish CML as a theory of the brain, under the working hypothesis that the brain is physical and intelligence should, and can be explained by Physics. The argument in Part II posits that the brain in its capacity as an information processing device is a causal thermodynamic device, fundamentally different from any Turing computer, and cannot be studied or understood in the context of the Turing theory of Computation unless causality and thermodynamics are appropriately simulated. Also argued in Part II is that any physico-mathematical description of the brain must be carried out in the context of a theory that supports full detail, rather than other theories of Physics or of the brain where simplifying assumptions are made, such as statistical distributions for statistical methods or assumptions of smoothness and differentiability for differential methods.

## 6. A Causal Model of the Brain

We argued above that the brain is an organ and that its behavior is causal and analog, just like any other organ in the body. As an organ, it works continuously, and we are not aware of its doings any more than we are of what our liver is doing. The organ part corresponds to the unconscious, or to the host in a host-guest model (Pissanetzky, 2010). The organ part is of a causal thermodynamic nature. Its task is to organize the guest information by associating causal pairs and making the memory associative. It encompasses the entire brain, every neuron is constantly trying to associate causal elements represented by its dendritic connections. Neuron spikes are not involved. This task is carried out constantly, without effort, and without us being aware of it or being capable of controlling it. It corresponds to the legendary perception of the brain as a massively parallel “computer,” except that “massively parallel organ” would be more accurate. This task is not computation, it is not related to computation, and is fundamentally different from any Turing machine. However, the connectome itself, when considered as a network, is an algorithm, mechanically segregated from the causal input by the binding process, where the neural cliques correspond to behaviors. When the neurons fire, the behaviors execute, for example the motor signals that pronounce “mother” in the example above are sent to the muscles.

The task of binding is absolutely essential for intelligence. It is the core engine that generates a *global response* to the local actions of each one of the billions of neurons that work simultaneously, locally, and in complete isolation except for their immediate connections, but without any awareness of the global consequences of their actions. The globalization of the local action is formalized by Eq. (3). This equation is the critical detail that has been missing in Science so far. It expresses a global effect - the optimization of the global action functional - as the algebraic sum of an undetermined and possibly very large number of positive independent contributions, coming from independent parts that have only been programmed by evolution to perform a simple local association without any external director. Eq. (3) is what allows the human brain to be so large.

The globalization of the local action propagates causal association to the entire brain. This phenomenon is known as *binding*. It has a place in history. The *binding problem* was started by Bertrand Russell over a century ago, but remained unsolved until now. The failure to solve the problem for so long is directly traceable to the omission of causality. The solution to the problem of binding was proposed in the context of CML theory. It was also proposed that binding is a universal phenomenon, and is critically necessary for intelligence because it allows the globalization of associations that would otherwise remain local and the generation of global behaviors that respond to global conditions. But binding by itself is not sufficient to produce intelligent behavior. Phenomena of emergence and self-organization observed in many complex systems that aren't intelligent are also due to binding. Necessary and sufficient conditions for intelligence are discussed in Section 7.

It has been proposed that, as global binding happens “unintentionally,” in the sense that evolution only intended to preserve resources without sacrificing memory and not to create intelligence, it would then be fair to say that intelligence is a thermodynamic side effect of resource preservation (Pissanetzky, 2011a). Of which, of course, evolution has taken after-the-fact advantage.

The Eureka effect, also known as the aha! effect, refers to the common human experience of suddenly understanding a previously incomprehensible problem or concept. We say we “have an idea” when an algorithm is delivered from unconscious to conscious. We feel a sense of surprise, because we now “know” the solution that we didn't expect. We say this at the precise point where we feel “ready” to act on that idea, and we feel ready to act because the binding has finished and the connectome for our algorithm is complete. There is no “transmission” of bound information from unconscious to conscious. It happens in place. The associative memory itself serves both as a substrate for the unconscious binding and the conscious algorithm, including the neural cliques, and cliques of cliques. The notion of having an idea, or suddenly understanding something or discovering something occurs when the host finishes its work and feeds the results to the guest. That is, when the unconscious thermodynamic process is complete and its output is passed to conscious.

There is no need for any separate bootstrap mechanism or for the presence of any loop that never ends and keeps the brain working. When the brain is represented as a causal set, there is an input and an output

of information, as discussed above. The input is “all the information that enters the brain,” irrespective of the origin of that information or of the “meaning” we may want to attach to it. Similarly, the output is all the information that the brain puts out, irrespective of its destination. But these considerations also imply that external feedback loops automatically arise, as a consequence of the interaction of the brain with its environment. There can be a large numbers of such loops. For example, consider the sensory-motor loop. A baby shakes her arms and legs as directed by output from her brain. But the baby can also see and feel the arms and legs as they move, and generate input for the brain. This is an external feedback process. The brain stores initial input, then binds it with additional input coming from the external sensory-motor loop, and creates additional output for the limbs. The resulting feedback loop results in sensory-motor integration, where the baby starts with, perhaps, random movements and progressively learns how to control her limbs, how to grab, etc. Another loop such as this may be responsible for eye-movement coordination. Of course, there is no closed-form mathematical proof of our statements. They can only be verified by construction, and this task needs black-box experiments.

The brain is known to contain many internal feedback loops. We can’t explain exactly what their function is, but it would be perfectly possible, as far as this theory is concerned, that they merely serve the purpose of reusing resources in a way similar to how computers reuse resources by “calling” subroutines or methods or instantiating objects of a common class to avoid the repetition of code.

## 7. The Conditions for Intelligence

The phenomenon of binding, defined in Section 4.3 as the generation and globalization of local associations of elements of information, is by itself not sufficient for intelligence. There are other conditions that need to be satisfied before a complex system exhibits intelligent or adaptive behavior.

There are three conditions that need to be satisfied in order for a dynamical system - any dynamical system - to exhibit intelligent and adaptive behavior. The first condition is that the causal space must be large, and the larger it is the more intelligent the system will be. The causal space is the collection of all state variables that describe the dynamics of the system. The state variables hold information and represent the memory of the system, and a state of the system is the state of the memory at any particular instant of time. The larger the causal space the larger the memory and the larger the complexity and degree of detail of the patterns it can create and hold. This condition appears to be satisfied quite well among known adaptive systems, with motor proteins at the bottom, followed with worms and other animals, apes, and humans. It is generally recognized that the bigger the brain the more intelligent the animal is, although of course differences in implementations cause fluctuations.

The second condition is that the memory must be autobiographical, where information acquired by interaction with the environment is stored and does not “disappear” by being erased or overwritten by new information. In a dynamical system such as a hot gas in a bottle, the causal space is enormously large, but the gas is not intelligent because states get overwritten: two molecules in a certain state collide and create a new state, but keep no memory whatsoever of the previous one. By contrast, the brain creates new variables for new states, rather than overwriting existing ones. The brain creates new state variables by constantly making new connections to store learned information, and keeping all memories for a very long time.

All three conditions are necessary, but the third one is the most critical: it is the requirement that an external process must exist that will optimize action and bind the information being processed. A system such as this, where two different processes coexist and interact, is known as a *host-guest* system (Pissanetzky, 2010). The term *substrate* is sometimes used to designate the host (Koene, Hutter, and Hales, October 2012). When information is learned by way of the system interacting with its environment and growing its memory, the acquired information carries energy and entropy with it. Entropy is a measure of the degree of uncertainty in the information. As natural systems constantly learn and acquire more and more information, more and more energy, entropy and uncertainty accumulate in the system, making it “hotter” and more uncertain and disorganized, and the need arises for the external process to constantly operate and remove excess energy. This process represents the *fundamental principle of least-action*, well known in Physics. When action is minimized, excess energy and entropy are removed from the information, the information becomes more

certain and better organized. Patterns arise when uncertainty is removed and the information is invariant. The process that minimizes action is known as causal logic (Pissanetzky, 2012f), and an implementation of causal logic on hardware or software is known as a *causal processor*. A causal processor is a thermodynamic device, not a computer, but it can be simulated on a computer or on specially designed hardware. Once production starts, causal processors will be small and inexpensive devices. They will be rated by problem size, with execution time independent of the size of the problem, and will be stackable for larger problems without increasing the time interval needed for completion. More details are covered in Section 10 of the present paper.

A *scaling problem* is the classical symptom that diagnoses the accumulation of entropy in dynamical systems that do not incorporate a causal processor. The system appears to “work” for small problems but fails at a larger scale because the excess entropy and uncertainty are not being removed. Scaling problems are very common in AI systems.

As an example of application of these ideas we propose the prediction made in 2011 (see Pissanetzky, 2012c, § 3.8), that the length of dendritic trees in the brain should be as short as possible. The prediction was made under the single assumption that the brain as an organ uses CML to create intelligent and adaptive behaviors, and at a time where a non-optimal  $4/3$  power law was the accepted norm in Neuroscience. The prediction was independently confirmed in 2012 by Cuntz, Mathy, and Häusser (June 2012), who proposed an optimally short  $2/3$  power law with a wide experimental base that includes humans and other species. This is a remarkable, very early success of the theory, and demonstrates its potential power and usefulness as a research tool in Neuroscience. However, it does not prove by itself that causal logic actually exists in the brain, or how it is implemented if it does. That task, is for Neuroscience to undertake. The profound consequence of the discovery of CML illustrates very clearly the nature of the contribution our work can make to Neuroscience: to provide a theoretical context within which neuroscientists can place their discoveries and relate them to one another, and help Neuroscience to become a more integrated and less descriptive science.

## 8. The Neuroscience Application of Causal Theory

The generation of millions of papers in Neuroscience working on various parts or levels within the brain has helped us to define most of its circuits and many of the higher level functions and principles. Due to the complexity of the system we have arrived at an impasse in regards to how far we can proceed to complete understanding without large scale integrative Neuroscience. Current projects now emerging for the ground work in whole Brain simulation or aspects of it, are: Human Brain Project (HBP), the US BRAIN project, Darpa Synapse, Brain Corporation, Neurogrid, Allen Brain Atlas and the IBM ultrascaleable tissue simulator. With current technology these will be restrained in regards to how much of the brain’s detail can be simulated. Currently whole cell models are able to derive a replication of the phenotype from genotype in entire genetic detail using physical principles (SimTK framework). By contrast, whole brain emulation (WBE) aims at a more efficient route to acquire the minimal requirements to run a workable copy of our conscious operations with as few of these detailed constraints as possible.

We will briefly summarize the difference between these projects, and why the area of neural replication needs to be clarified as having a validity to bridge the gap between Simulation and AGI. Also we will attempt to clarify why AGI is directly relevant to WBE or brain simulation at all. Most importantly we aim to present in a more sequential method the way in which all these disciplines will be required for the success of WBE, and how a natural theory of information is both consistent with current Neuroscience, and may possess a proposed explanatory power for Neuroscience discoveries not attainable by other means.

### 8.1 The Difference between Brain Simulation, Emulation, and Replication

*Whole brain emulation* (WBE), has as its priority to take a snapshot of the mind in whatever detail is required and try to engineer a workable running system with translation of the system onto a substrate more robust than DNA. The problem of acquiring the WBE data is hard so is a priority. WBE aims to get the copy to run as it was, i.e. remember its past, and have conscious input/output processes. For the copy to be able

to proceed with evolving its own intelligence, it may be preferable to have the principles of intelligence to simplify the entire process. So for such an aim, there is often dialogue between WBE and AGI.

*Brain Simulation.* Henry Markrams Blue Brain or Human Brain Project (HBP) is a current primary example of simulation. HBP does not have an agenda for WBE scanning. It is interested in biological detail and a general system model. To achieve aims with current technology there is a cut-off limit in regards to specific details of cellular replication. HBP also proposes to simulate the brain's lifespan from neuro-development to ageing. This is one step from a generic bootstrap intelligence model with a causal direction, and it is hoped we might see generic principles of intelligence emerge without explicitly programming them.

*Replication.* At its purest level, replication seeks to recreate the fundamental physical conditions of the brain in our natural environment to explore the possibility of discovering the deepest principles of intelligence. It is not currently feasible to build a physical copy of the brain so replication projects are minor in scale. A Neuroscience example is Colin Hales aims to develop his work on ion channels (Hales, Grayden, and Quiney, 2011) into physical replication experiments. Hales general view also aims to find a similar natural law for information, in the manner laid out in Hales (2011).

The difference between this pure replication approach and a simulated replication as proposed here will be clarified further. As replication tends to be based on the fundamental principles of the natural world, it would be useful to develop a classification between simulation and replication which we will call SimREP. SimREP would be the neuroscience analogy to Simbios and SimTK currently used for cells, which rely heavily on physical principles (Karr et al., 2012). Although we view SimREP as theoretically foundational, we cannot be sure at this stage if it will be as physically foundational as replication experiments or pure science theories like CML. Pure replication, does not have to represent every detail, but proposes we will benefit from staying on track with the most general laws of physics, as we can derive principles on intelligence directly from those which can translate across substrates like DNA into fully complex systems.

SimREP will be proposed as a CML-type foundational basis to help determine whether WBE is solid or has a foundational problem. Theoretically SimREP asks us to be open to the conclusion we might require different software and hardware platforms, to our current tools, at least for WBE to be capable of operating as an intelligent entity it may have to be, but the conclusion to this question is as unknown as the current state of physics. In summary, simulation, replication, emulation projects (and in between categories) have different priorities and roadmaps, which crossover with each other. It is preferable for the efficiency of WBE if the requirement for a SimREP platform, or even many facets of neuroscience are found to be unnecessary. A summary in regards to what we might be able to strip out from the biological substrate are covered elsewhere (Deca, 2011).

So currently all these approaches are on the table and may support each other in various ways. First of all we need to justify CML for Neuroscience in the first place, and the validity of SimREP as a concept. This will be attempted by determining if CML is both general enough to provide a description of the natural world, and from there review if we can determine its consistency with the basic components in intelligence signals that are sought to emerge in contemporary brain simulations. i.e. self oscillation and spreading waves. We will propose it has explanatory power to describe well in natural terms, the integration of these components where we define perceptual markers of mammalian intelligence in terms of the most emergent and complex signals in neuroscience which are the spectrum of event related potentials (ERP). See Lanzalaco and Pissanetzky (2013) for more details.

## 8.2 Why do we need a Causal Mathematical Language in Neuroscience?

A natural theory of information has to be successful in the translation across many substrates. E.g. from fundamental physics - DNA - phenotype - simulation. The first issue that is put forward in CML is based on overtly simple physics, while a proposed application such as brain simulation project has considerable complexity. First of all the scope of CML is such that it does not miss in between stages across substrates, as the physical definitions are operational both in live physical actions (such as operation of cells, or a brain). They are also operating within the information space of discrete substrates such as DNA, as well as the in-between stages of the transition between genotype and phenotype, such as epigenetics and principles of

development. To support this perspective, pioneers in biology have proposed a growing body of work that such causal physical principles are to be found operational within the genetic information space (Gjuvsland et al., 2013; Noble, 2012; Eigen, 2013). As noted previously, SimRep type projects active in cellular biology are also forced to simplify their computational domains by deferring to natural physical principles.

Currently we can only guess how such a scheme would operate through the principles of genetics. But these principles do have a causal style. Genetic crossover does integrate a causal set of genes with symmetrical rules, giving rise to a hierarchy with reduced entropy, e.g. least action is the effect of the cell, organism, mind etc., gaining fitness as efficiency in the environment. The reproduction cycle and crossover (a cyclic permutation inducing symmetry) selects the most efficient breeds for system re-integration, which then removes entropy from that system, giving a causal direction to the evolutionary process. So CML although it appears simple is proposed as a complexity theory by codifying natural laws of physics applied to information. This would in theory, give rise to a complex evolutionary process when setup to run within an information space where it can self modify its own hardware. With the current limited resources there has been some preliminary success in the application to general Neuroscience (Lanzalaco and Pissanetzky, 2013) by means of review and in unraveling the high level processes of the mind by experiments (Pissanetzky, 2009, 2011a, 2012f). More recently this method has been applied to physico-mathematical problems that involve the application of high brain function (Pissanetzky, 2012d). Also, as has been pointed out by the reviewer, the general concept of signals and systems, and that systems can be described in terms of their effects on signals is a well-known method that can be applied at all levels. A possible equivalence of approaches is certainly hinted at in Wallace (2013).

### 8.3 The Physics of Causal Theory in Neuroscience

Here we will summarize current work (Lanzalaco and Pissanetzky, 2013) that contains more neuroscientific detail about how it is possible to propose that CML theory has a solid consistency with Neuroscience, and help to complete halted projects where the progress had, as is common in Neuroscience, fragmented to a dead end.

This is not a selective process, where we take parts of Neuroscience to fit our theory. The CML theory and its principles appears to make sense of the most general aspects of Neuroscience that are currently agreed on, but the integration of these to a complete theory for the mind had remained incoherent. Its not proposed this is the final solution either. There may still be deeper general aspects of neuroscience to be found, so we can not be sure, but at this stage the mammalian system has been taken apart and studied repeatedly on so many levels it can be proposed we have a good general idea of its input and output processes. There are still macroscopic unknowns such as the glial system and principles of neuro-development for which a covering attempt is produced in light of current data (Lanzalaco and Pissanetzky, 2013). As the sub-neuron detail and genetic substrate are still not completely mapped there is less certainty and we have to remain open to concepts such as neural coding factors that could affect structural conformity in the neurons cytoskeleton. There may also be wider ranging genetic and hormonal aspects effect on the brains primary components not yet known.

The general areas of Neuroscience where we focused on, were the set of the highest level “intelligence signals”, the event related potentials (ERPs). ERPs are proposed as such signals as they are associated with both information and the complex processing of it. ERP signals represent single sensory modalities associated with cortical processes, such as response inhibition, response conflict, error monitoring, novelty detection, object recognition and intentional deviation. Multi-modal integration such as context updating, action selection, perception stabilization, maintenance in working memory, syntax processing and generation of expectancies that are associated with conscious awareness (Lanzalaco and Pissanetzky, 2013). So ERPs are our current biophysical markers for the general intelligence processes in terms of how we can measure intelligence from the systems input and outputs.

What are the ERPs composed of? This is a complex area of Neuroscience. For more easy terms for cross discipline use here they contain a spectrum of two extremes of signal (spreading waves and oscillations) for which there are sets and coalitions of different frequencies of these, as well as signal types where the

line is blurred between these two extremes. The more detailed analysis for this area of ERP (Lanzalaco and Pissanetzky, 2013) revealed the thermodynamic theory and the associated definitions with CML are surprisingly consistent across the translation from natural mathematical language to our full Neuroscience complexity. Such a concept is not controversial and was originally started by Papo (2013), who sought to define the entire system thermodynamically in terms of its resting state. The action functional is really now just the proposed current implementation attempt that tries to codify the principles of causality, least action, symmetry and thermodynamics to information for the entire system. It is very foundational, yet the same predictions for Neuroscience to both authors arrived later. The power law for Cuntz neurons (least action) was predicted as a result from the minimization of action within the block system. Coupled oscillations in the brain are a function of entropy which are described by cyclic permutations, where the use of the cyclic permutation in the functional predated these conclusions (Lanzalaco and Pissanetzky, 2013).

The overall integration of these extremes within ERP was then derived as a larger scale thermodynamic “information engine” structure in terms of the glia which develop the mammalian brain. Such a finding might have predictive power for the questions in glia that surround WBE (Deca, 2011). It is early days for such a broad and controversial approach, but in essence if agreement is found, an integrated theory of the brain is proposed to have been derived from the natural mathematics laid out in CML that is consistent with the generally accepted details of the mammalian brains input and output.

### PART III

In Part III, the notion of black-box brain experiments is introduced. The first experiment of this kind and the discovery of the action functional are described, and several other experiments that are either successfully completed or under way are discussed. Emphasis is placed on the variety of applications and the wide scope of the causal theory, because they attest to our conjecture that the causal theory is indeed a theory of intelligence and of the brain, and has as a consequence a scope and range of applications at least as great as our own cognition. Current limitations, and the steps necessary to overcome them, are near the focus of the discussion. A preliminary project plan is included, and an application for a patent is recommended.

## 9. Black-box Brain Experiments

For the purpose of a black-box brain experiment, the brain is considered as an implementation of a causal dynamical system that receives some input information, processes it, and generates some output information. Only the input/output relationship is of concern. The details of the implementation, no matter how complicated they may be, are irrelevant. The output from the box is known as the “measured” information.

The same input information is separately supplied to a causal virtual machine, such as the one proposed in Section 4.3 of Pissanetzky (2012f). The resulting output is considered to be the “prediction” and is expected to be identical to the measured output.

The black-box concept is expanded in the next Section. The first experiment of this kind was performed by one of us (SP) years ago. It was designed as a research tool at a point where extraordinary self-organization had already been observed in causal sets (by SP) but was not well understood, and it led to the discovery of the action functional given in Eq. (3).

After the discovery of the action functional, an experimental program was started to conduct more experiments of that kind, the main purpose being to understand the scope of the finding and write a corresponding theory. Several sources of data for the experiments were identified, and are discussed in Section 9.3. Several experiments were indeed performed, as discussed in Section 9.4. All experiments were designed subject to severe limitations in computational power and lack of resources. Other possible types of brain experiments that were not tried yet are in Section 9.5, and the project plan schematic of Section 10 proposes a possible approach to continue the research.

Contrary to what it may appear, plenty of detailed and well-documented information for experiments of this kind exists in the literature for the case where the black box is the entire brain. For experiments involving parts of the brain, it is obviously more difficult to obtain precise measurements of input/output, but it may

still be possible to carry them out in small animals or in the context of the projects mentioned above, followed with an effort to develop better tools for this research.

### 9.1 The Concept of Black-box Brain Experiments

Figure 1 illustrates the concept of a black-box experiment. The black box contains any physical device that can process causal information. It is called a black box because we do not need to know what's inside. The device receives a certain well-defined causal input, and generates a certain causal output, both of which can be observed and measured. The device can have an associative memory, but then the content of the memory must be considered as part of the input. This is the experimental part, and the output is the observed result.

Separately, the same causal input is supplied to a CML virtual machine. The machine is extensively discussed in Section 4 of Pissanetzky (2012f), but is not covered here. The machine knows only the input information, but knows nothing about the device in the black box, is not an emulation, simulation, or replication of the device, and contains no computer and no computer program. The machine is a causal processor that operates thermodynamically in accordance with the causal theory. The machine is equivalent to function  $\mathcal{E}$  discussed in Section 4.1.

However, it is easy to implement the virtual machine on a computer, in which case it becomes a simulation of the thermodynamics of CML, but only the thermodynamics, never the device. The computer implementation was also discussed in §4 of Pissanetzky (2012f). In this case, the machine does need a program, simply because any computer needs a program, otherwise it wouldn't work. The program performs only two tasks: the input and output of the supplied and resulting information, and the minimization of the functional in Eq. (3). The program never changes, and never has any knowledge at all of the device in the black box. It is fixed, it has to be developed only once and then used in perpetuity for all experiments.

The two outputs are, then, compared. If they are similar, then quantitative conclusions can be drawn regarding details of the internal implementation of the device in the black box. Several black-box experiments of this type were performed, and agreement was found in all cases. The most notorious quantitative prediction is the optimally short length of dendritic trees, later independently confirmed by a team of neuroscientists. Without CML, even with the power law now known, and even if a connection between the power law and intelligence were to be discovered, it would still take neuroscientists considerable effort and time to explain the connection.

One of the experiments below is for the vision system, others are for high function in the human brain. We don't have access to resources adequate to perform these experiments in selected parts of the brain. The issue of limitations is discussed some more in Section 10.

Table 1: Black-box experiments

Name	Section	input	output
Foundational experiment	9.2	code	refactoring and OO design
European Example	9.4	code	refactoring and OO design
Image recognition	9.4	retina	identified object
Theories of Physics	9.4	measurements	law of Physics
Euler Equation	9.4	measurements	law of Physics

Experimental results from several different classes of black-box experiments are discussed in Section 9.4. Table 1 lists the names of the experiments and their respective input and output. These experiments indicate that the notion of brain analysis is not just convenient, but necessary. The brain receives causal input from sensory organs and other afferent nerves in a form that is easy to formalize as a causal set. The observable output consists of objects that we can understand and use to make decisions and take actions. However, the input for the European Example (§9.4) consists of a computer program, the input for Image Recognition (§9.4) is a table representing points on a plane, the input for Theories of Physics (§9.4) is a set of equations, and the input for Euler Equation (§9.4) is also a set of equations. All these tasks are considered as high brain function.



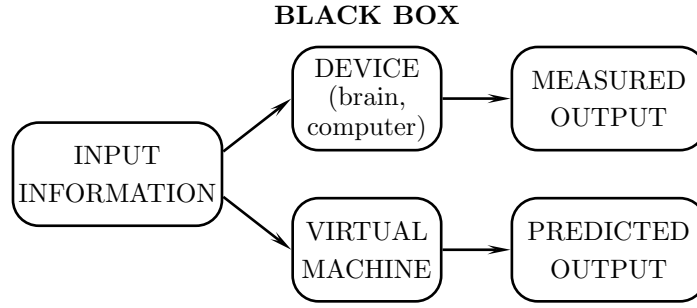


Figure 1: A black-box brain experiment. The “device” in the black box is any unknown dynamical system that receives some input information, processes it, and generates some output information. A causal virtual machine receives the same input information and generates its own output. Predicted objects calculated by the virtual machine are compared with the natural objects and found to confirm the theory.

If that kind of input is presented to a human analyst, the analyst’s visual system would first convert the program, the table, the equations, or the tasks into action potentials, and a great deal of processing would have happened in the analyst’s brain before the action potentials are again interpreted as a program, equations, a table, or tasks. This processing is addressed in the image recognition experiment, but excluded from the others. In all cases, the input is represented as a causal set.

## 9.2 The Foundational Black-box Brain Experiment

The original motivation for this work was Sergio Pissanetzky’s (SP) interest in the automation of the GUAPs, the Great Unsolved Automation Problems of Software Engineering. The GUAPS are loosely defined as a class of problems that are easy for humans but difficult for computers to solve (Pissanetzky, 2011b). *Refactoring* (Opdyke, 1992), or the organization of computer code so it becomes more “understandable” to the developer, is one of the GUAPs. There are many tools designed to help the developer to refactor code in different languages, but in the end, it is always the developer who makes the real decisions, the decisions that deal with the “understandable” part. A full solution of the problem of refactoring would entail a solution to the problem of meaning, but such solution was not available at the time.

The first experiment was carried out by SP around 2007. A simple C program with only 18 lines of code was chosen for the experiment. The developer, in this case SP, was asked to refactor the code, given in the form of “spaghetti code”, badly disorganized and nearly impossible to understand. Any experienced developer can refactor a program that short in a few minutes, and do so automatically, without any conscious input or intelligent directives. But no experienced developer had been able to code the decisions that her mind makes during the process. The program is reproduced in mathematical notation in Eq. (20) of Pissanetzky (2012f).

At that time, SP had already noticed the experimental fact that certain sparse “canonical matrices” that represent code had remarkable self-organizing properties. Canonical matrices are equivalent to causal sets, so the finding was the same as saying that causal sets had remarkable self-organizing properties. This statement is itself remarkable. To investigate the matter, SP created a setup where a canonical matrix corresponding to the initial disorganized version of the program was constructed, and any refactoring decisions made by SP’s mind were reflected as operations that caused changes on the canonical matrix. This setup made it possible for SP to directly observe the mathematical effects of his refactoring decisions on the matrix.

Contrary to what it may appear, the experiment was not easy. The conclusion, is that the SP’s brain was trying to minimize the total length of the flux lines in the matrix. The process consisted of a series of permutations among neighboring statements in the program, all of which were conducive to the minimization

of that length. This conclusion represents the discovery of the action functional, now proposed in Eq. (3), and led to the definition of CML as a thermodynamical process that minimizes the action functional. These are the conclusions as they stand today, but they were not easy to arrive at the time. It took several years of intense research to discover the significance and wide scope of what had been found, and to elaborate the corresponding theory as it stands today.

More details about this experiment were discussed in Section VIII.A of Pissanetzky (2010) and reviewed in Section 4.3 of Pissanetzky (2012f).

### 9.3 Sources of Data for Black-box Experiments

The intent of the first experiment was to discover the secret of the intelligent behavior that a developer exhibits when refactoring code. The experiment had not been planned to be observer-independent or to have rigorously controlled parameters. After that goal had been achieved, the natural next step was a search in the literature for other possible sources of causal input and corresponding output obtained by a human that had been independently obtained and documented in the literature. Luckily, there are several sources and plenty of carefully documented cases.

One of such sources, is object-oriented (OO) code. In the course of development of OO code, a human analyst receives a *problem statement* that describes some physical system, such as a business, or a machine, or perhaps a problem of Physics that needs to be solved, and is asked to create an object-oriented model of that system. The model is subsequently used to develop software and documentation. All the work is causal, and the process amounts to a carefully controlled experiment where the given problem is the causal input, the analyst's brain is the device in the black-box that provides the transformation dynamics, and the resulting OO code and documentation are the output.

The second source of carefully documented experiments is provided by the theories of natural science, that is, by the theories themselves, not the measurements on which the theories are based. A theory of science also consists of objects. In this case, the input consists of the experimental measurements that support the theory, the scientist's brain is the device, and the output is the theory.

The third source of interesting brain experiments, is image recognition. In the Hofstadter's example, the causal input consists of the 100 million dots plus a collection of causal imagery in memory, the output is 'mother', and the device is the retina plus the brain of the observer. More in general, a television picture that consists of illuminated pixels can be the input. The eye and brain of the person who is watching the image provide the dynamics, and the output is the recognition of the objects present in the scene.

Experiments in all these 3 areas were initiated. Some were completed, but some run into computational limitations and were carried only partially. The limitations arise from the need for special hardware, as discussed in Section 4.3 of Pissanetzky (2012f). The hardware is very simple, but was not available at the time and is still not available now. All the work that was completed, was also successful, no disagreements were observed. The following Section reviews some of this effort.

### 9.4 Observer-independent Black-box Experiments

Following the success of the foundational black-box experiment and the discovery of the action functional, an effort was initiated to provide additional experimental evidence that is more advanced and observer-independent. Software was developed for that purpose.

#### THE EUROPEAN EXAMPLE

A problem on object-oriented software analysis, design and refactoring was selected. The problem statement describes a Local Area Network (LAN) with 4 stations of different types interconnected in a token ring. The statement describes the actions that each type of station is expected to take, such as allow a user to enter a message, receive a message, retransmit it to the next station, print it, destroy it, or report the condition of the network. The statement is, of course, causal, and is converted to the form of a causal set, consisting of ordered pairs given in any arbitrary order. An experienced developer is in the black-box. The developer is

asked to produce a design for object-oriented Java code. The problem is used in European universities to teach refactoring and OO analysis and design. The final Java code is public (Demeyer, Rysseberghe, and et. al., 2005).

The resulting OO design is compared with the output of the causal virtual machine, as schematically represented in Fig. 1. As usual, the machine is fixed. The machine is trained with the given causal set as input, and its only task is to minimize the action functional and output the resulting OO design. The machine knows nothing about Java, or about the theories of object-orientation or refactoring. The purpose of the experiment is to demonstrate that the virtual machine accurately emulates the high brain function of the developer in the black-box.

It would, of course, be of great interest to obtain the Java code itself as the final product. This is simple engineering. It would involve training the virtual machine with the content of a Java user's manual. We anticipate that such machines will soon be available, already pre-trained on various languages, and will be of enormous commercial interest. Similar machines will also be developed for applications in high-performance computing, such as the one reported in [Hon, Liwen]. However, for the time being, training a causal virtual machine with a user's manual is beyond available resources. For the purpose of the experiment, the design, not the code itself, was the output to be considered.

The experiment, results, and interpretation are discussed in Section VIII-C of Pissanetzky (2010). Additional details are available in the supplementary material, and consist of several files with names starting with "The European Example", accessible from [www.SciControls.com](http://www.SciControls.com). Available are the original Java code, the execution trace considered for the experiment, the original causal set, the set of least-action permutations, and the interpretation and fractal hierarchy. The causal set of size 33 was obtained by application of function  $\mathcal{E}^{-1}$  directly to the Java code. The causal machine partitioned the set into 19 blocks, all of them belonging to just 4 classes. The agreement between the resulting classes, instantiated objects, and inheritance structure, and the hierarchy of block systems generated by the causal machine, was excellent.

This experiment strongly confirms several important conclusions. (1) That CML, as implemented on the causal virtual machine, can indeed duplicate high brain function in humans. (2) That CML indeed creates semantics and meaning in the same way we humans do. The whole purpose of techniques of object-orientation and refactoring is to improve the *understandability* of machine code for human developers. This task can only be achieved by creating meaning in the same way humans do. (3) That CML is the key for both AGI and brain research. It is not about computational efficiency. The point is that, as confirmed by prediction and confirmed by Cuntz, Mathy, and Häusser (June 2012), the human brain actually runs CML in the unconscious, and the structures and meaning arrived at by CML are the same as the structures and meaning arrived at by humans.

## IMAGE RECOGNITION

One of the most exceptional feats of the human intellect is its ability to recognize the objects in a scene in very little time and with so little effort, and to do so continuously, without interruption during the entire day. When light from the scene impinges upon a light sensitive element in a camera or retina, the element generates an electric signal that acknowledges the fact that causal information from the external world has just been acquired, where the light is the cause and the signal is the effect. Hence, (light, signal) is an ordered cause-effect pair. All sensors operate this way. When many sensing elements are operating at the same time, the resulting output is a causal set. There is no need for any conversions.

In the brain, the causal set captured at the retina is transmitted to the brain and interpreted by causal logic, resulting in the identification of the objects present in the scene. This process is extremely simple, basic, and fast, yet such a large amount of effort has been invested for so many years in designing image recognition signals with only partial results.

The experiment was presented in Section 5 of Pissanetzky (2008) and further discussed in Section VIII-B of Pissanetzky (2010) and in Section 4.5 of Pissanetzky (2011a). Unfortunately, the experiment is still ongoing due to lack of resources. The causal set has 1433 elements, too many for our capabilities. Still, one major feature of the image has been correctly detected: the edges of three objects composed of points,

present in the scene, indicating that edge detection works without having been explicitly programmed. It is not currently possible to proceed to further details. Recall that the causal processor is generic, and has not been programmed in any way, certainly not with any edge-detection algorithm, and yet it detects edges. The results are available as supplementary material (Pissanetzky, 2012g).

## THEORIES OF PHYSICS

An even more extraordinary feat of human intelligence is its ability to abstract sentient evidence and configure it as a formal theory that allows complex predictions to be made. More precisely, the unparalleled ability of humans to construct theories of nature. However, as our working hypothesis describes the brain as a thermodynamical implementation of causal mathematical logic, with the ability to formalize its own abstractions and create meaning the same way intelligent humans do – even scientists – it is therefore appropriate and necessary to test whether such an ability does indeed pertain to CML as well.

Limitations discussed elsewhere and related with the need for simple but special hardware have prevented us from undertaking any large scale attempt at answering the question. In a very small scale, and only as a proof of concept, one of us (SP) has solved the Euler equation as a black-box brain experiment with a human scientist in the black box, and a causal virtual machine trained with a causal set that describes the causality involved in a simple measurement, but knows nothing about Physics or Mathematics, or even differential equations. The computational experiment was carried out in February 2012. The results are available as supplementary material (Pissanetzky, 2012d).

## EULER EQUATIONS

This experiment is interesting because it illustrates the first use of the mathematical limit *within* a causal set, hence extending CML to applications with differential equations. We humans can deal with the infinite and the infinitesimal using 100% finite means. We can define the concept of mathematical limit and use it to “invent” real numbers, continuous functions and manifolds, integrals and derivatives, and all the powerful properties of continuous things without ever having to write, calculate, display, or draw those infinite things. We always prove the properties of those entities by way of the mathematical limit. We approximate real numbers as integers, we use integers in computers for real calculations, we solve differential equations in time steps, or with finite elements. Our means are finite, yet we deal with the infinite.

Causal logic can too. Discrete, finite causal logic can deal with continuous entities the same way we humans do. If it didn’t, then our conjecture that causal logic is a theory of intelligence would be invalid.

Euler equations are a set of three differential equations that describe the rotational motion of a rigid body. Euler used differential calculus to derive them from Newton’s equations, who in turn derived his from experiment. For the purpose of this black-box brain experiment, we consider an imaginary scientist who lived before Newton and performed some experimental measurements on rotating rigid bodies, but did not understand his measurements, didn’t know Newton’s equations, and was not able to derive the equations of motion. The scientist expressed his measurements as simple formulas involving real numbers that he could measure, but whose meaning he couldn’t understand. He took measurements at short time intervals, and knew that shorter intervals produced more accurate results. That’s all the scientist knew. And that’s all the causal machine was told, yet it correctly derived Euler’s equations.

For the purpose of the black-box experiment, we extracted the causality from that scientist’s expressions and reduced them to a causal set with 21 elements and 18 ordered pairs. As usual, the elements have no meaning. After causal processing, three difference equations are obtained, where the time step appears as a factor. The three differential Euler equations are obtained upon taking the limit for infinitesimally short time steps. The experiment is described in detail in the supplementary material (<http://www.scicontrols.com/Publications/Euler.pdf>).

## 9.5 Other Types of Black-box Experiments

Just as the brain is considered as a black box with an input and an output, any functional part of the brain can as well be considered as a black box with an input and output. This idea gives rise to the notion of *brain analysis*, discussed in Section 9.1. Algorithms and a computer implementation are discussed in Section 10.

Black-box brain experiments can also be carried in reverse, that is, start from the given output and calculate the input necessary to produce that output, which is equivalent to application of function  $\mathcal{E}^{-1}$ . This approach provides additional information, helps with causal studies, and it may be the usual case when not all input causality is known.

A different type of brain experiments involves *in-vitro* cultures of live neurons ground from animal brains and placed on a Petri dish with multi-electrode arrays, also known as brain-on-a-dish experiments, where the behavior of individual neurons can be directly observed. The experiments are designed to observe and measure spontaneous network formation and neuronal behavior. The electrodes detect the action potentials of individual neurons as they spontaneously organize in the culture, create a network, and become active. This experiments were proposed for CML research (Pissanetzky, 2010, § VIII.D). See also (Chen et al., 2010) for cultured hippocampal networks.

Nothing has been said yet about sensory-motor integration, an all-important topic critical for gaining understanding of the operation of the brain. *Integration* is the key here. Integration is precisely what CML does.

## 10. Project Plan

The greatest strength of a scientific theory is that it can be falsified, but has never been. A theory is scientific only when it satisfies Lee Smolin's three conditions: it must be confirmable, it must be falsifiable, and it must be parsimonious. The third condition is also known as Occam's razor for competing hypothesis. Causal Mathematical Logic (CML) is clearly confirmable. All the experiments of the previous section confirm the theory. The theory is parsimonious. It has only one fundamental principle, causality; only one hypothesis, the action functional; only one physical system, information; only one space, causal space; and only one law, the logic. And the theory can be falsified. Any of the experiments could have falsified it. But the theory is weak. It is weak because of practical limitations that have permitted only small-scale experiments, and have limited the range and the scope of the experiments. This project plan offers the steps that need to be immediately taken in order to make the theory stronger by trying to falsify it. There are only two outcomes possible: either the theory is falsified and discarded, or a stronger theory will emerge with all the technical means to undertake large-scale developments in Neuroscience.

The greatest strength of causal logic is its power to integrate fragments of knowledge or fact into a functional whole. An attempt to harness that power immediately spawns hundreds of projects in many different disciplines including AGI and Neuroscience.

A schematic preliminary project plan has been prepared in the supplementary material (Pissanetzky, 2012e). The project only covers immediate actions needed to jump-start a project on CML and expand it to a point where CML is sufficiently developed for application in brain research and AGI. The project includes requirements for hardware and software. The software can be developed separately if preferred, but it will be difficult to test or make it operational without the hardware. Hence, the hardware needs to be developed first.

The required hardware is a *causal processor*. It consists on an implementation of the virtual machine proposed in Section 4.3 of Pissanetzky (2012f) on a programmable processor such as an FPGA or GPGPU unit. There is still no design for it, but a crude estimate indicates that it may contain more than 100,000 neurons per unit. Units are inexpensive and stackable with no penalty paid for time of execution, except that a slight latency associated with inter-unit communications is to be expected. This way, a stack with millions of neurons can be obtained at low cost. It will run an estimated 1 million times faster than the human brain. In time, much faster LSI processors with many more neurons per unit will be produced, and a stack of them will contain as many neurons as the human brain and will execute many millions of times faster.

The causal processor's only function is to calculate values of function  $\mathcal{E}$ , as defined in Eq. (5), for a given causal set. The given causal set provides *training* for the processors, or one can also say that the processor *learns* or gains experience from the environment. The causal set is the only contact between the processor and the environment. The processor is fixed. Once designed and tested, it will work for all types of problems and it never needs to be modified or re-programmed again.

Once the hardware is functional, there will be many projects that can start simultaneously and run in parallel by different teams, depending on priorities. Some ideas follow:

- Repeat the foundational experiment (§9.2) and the other initial experiments performed by SP, but on a larger scale. This will serve to test the processor, verify the experiments, and develop techniques and gain experience in causal analysis.
- Solve the senses, particularly vision, hearing, and olfaction. Build systems for image and voice recognition based on causal theory, not on engineering, that will operate millions of times faster than the brain. Build smell sensors better than dogs for drugs, explosives.
- Brain-on-a-dish experiments. This type of in-vitro experiments with live neurons ground from an animal's brain emphasize that the neuron's behavior is local, determined by the neuron itself and not by any global requirements such as intelligence, or by the network they participate in, or even by the surrounding morphology. Neurons and dendrites on the dish can be observed and measured, and a CML model of the experiment can be built and compared with measurements.
- Candidates for black-box experiments are cases where either input or output are known, and at least something is known about their counterpart. In the Hofstadter's example, the input is given, and 'mother' is known about the output. Examples where the output is known or measurable include gait analysis and sensory-motor integration, not just in humans or animals but also in robots.

Most importantly, once the hardware is ready and basic experience in causal analysis is gained, then CML will be ready as a new tool for brain research.

### 10.1 Converting Existing Simulations into Causal Models

It is easy, in relative terms, to convert legacy code or existing brain simulations in computer software to causal models. A conversion module has to be developed that can convert both ways between each programming language of interest and causal models, or alternatively between the corresponding virtual machines and the causal virtual machine. This task, writing the conversion module, has to be accomplished only once for each language. It is anticipated that, with time, readily available conversion modules will exist for many languages. As a by-product, the modules will also allow direct translation between the programming languages themselves.

Generic, language-independent transformations suitable for all languages have been listed in Table I and discussed in Section III of Pissanetzky (2009). Admittedly, these transformations are cumbersome, but the task has to be done only once in a life-time, and the result is permanent.

An issue that will arise in practice, is the use of heuristic assumptions made by developers and firmly planted in code. As models from different origins are integrated into a general causal model, incompatibility issues between various heuristics will arise, and the heuristics will have to be eliminated or made compatible. The only information that needs to be entered to the causal model is data, measured, observed data, not interpretations. However, as humans have created these heuristics, humans can recognize them and eliminate them, and so also can CML. With time, as more experience is gained, it is anticipated that a fully automatic CML-based model conversion capability will be within reach.

### 10.2 Training Programmers in Causal Modeling

Programmers will still be needed to enter observational information into the causal models during their development. They will operate the GUI, test and maintain the system, upgrade it when necessary or if better hardware becomes available, and provide feedback to neuroscientists. But programmers have preferences.

Programming languages constantly change. Research demands come in a great variety but must remain flexible. This constant pressure for change is hard for programmers, and good programmers are hard to find.

Such misfortunes are alleviated by the proposed GUI and conversion modules. Programmers will not have to learn “causal modeling” as if it were yet another programming language. They will be allowed to work in their own preferred languages, and in a stable environment. The object models for their programs will be done automatically by CML via the conversion modules, specific to their preferred language. Multi-language teams of developers will become common in following years. With causal modeling, programmers will be relieved of much of their non-essential work. They will be able to better concentrate in their specific task, which is to supply new information to computers. They will become more efficient and productive as their work simplifies. There will have to be a transition period, however, these changes cannot happen immediately.

With time, after a transition period and as CML becomes more powerful and causal research advances, natural languages, where words and sentences and books are also invariant observables of the same kind we discussed in Section 3, will become part of the GUI. It will then be possible for neuroscientists to communicate directly with the models. This is part of our vision, but we are not there yet. We have to get there first.

### 10.3 Creating a GUI for Causal Modeling

A graphical user interface (GUI) for work in causal modeling is proposed. The features will include various forms of input of causal information, such as files and models in other programming languages, visualization methods, and testing.

Such a GUI will give neuroscientists unprecedented flexibility for programming language selection. New languages specific to brain simulation will be created and easily interfaced into central causal models without the usual design cost and overhead associated with man/machine communication. It will allow programmers to work in their languages of choice, with the only limitation that an automatic converter for that language to and from causal models must exist.

Testing will also be included in the GUI. Testing is very special in causal logic: in short, it works just like a human. What does a child do when information is incomplete? He asks for more. So also does CML. What does a child do when information is contradictory? He asks why. So also does CML. With time, testing will be more of a conversation than the traditional case-by-case testing used in Engineering. These features have been discussed, with the help of examples, in Section IV “Basics of MMC supervised learning” of Pissanetzky (2009).

## 11. Conclusions and Outlook

Discussed in this paper is a new, fundamental theory of Physics grounded on the universal principle of causality as its only principle. The theory has only one postulate, the action functional, and one model, a collection of ordered (cause, effect) pairs known as *information*. It proposes that natural phenomena ranging from the elementary to the complex, and including life and thought, are expressions of the mathematical properties of ordered pairs and must be understood as such.

The result is a scale-free, detail-tolerant theory that describes observables as fractal hierarchies of algorithms, or behaviors, allows for their quantitative calculation, and seamlessly extends across scales and substrates from the fundamental Mathematics of causal pairs to elementary systems to complex systems. This is a theory where lossless physical models of variable granularity and complexity can be constructed and integrated together, from the finest details, to high-level models that emphasize global features and observables. The theory is a unifying principle for information.

In Neuroscience, the theory is proposed as a universal, mathematical, non-explicitly programmed, computer-readable and human-interfaceable language, with a graphical user interface that will allow neuroscientists to express their observations and measurements and integrate them in all scales, where

elements of information are automatically associated and compressed to create associative memories, and high brain function including signals of intelligence emerges from models.

Black-box experiments are central to addressing the issue of the external observation of life signals. They bridge the gap between internal function and externally observable signals. In simulation, the requirement for life signals to emerge rather than being explicitly programmed is the same requirement of self-programming.

Because simulation is approaching pure science, we expect it to help us understand the entire system more accurately. Whole-brain emulation (WBE) places more intrinsic demands and seeks a system strip down. For example developmental space is not a priority. The WBE system is a copy of our information and we require to restart the processes of intelligence with the least running description. WBE will likely require a subset of the principles which will be derived from tissue simulators, and tissue simulators will benefit from predictions made from the principles of intelligence. It appears then that from a logical sequence point of view the progress for system simplification could be hindered without a natural theory of information and intelligence. The issues here for such a fundamental theory are how it should be derived: (1) from a pure science/replication perspective such as SimREP; or (2) recruiting a basic theory like CML to assist with consistency checking simulation output.

When we start up a WBE, the information for the neural components will need to have been compressed into the set of simulation objects and integrated frameworks deemed suitable for them to be run. As the machinery is the information, a temporary dualism between process and information will occur during the phase we derive the design of these objects and frameworks from general principles. Hopefully we can resolve this and get our theory of natural information right to predict the foundational methods. If such a theory is solid, the WBE system should require less peripheral support to compensate the wrong foundation, and so transferring biology to a new substrate will have more freedom in the physical world to direct itself with a natural evolution. If we get it wrong, decades spent on development may be resting on a house of cards.

The ideal sequence at its simplest form is : (1) agree on a starting basis for the natural principles of intelligence such as CML; (2) run experiments to determine their consistency with SimREP; and (3) use derived principles for Simulation objects and frameworks for WBE.

In reality it will be far more complex than this. There will be two steps forward one step back, a mixup of various directions as well as leftfield insights, which feed off and back into each other. We cannot depend on having the ideal sequence, but we propose that the sequence is still important enough there should be ongoing discussion on whether or not a fundamental theory of physical intelligence should be integrated a priority to the WBE roadmap.

Practical issues such as developing special hardware, building a prototype, training scientists and technicians in the new discipline of causal analysis, building a development platform, interfacing with legacy software, and training causal machines have been considered. A complete program for the development of new causal systems has been proposed, including hardware, software and human training. Also proposed is the design of massively parallel specialized hardware of a high efficiency, that works much faster than the human brain and is of a general nature, yet inexpensive, simple, small, light, portable, and stackable as the size of the models increase.

A concerted program based on black-box experiments is proposed to make more life signals externally observable. The program should be directed, of course, to interests in Neuroscience, although applications in robotics, in Software Engineering, such as solving the GUAPs, and some general-purpose applications can also be considered and may attract additional funding.

## Acknowledgements.

To Dr. Liwen Shih, Dr. Peter Thieberger, Juan Carlos Kuri Pinto, and many others, for their support and many contributions. To Carey Carlson. A few sentences used in the paper were inspired by his writings or are literary quoted. To Carla Lanzalaco for her contribution to Genetics.



## References

- Berut, A.; Arakelyan, A.; and et al., A. P. 2012. Experimental Verification of Landauer's Principle linking Information and Thermodynamics. *Nature* 483:187–189.  
Abstract: <http://www.nature.com/nature/journal/v483/n7388/full/nature10872.html>.
- Chen, W.; Li, X.; Pu, J.; and Luo, Q. 2010. Spatial-temporal dynamics of chaotic behavior in cultured hippocampal networks. *Physical Review E* 81(061903).
- Cuntz, H.; Mathy, A.; and Häusser, M. June 2012. A scaling law derived from optimal dendritic wiring. *PNAS*, 2012, DOI: 10.1073/pnas.1200430109 1–5. Available electronically from <http://www.pnas.org/content/109/27/11014>.
- Deca, D. 2011. Available tools for whole brain emulation. *Int. J. of Machine Consciousness* 04:67.
- Demeyer, S.; Rysselberghe, F. V.; and et. al. 2005. The LAN-simulation: a refactoring teaching example. *8th Int. Workshop on Principles of Software Evolution, Lisbon* 1:123–134. Code and teaching materials available from [www.lore.ua.ac.be/Research/Artefacts/RefactoringLabSession](http://www.lore.ua.ac.be/Research/Artefacts/RefactoringLabSession).
- Eagleman, D. 2011. *Incognito*. New York: Pantheon Books.
- Eigen, M. 2013. *From Strange Simplicity to Complex Familiarity*. New York: Oxford University Press.
- Fuster, J. M. 2005. *Cortex and Mind*. New York: Oxford University Press.
- Gardner, A., and Conlon, J. P. May 2013. Cosmological natural selection and the purpose of the universe. *Complexity* doi: 10.1002/cplx.21446.
- Gell-Mann, M. 1994. *The Quark and the Jaguar*. New York: W. H. Freeman and Company.
- Gjuvslund, A.; Vik, J.; Beard, D.; and et al. 2013. Bridging the genotype-phenotype gap: what does it take? *J. Physiology* 591(8):2055–2066.
- Hales, C. G.; Grayden, D. B.; and Quiney, H. 2011. The electric field system of a macular ion channel plaque. In *33rd Annual Conf. of the IEEE EMBS, Boston, MA*.
- Hales, C. G. 2011. On the status of computationalism as a law of nature. *Int. J. of Machine Consciousness* 3(1):55–89.
- Hawkins, J. 2004. *On Intelligence*. New york: Times Books.
- Karr, J.; Sanghvi, J.; Macklin, D.; and et al. 2012. A whole cell computational model predicts phenotype from genotype. *Cell* 150:389–401.
- Kauffman, S. 2011. Answering Descartes: Beyond Turing. In *European Conference on Artificial Life (ECAL)*, 11–22. Available from: <http://mitpress.mit.edu/sites/default/files/titles/alife/0262297140chap4.pdf>.
- Koene, R.; Hutter, M.; and Hales, C. October 2012. Intelligence, substrates, and computation. Possibilities for the future. Available electronically at <https://www.youtube.com/watch?v=KgcBvFpCD2k>.
- Landauer, R. 1996. The physical nature of information. *Physics Letters A* 217:188–193. Available electronically from <http://www.sciencedirect.com/science/article/pii/0375960196004537>.
- Landauer, R. 1999. Information is a physical entity. *Physica A: Statistical Mechanics and its applications* 263:63–67. Available electronically from <http://www.sciencedirect.com/science/article/pii/S0378437198005135>.
- Lanzalaco, F., and Pissanetzky, S. 2013. Causal Mathematical Language as a guiding framework for the prediction of "Intelligence signals" in brain simulations. *J. of Artifical General Intelligence* December 2013.
- Lin, L.; Osan, R.; and Tsien, J. Z. 2006. Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes. *Trends in Neuroscience* 29(1):48–57. Available electronically from <http://www.ncbi.nlm.nih.gov/pubmed/16325278>.
- Noble, D. 2012. A theory of biological relativity: no privileged level of causation. *Interface Focus* 2(1):55–64.

- Opdyke, W. F. 1992. *Refactoring Object-Oriented Frameworks*. Ph.D. Dissertation, Dep. of Computer Science, Univ. of Illinois, Urbana-Champaign, Illinois, USA. Available electronically from <http://www-public.it-sudparis.eu/~gibson/Teaching/CSC7302/ReadingMaterial/Opdyke92.pdf>.
- Papo, D. 2013. Why should cognitive neuroscientists study the brain's resting state? Opinion article doi: 10.3389/fnhum.2013.00045.
- Pissanetzky, S. 2008. A new type of Structured Artificial Neural Networks based on the Matrix Model of Computation. In *Proc. Int. Conf. in Artificial Intelligence (ICAI 2008)*, 251–257. Abstract: <http://www.scicontrols.com/Publications/AbstractSP2008C.pdf>.
- Pissanetzky, S. 2009. A new Universal Model of Computation and its Contribution to Learning, Intelligence, Parallelism, Ontologies, Refactoring, and the Sharing of Resources. *Int. J. of Information and Mathematical Sciences* 5:143–173. Available electronically from <https://www.waset.org/journals/ijims/v5/v5-2-17.pdf>.
- Pissanetzky, S. 2010. Coupled Dynamics in Host-Guest Complex Systems Duplicates Emergent Behavior in the Brain. *World Academy of Science, Engineering, and Technology* 68:1–9. Available electronically from <https://www.waset.org/journals/waset/v44/v44-1.pdf>.
- Pissanetzky, S. 2011a. Emergence and Self-organization in Partially Ordered Sets. *Complexity* 17(2):19–38.
- Pissanetzky, S. 2011b. Emergent inference and the future of NASA. Workshop, NASA, NASA Gilruth Center, Johnson Space Center, Clear Lake, TX. Available electronically at <http://www.scicontrols.com/Publications/AbstractNASA2011.pdf>.
- Pissanetzky, S. 2012a. A case study: the European Example. Available electronically at <http://www.scicontrols.com/Articles/EuropeanExample.htm>.
- Pissanetzky, S. 2012b. Causality, symmetry, brain, evolution, DNA, and a new theory of Physics. *Bulletin of the American Physical Society* 57(10):E1.0003. Available electronically from <http://meetings.aps.org/link/BAPS.2012.TSF.E1.3>.
- Pissanetzky, S. 2012c. The Detailed Dynamics of Dynamical Systems. Available electronically at <http://www.scicontrols.com/Articles/TheoryOfDetailedDynamics.htm>.
- Pissanetzky, S. 2012d. Differential Euler Equations obtained from Causal Mathematical Logic. Available electronically at: <http://www.scicontrols.com/Publications/Euler.pdf>.
- Pissanetzky, S. 2012e. Project Plan for a core implementation of Causal Mathematical Logic. Available electronically at: <http://www.scicontrols.com/ProjectPlan/CoreImplementation.htm>.
- Pissanetzky, S. 2012f. Reasoning with Computer Code: a new Mathematical Logic. *J. of Artificial General Intelligence, special issue on Self-programming* 3:11–42. Available electronically from <http://www.degruyter.com/view/j/jagi.2012.3.issue-3/issue-files/jagi.2012.3.issue-3.xml>.
- Pissanetzky, S. 2012g. Separating points. Available electronically at <http://www.scicontrols.com/Articles/PointSeparation.htm>.
- Pissanetzky, S. 2012h. The theory of Detailed Dynamics and the Combinatorial Explosion. Available electronically at <http://www.scicontrols.com/Articles/TheTheoryAndTheCombinatorialExplosion.htm>.
- Pissanetzky, S. 2013. The unification of symmetry and conservation. *Bulletin of the American Physical Society* 58(3):N2.0001. Available electronically from <http://meetings.aps.org/link/BAPS.2013.TSS.N2.1>.
- Wallace, D. 2013. Thermodynamics as control theory. Available electronically at <http://philsci-archive.pitt.edu/9904/>.
- Wissner-Gross, A. D., and Freer, C. E. April 2013. Causal Entropic Forces. *Physical Review Letters* 168702:1–5. Available from: <http://www.alexwg.org/publications/PhysRevLett.110-168702.pdf>.