From Simulating to Duplicating the Brain

Abstract

The philosophy of AI has a long-standing tradition of discussing brain duplicates and brain simulations as well as a tendency to blur the lines between the two. The distinction between simulating and duplicating the brain has become increasingly important with the emergence of ”neuromorphic computers”, hardware operating with bio-inspired mechanisms and interconnecting artificial neurons and synapses. This paper explores what it means to duplicate the brain rather than merely simulate it. I claim that while simulations share only a mathematical structure with their targets, duplicates have the same relevant measurable properties and are governed by the same relevant causal processes as the target system. I propose six criteria for brain duplication that are often not met by simulations, thereby offering a clearer understanding of the complexities involved in achieving brain duplication in artificial systems. Furthermore, this paper explores whether neuromorphic computers can duplicate neural *computations* by exploring the notion of models of computation. I submit that analog neuromorphic computers that use memristive technology are candidates for duplicating neural structures with respect to simple models of neural computation. Finally, I discuss five possible objections to my view.

1 Introduction: The Case of Neuromorphic Hardware

There is a long tradition of philosophical arguments revolving around the questions of *simulating* versus *duplicating* the brain. Critics of computational functionalism argue that not all cognitive capacities can be accounted for computationally, and that for these capacities to be present in an artificial system, it needs to *duplicate* rather than merely *simulate* the causal powers of the brain—paradigmatic examples of such capacities being consciousness and intentionality. For example, Chalmers (1995, 1996, Ch. 7) famously claims that replacing neurons one by one with digital computer chips to create a whole-brain *simulation* would leave the subject’s conscious experience unaltered. Chirimuuta (2024) objects, asserting that ”the perfect input-output equivalent of the neuron would have to be something more like a *duplicate* [...]” (Chirimuuta, 2024, p. 265, emphasis added), and that for that “we would still need to go well outside the realm of [...] Chalmers’s electronics substitution scenarios” (Chirimuuta, 2024, p. 265). While the concept of brain duplication has so far been largely speculative, advancements in technology prompt reevaluation of certain concepts. Intuitively, duplication of the brain implies building a physical brain. Presumably, this amounts to developing artificial neurons and synapses and connecting them in a brain-like manner. This intuitive understanding was sufficient when brain duplication was far in the future. However, with Moore’s law reaching its limits and standard digital hardware facing constraints in terms of speed and energy efficiency, computer engineers are redirecting their focus to systems beyond the von Neumann architecture. Chips that operate with mechanisms similar to those in the brain have garnered significant interest. Known as *neuromorphic hardware*[1], these chips consist of interconnected physical ”neurons” and ”synapses” and use neural computation strategies. Let me provide you with a list of examples that should make clear that the intuitive notion of ”building physical brains” is no longer sufficient to decide whether neuromorphic chips can be used to duplicate the brain, and that a more thorough analysis of duplication is necessary.

In 1991, Mahowald and Douglas presented an analog integrated circuit with some of the functional characteristics of a biological neuron. This system uses physical dynamics to respond to a given input current with the same spiking pattern as a neocortical neuron (see figure 1). Since its discovery in 2007, significant attention has been directed towards the memristor.[2] A memristor is a two-terminal electrical device (similar to a resistor), whose conductivity changes as a function of the current flowing through it. Memristors are prime candidates for artificial synapses because their electronic features allow the hardware implementation of long-term synaptic plasticity (Serrano-Gotarredona et al., 2013). Large scale neuromorphic systems under construction today such as SpiNNaker2 (Mayr et al., 2019) comprise up to 10 million hardware ”neurons”, which is about the same number as the neurons of a zebra fish’s brain (Hinsch and Zupanc, 2007). Furthermore, research has progressed beyond circuit-based electronics. Inspired by ion-mediated voltage transfer in the brain, neuromorphic chips that use chemical ion concentration gradients for processing have been proposed (Agarwal et al., 2017). As a first step towards neuromorphic prostheses and brain-machine interfaces, Keene et al. (2020) directly connected an organic neuromorphic device with dopaminergic cells to create a biohybrid synapse that exhibited neurotransmitter-mediated synaptic plasticity (see figure 2). Their research aims at building hardware that can interface with living tissues and adapt to biofeedback. Furthermore, neuromorphic research is not restricted to the neuronal level. Meng et al. (2024) propose an ionic neuromorphic device that mimics the dendritic signal processing in the human brain. This list should not be regarded as remotely representative of the diversity of the current neuromorphic research. Instead, it aims to illustrate how significantly the state-of-the-art has evolved, thereby highlighting that an intuitive understanding of duplication is no longer sufficient in the context of neuromorphic computing.

Although the distinction between simulations and duplicates is common in philosophical discussions, it has rarely been articulated in detail. While a substantial body of literature exists on simulation (cf. Hartmann, 1996; Guala, 2005; Parker, 2009), there are no well-developed accounts

[1] In 1990, Carver Mead first introduced the term ”neuromorphic” to characterize computational systems composed of highly interconnected electronic circuits that emulate the neurobiological architecture of the nervous system. Since then, the field has evolved, encompassing a diverse array of approaches unified by a common goal: To develop brain-inspired hardware that capitalizes on the remarkable energy efficiency and real-time processing capabilities of the brain. Systems range from standard digital hardware optimized for parallel processing (cf. Merolla et al., 2014; Davies et al., 2018) to analog or mixed signal chips (cf. Pehle et al., 2022; Neckar et al., 2018) to bio-plausible neurons and neural implants (cf. Mahowald and Douglas, 1991; Abu-Hassan et al., 2019). For a comprehensive overview of this field, see Indiveri (2021).

[2] The term ’memristor’ is a concatenation of ’memory’ and ’resistor’. It is worth noting the distinction between the memristor as a hypothetical fundamental circuit element—originally postulated by Chua (1971) as the missing fourth passive circuit element with idealized properties—and memristors as realized in contemporary devices. The latter only approximate some of the characteristics predicted for the ideal memristor and typically exhibit behaviors and limitations not present in the original theoretical conception.

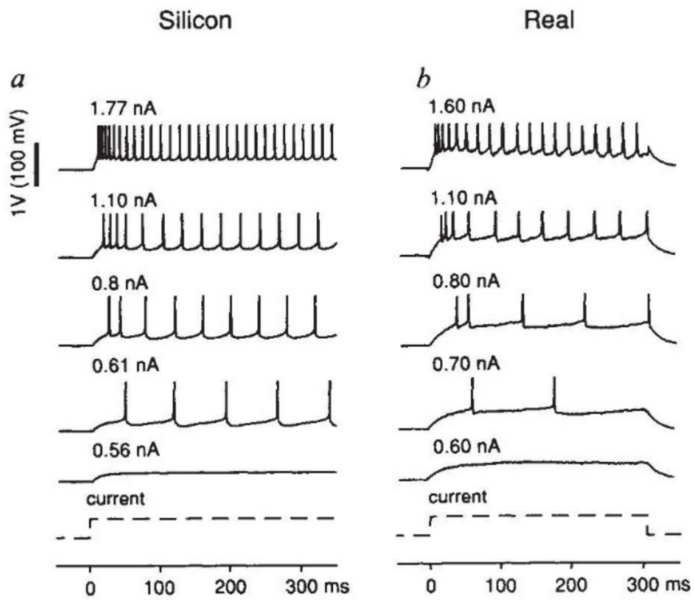

Figure 1: Comparison of response to intrasomatic current injection of the silicon neuron and a neocortical neuron. Voltage scale represents 1 V for the silicon neuron, and 100 mV for the neocortical neuron. Modified from Mahowald and Douglas (1991).

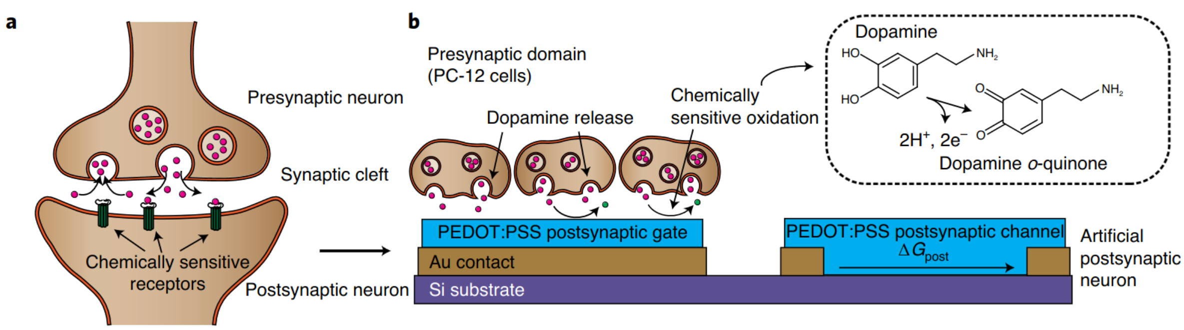

Figure 2: Design and performance of dopamine-mediated organic neuromorphic device. Schematic comparison of a biological synapse (**a**) and neurotransmitter-mediated neuromorphic device (**b**). Modified from Keene et al. (2020).

of duplication.[3] The objective of the present paper is to clearly distinguish simulation from duplication in a way that is applicable to actual systems such as neuromorphic computers. The paper is structured as follows: I start by presenting an account of scientific modeling that considers models to be mathematical structures together with an interpretation of variables and parameters (IVP), as well as an interpretation of transition rules (ITR). The interpretation relates the abstract mathematical structure to the concrete target system. Section 3 provides a general definition of simulation

[3]Searle (1980) characterizes duplication as involving systems "with the same causal powers" (Searle, 1980, p. 423) as the target, yet does not elaborate on how to interpret or operationalize this phrase. Later, I argue that such a definition is too demanding for many theoretical contexts. Somewhat related ideas appear in discussions of experiments versus simulations. However, these accounts do not connect their characterizations to duplication, and limit themselves to describing experiments as involving systems with "the same 'material' causes" (Guala, 2002, p. 67), "the same stuff" (Morgan, 2005, p. 323), or being of "the same kind" (Roush, 2018, p. 4902) as their targets.

and duplication: Simulations are in a certain sense isomorphic to the target, whereas a duplication additionally must have the same relevant measurable properties and work according to the same relevant causal processes. In section 4, I turn to *brain* duplicates and apply the concepts developed in section 3 to brain models. For this, I examine the model developed in Izhikevich and Edelman (2008) and derive criteria that a system must fulfill to be considered a brain duplicate. In the context of neuromorphic computing, a further question is relevant: What is required for duplicating (neural) *computations*? Section 5 and 6 address this issue. I submit that analog neuromorphic computers that use memristive technology are candidates for duplicating neural structures with respect to simple models of neural computation. In section 7, I discuss five possible objections to the proposed account of duplication.

2 Choosing Relevant Causal Powers: Scientific Models

In an early attempt to distinguish simulation from duplication, Searle (1980) focuses on ”machines with internal causal powers equivalent to those of brains” (Searle, 1980, p. 1), in contrast to mere brain simulations. This passage serves as the starting point for defining duplication. In my opinion, there are three interpretations of this sentence: On a *first* reading, duplicating a specific target system might amount to *duplicating all its causal powers*, i.e. sharing all its ”properties that are displayed in causal processes” (Ellis, 2010, p. 1). However, upon closer examination, this definition turns out to be too strong. For example, duplicating all the causal powers of a native Chinese speaker's brain would amount to creating an exact copy of their brain, including all their memories, emotions, and perceptions. Duplication in this sense is certainly not necessary for understanding Chinese. *Second*, ”duplicating the brain” might be understood as “building a brain,” in the sense of artificially creating something that falls under the *natural kind* of brain. However, this definition also seems too demanding. To my understanding, total artificial hearts should be regarded as duplicates, even though they do not fall under the natural kind of heart. Additionally, individuating kinds is a notoriously difficult task. Therefore, I opt for a *third* interpretation that best captures the idea of duplication. Instead of duplicating *all* causal powers, it suffices to duplicate only causal powers that are relevant to the phenomenon under consideration. I elaborate on this idea in the remainder of the paper.

Distinguishing between the relevant and irrelevant causal powers of a given target is an empirical task central to the process of *model* building. Scientists invest significant time and effort in identifying causal processes that are relevant to a given phenomenon and distinguishing them from irrelevant details. Therefore, it is natural to revert to scientific modeling when defining simulation and duplication. In this section, I present a standard account of scientific modeling intended to clarify this distinction.[4] My explanations regarding models largely follow the ideas in Weisberg (2013) and Frigg and Hartmann (2020). I only consider models in the sense of an abstract math-

[4]Even though I believe that my understanding of what a model amounts to, is close to what seems to be the minimal consensus in philosophy of science, I am not claiming to give a complete account of scientific modeling. Rather, my definition could be understood in a more pragmatic sense: It is supposed to provide us with the concepts necessary for making explicit the distinction between simulation and duplication.

ematical structure together with an interpretation, keeping physical models aside.[5] Furthermore, I focus on models in the natural sciences, since the focus of this paper is brain duplication, and I consider the brain sciences, including computational neuroscience, to be natural sciences.[6]

A model is a mathematical structure paired with an interpretation connecting it to the physical world. Typically, the mathematical structure is defined by specifying the parameters and a set of variables together with an equation that defines their evolution. Variables can be either discrete or continuous, and can take values whose range needs to be defined by the model. Similarly, the evolution rules can be discrete or continuous, corresponding to algorithms or differential equations, respectively. Some famous examples of mathematical structures include Kepler's equation and the Lotka-Volterra equations. However, models are not just isolated abstract entities; they stand in a representational relation to concrete objects or systems and are intended to explain real-world phenomena. The real-world system that the model is designed to represent and whose evolution it is supposed to explain is called the target system. For models to represent their target system, we must endow the mathematical structure with an interpretation. This interpretation is twofold. *First*, we need to interpret the parameters and variables by specifying the measurable properties[7] they represent. For example, the variables in Kepler's equation represent the positions of planets and the variables in the Lotka-Volterra equation represent the population densities of predators and prey animals. This is called the *interpretation of the variables and parameters (IVP)*. Next, we must interpret the specific transition rules by specifying the causal processes they describe. For example, Kepler's equations describe the processes of gravitational motion, and the Lotka-Volterra equations describe predation and mating processes. This is called the *interpretation of the transition rules (ITR)*. Combining the same mathematical structure with a different interpretation yields a different model. For example, the Kepler equations can be used in electrodynamics to model hydrogen atoms. In this case, the same variables are not assigned to positions of planets, but positions of electrons, and the transition rules do not describe gravitational motion, but electrostatic Coulomb-attraction (see Nieuwenhuizen and Liska, 2015).

Let us consider an example to illustrate the concepts that we introduced: The *forest-fire model* proposed by Drossel and Schwabl (1992). This model uses a cellular automaton to describe the spreading behavior of a forest fire. Each cell is designated as a number between zero and two, where zero stands for an empty space on the forest floor, one stands for an unburned tree, and two stands for a burning tree. The initial grid is randomly populated with either trees or empty spaces, and one tree in the center is set alight. After each time step, the grid is updated according to the following four rules.

1. A one (unburned tree) (at $t = i$) that has no twos (burning trees) in a neighboring cell stays a one (at $t = i + 1$).
2. A one that has a two in a neighboring cell becomes a two.

[5] On my account, physical models might rather be understood as *duplicates* of their target.

[6] The analysis proposed in this paper could be applied to duplication in the social sciences, if the notions of measurable properties and causal processes were adopted appropriately. However, this is beyond the scope of this paper.

[7] Measurable properties are properties of a system that can be directly observed or indirectly measured. Examples of measurable properties are the temperature of a room, the voltage across a capacitor, the spin of an electron, and the state of a cell being alive or dead.

3. A two becomes a zero (empty space).

4. A zero stays a zero.

The grid for a random initial distribution updated by following these rules is shown in figure 3 for $t = 0, 2, 5, 10, 25$. This model can be used to investigate the evolution of forest fires.

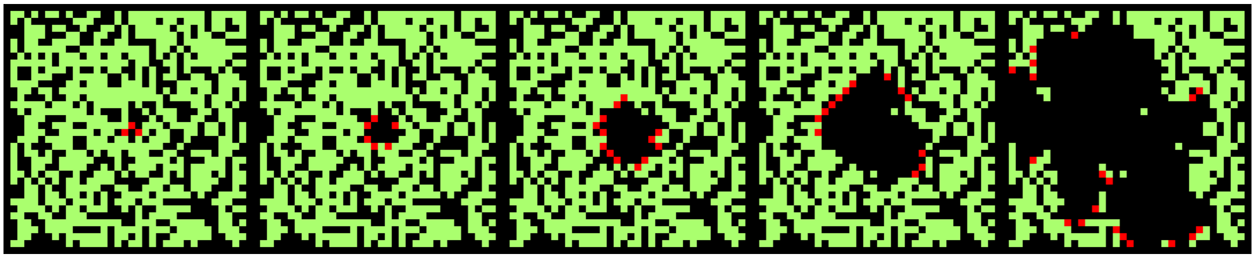

Figure 3: Depiction of cellular automaton at time steps $t = 0, 2, 5, 10, 25$. The green cells represent unburned trees, black cells represent empty spaces, and red cells represent burning trees. Tree population desity was chosen to be $p = .65$.

The *mathematical structure* of the forest fire model is that of a cellular automaton, comprising $n \times n$ variables, one for each cell. To make it a model of a forest fire (rather than that of an infectious disease, let us say), we need an interpretation of the variables and parameters (IVP). To accomplish this, we assign measurable properties of the real-world object to the mathematical variables of the cells. More specifically, we assign the property of whether a certain area is populated by healthy trees, burning trees, or no trees at all. The only parameter that needs to be fixed is the tree population density p. At each time step, there is a mapping between cells with zeros and real empty spaces of the forest floor, cells with ones and real unburned trees, and between cells with twos and real burning trees. Furthermore, we need an interpretation of the transition rules (ITR). The physical processes that motivate the specific transition rules are the thermodynamic processes of heat transfer and exothermic reactions, and thus, the fact that burning trees set neighboring trees alight. The model consists of both: A cellular automaton *as well as* an interpretation, i.e. an IVP and an ITR.

Which mathematical structure is best for describing a given phenomenon? A mathematical structure is a good candidate for a model when it is somehow ”similar” to its target system. On a standard account, there must be an *isomorphism* between the mathematical variables of the abstract structure and the measurable properties of the real-world system. The state of the forest fire model, together with the respective time step, yields the ordered tuple $\{\{m_1, m_2, ..., m_{n^2}, t_m\}_i\}$ where m_i is the variable of individual cells, and t_m denotes time. A cellular automaton can be used to model real forest fires because there is an isomorphism between the set of tuples $\{\{p_1, p_2, ..., p_{n^2}, t_p\}_i\}$ obtained through ecological observations of real forest fires and the set of tuples $\{\{m_1, m_2, ..., m_{n^2}, t_m\}_i\}$ obtained through the Drossel-Schwabl algorithm. This means that for every time t_m we can compute the state of the cellular automaton $\{m_i\}$, and if we then look at our real-world system and observe the state of the trees of the forest $\{p_i\}$ at time t_p, we will find that the tuples $\{\{m_1, m_2, ..., m_{n^2}, t_m\}_i\}$ obtained through mathematical calculations and

$\{\{p_1, p_2, ..., p_{n^2}, t_p\}_i\}$ obtained through measurement on a physical system are isomorphic to each other.

In summary, it can be stated that models consist of a *mathematical structure* paired with an IVP that links variables and parameters to measurable properties of real world systems, and an ITR that states the causal processes that govern the behavior of the target system and motivate the use of the evolution rule.[8]

3 Simulations versus Duplication

Having established the relevant terminology, we now turn to the distinction between simulation and duplication. As we have seen, models are mathematical structures equipped with an interpretation. By contrast, both duplicates and simulations are typically *concrete entities of the physical world*. Simulations are usually processes inside a computer, whereas duplicates can be any type of real-world phenomenon. Intuitively, "a simulation imitates one process by another process" (Hartmann, 1996, p. 5). A straightforward way to cash out this notion is to require that simulation and target share a common mathematical structure. More precisely, I define simulation as follows:

> A system S simulates a target T relative to model M of T if and only if, the measurable properties of S map onto the mathematical states of M.

The construction of a simulation of a target T is a two-step process. First, the target must be modeled. Then, we need to find a system S that imitates the target in the sense that its measurable properties $\{\{s_1, s_2, ..., s_{n^2}, t_s\}_i\}$ map onto the mathematical states $\{\{m_1, m_2, ..., m_{n^2}, t_m\}_i\}$ of the model of T.

On the other hand, *I define a duplicate of a target T, with respect to a model M of T, as any system that M is a model of.* This means:

[8] The notion of a scientific model presented in this paper is not original in any respect but one. To my understanding, no account of modeling differentiates IVP and ITR. Introducing this distinction is valuable beyond its utility in differentiating between simulations and duplicates, in that different models might have the same IVP but different ITRs. Examples of this are not difficult to find in science. To a first approximation, Kepler's model of planetary motion and the relativistic two-body problem yield the same equations (mathematical structure) and represent the same states, viz. the positions of planets (IVP). However, they differ in their ITR in that they disagree on how the change in the planetary position is brought about. Kepler's models posit causal processes that, according to our current understanding, are not actually at work. For instance, Kepler's theory suggests that gravitational forces constrain a planet to an elliptical orbit and counteract the centrifugal force. Einsteinian physics, on the other hand, explains the elliptical path by the bending of space and the fact that objects that are *not* subject to external forces move along geodesics. Another example is the comparison between the Bohr-Sommerfeld model and the quantum mechanical model of the hydrogen atom. Both models accurately describe the position of the electron within the potential of a positive charge (IVP). However, the Bohr-Sommerfeld model interprets the equations as describing classical processes of electron motion, whereas the quantum mechanical model refers to quantum mechanical processes. Introducing the notions of IVP and ITR thus allows for a more nuanced discussion of scientific models.

A physical system D duplicates a target T relative to model M of T if and only if

(1) there is an isomorphism[9] between the measurable properties of D and the mathematical states of M,

(2) the measurable properties of D are those specified by the IVP,

(3) D is governed by the causal processes specified by the ITR.

The idea is this: Even if D is not the *intended* target, it is still a *possible* target of the model of T. Both, the intended target and its duplicate are correctly described by the same model. (1) then follows from the fact that there is an isomorphism between the mathematical structure of a model and the measurable properties of any system that it is a model of. (2) follows from the fact that in order for D to be a possible target of the model, it has to have the measurable properties specified in the IVP. Similarly, (3) follows from the fact that only systems governed by the type of causal process specified in the ITR are correctly described by the model. Conditions (2) and (3) I call the *IVP* and the *ITR criterion* respectively.[10]

To clarify these definitions, I now return to the forest fire model and provide an example. The variables are provided by the cells of the forest fire automaton. One cell can take values ranging from zero to two. The total state of the automaton at time t_m is given by the ordered set of variables, $\{m_1, m_2, ..., m_{n^2}\}$. Any physical system with at least $3 \cdot n^2$ degrees of freedom is, in principle, capable of instantiating an automaton with a grid of $n \times n$ squares, as long as there is a mapping ψ that maps the measurable properties onto the variables: $\psi : \{\{m_1, m_2, ..., m_{n^2}, t_m\}_i\} \mapsto \{\{p_1, p_2, ..., p_{n^2}, t_p\}_i\}$. How does the simulation carried out on my laptop and that produced the results shown in figure 3 represent a cellular automaton?[11] The random-access memory (RAM) in my computer is divided into capacitors that can be either charged or discharged, representing ones and zeros, and are used to save data. Two of these capacitors together represent one cell of the automaton: If both capacitors are discharged ($p =\downarrow\downarrow$), this corresponds to $m = 0$ (an empty space); if the first capacitor is discharged and the second is charged ($p =\downarrow\uparrow$), this corresponds to $m = 1$ (an unburned tree); and if the first is charged and the second discharged ($p =\uparrow\downarrow$), this corresponds to $m = 2$ (a burning tree). In this sense, capacitors represent single cells in the automaton. For each cell, there are two capacitors that represent the cell. If we combine the states of the different cells at one instance of time, we obtain the overall mathematical state of the automaton; if we combine the measurable properties of the different capacitors at one instance of time, we obtain the overall physical state of the computer. When claiming that there must be a mapping between mathematical states and measurable properties, this means that the capacitors mirror what their associated cell is doing. For example, if from one time step to the next, cell one turns from two (burning) to zero (empty), then the corresponding capacitors have to turn from charged discharged ($p_1 =\uparrow\downarrow$) to discharged discharged ($p_1 =\downarrow\downarrow$), and the same for every cell and every corresponding pair of

[9]An isomorphism is a mapping that is one-to-one, onto, and *structure preserving*.

[10]From (1), it follows that every duplicate of S is necessarily also a simulation of S. From now on, whenever I speak of a ”simulation” I mean a ”mere simulation”, i.e. a simulation that does not also meet both the IVP and the ITR criterion.

[11]The following description of the functioning of a computer is over-simplified and, strictly speaking, incorrect, but it nevertheless serves to illustrate the concept of simulation.

capacitors. If the physical system mirrors the mathematical structure in this manner, we can say that it is a simulation.

However, what happens if we move from simulation to duplication of the forest fire? A model assigns *measurable properties* of objects to mathematical variables and parameters. Only objects that have those measurable properties meet the IVP criterion. The forest-fire model uses cells to represent the property of whether a certain area is populated by healthy trees, burning trees, or no trees at all. Although the capacitors in my computer mirror the cells of a cellular automaton, they do not assume the relevant measurable properties and thus do not meet the IVP criterion. Furthermore, the model assumes certain causal processes that give rise to the mathematical structure. Only systems that evolve according to the causal processes of heat transfer and exothermic reactions satisfy the ITR criterion. Thus, computer simulations do not satisfy the ITR criterion either.

However, not all simulations must be computer simulations.[12] Think of the following scenario: In a pedagogical exercise designed to enhance understanding of the Drossel-Schwabl model, a group of students engages in a hands-on simulation activity. This simulation involves the use of a large chessboard that serves as the foundational landscape for exploring the principles of the model. To represent trees within the model, several candles were strategically placed on the board. To initiate the simulation, a single candle located at the center of the board is lit, which symbolizes the starting point of fire propagation. As the simulation progresses, students follow the designated algorithm outlined in the Drossel-Schwabl model by burning and removing burned candles in accordance with the specified dynamics. This simulation meets the IVP criterion, as it represents unburned, burning, and burned trees with objects with the relevant measurable properties. However, it does not meet the ITR criterion because it does not evolve owing to the laws of thermodynamics but due to the intentional actions of the students. Nevertheless, we can consider a similar scenario that counts as a duplicate, in that both criteria are met.[13] Imagine that the students do not use candles to represent trees, but rather matches that are arranged such that a burning match sets neighboring matches alight. The ”simulation” is then carried out by burning a match in the center of the board and watching the fire spread on its own. Such a system consists of objects that have the relevant measurable properties *and* is governed by the causal laws of thermodynamics and heat transfer. Thus, it meets all criteria for duplication.

4 Brain Duplication: Putting the Theory to the Test

In this section, I apply the concepts developed above to the brain model described in Izhikevich and Edelman (2008) to make plain the features a system must possess to be considered a brain

[12]”[Only] [i]f the simulation is run on a computer, it is called a computer simulation.” (Hartmann, 1996, p. 5)

[13]It is not possible to envision a system that satisfies the ITR but fails to meet the IVP criterion, since IVP and ITR are not logically independent criteria. Having the relevant measurable properties is a necessary condition for changes driven by appropriate causal processes. A system cannot evolve according to the relevant causal process of heat transfer if it does not contain parts that can be unburned, burning, or burned.

duplicate.[14] The following discussion omits some details and should be understood more as proof of concept than an in-depth analysis. Izhikevich and Edelman (2008) describe a large-scale model of the mammalian thalamocortical system that comprises one million spiking neurons and half a billion synapses that exhibit dopamine-modulated plasticity. Let us begin by identifying the *mathematical structure* of the model. The model consists of a set of two ordinary differential equations for each neuron:

$$C\dot{v} = k(v - v_r)(v - v_t) - u + I, \tag{1}$$
$$\dot{u} = ab(v - v_r) - au, \tag{2}$$

where C is the capacitance, and v is potential of the cell's membrane, u represents the voltage due to the accumulated effect of all inward and outward voltage-gated currents, and I is the combined dendritic and synaptic current. Additionally, there are four ordinary differential equations for each synapse:

$$\dot{c} = \frac{-c}{\tau_c} + STDP(\tau)\delta(t - t_{pre/post}), \tag{3}$$
$$\dot{s} = cd, \tag{4}$$
$$\dot{d} = \frac{-d}{\tau_d} + DA(t), \tag{5}$$
$$\dot{G} = (1 - G)/\tau_G, \tag{6}$$

where s denotes the synaptic strength measured as the postsynaptic potential, c is the activation of an enzyme important for plasticity, d is the concentration of dopamine, and G represents the synaptic conductivity. This results in a complex mathematical structure comprising more than two billion coupled differential equations.

The precise statements of the above equations are not relevant to the present argument. However, they serve to understand the model's *IVP*. The variables and parameters in equations (1) and (2) are assigned to objects that have a certain *potential* v and *capacitance* C and are *traversed by a current* I. Equations (3) to (6) are assigned to objects with a certain *conductivity* G, a (post-synaptic) *potential* s, a *dopamine concentration* d, and an *enzyme activation level* c. Izhikevich and Edelman (2008) describe their model as phenomenological, meaning it deliberately avoids proposing specific causal processes. However, the neuronal model used in Izhikevich and Edelman (2008) is described in more detail in Izhikevich (2003) as a computationally efficient version of the biophysically accurate Hodgkin-Huxley model. Therefore, it is fair to assume that the same causal processes that motivate the use of the Hodgkin-Huxley equations are used to interpret the neuronal transition rules (1) and (2). Thus, the ITR of the neuron model is the dynamics of voltage-gated ion channels, electrochemical gradients established by active transport mechanisms, and the passive electrical properties of the neuronal membrane. The interpretation of the synaptic transition rules (3) to (6) includes biochemical processes like dopamine mediation, exocytosis, and neurotransmitter reuptake.[15]

[14] This specific model has been chosen because it has been extensively discussed and offers a lot of detail. However, I do not want to commit to this as the definitive model but rather use it as an example to elaborate on my account.

[15] A more detailed description of the synaptic model can be found in Izhikevich (2007).

A *simulation* of this model would be any physical process that maps onto the two billion differential equations. Izhikevich and Edelman (2008) describe such simulations that have been run on a Beowulf cluster of 60 processors in great detail. Any brain *duplicate* must fulfill further criteria. It must include millions of objects with the measurable properties specified by the model (IVP). Additionally, it must function according to the neuro- and biochemical processes described above (ITR). From the requirement of fulfilling the IVP and ITR criteria, we can derive a couple of necessary conditions for a system to count as a brain duplicate: (1) It must consist of (vast amounts of) objects with the measurable properties of a neuron, i.e. an appropriate capacitance, potential and current. It must also comprise (vast amounts of) objects with the same measurable properties as synapses according to the model. (2) The neuron-like objects must have spiking potentials. (3) Dynamics must be based on ion transport mechanisms. (4) The synapse-like objects must be dopamine- and enzyme-mediated and (5) exhibit both short- and long-term plasticity. (6) Neuron- and synapse-like objects must be appropriately interconnected.

Further conditions could be derived through a more thorough examination of the model at hand. The objective of this section, however, is not to provide an argument for or an in-depth discussion of the model in Izhikevich and Edelman (2008), but rather to show that the terminology of the present paper can be applied to existing models to derive concrete criteria for brain duplication. Further research, both philosophical and empirical, is needed to fully determine which criteria should be met for a system to fulfill the IVP and ITR criteria.

5 Duplicating Computations

Some might argue that it is not necessary to duplicate the biological processes of the brain, and that it is more important to duplicate the *computations* it performs. To determine which systems are capable of duplicating physical computations, we must understand what models of computation are and how they relate to scientific models, as described in section 2. If computations are potential targets for duplication, they must be understood as *physical computations*. Given computational neuroscience's self-image as a natural science, only physical systems can be the target of scientific modeling. Therefore, a model of computation is a scientific model that describes a physical computation. What exactly do scientists mean when they, for example, say that "artificial neural networks (ANNs) [are] a model class well suited [for a complete picture of how cognition emerges]" (Doerig et al., 2023)? Unfortunately, philosophical literature on models of computation is sparse and not well integrated into broader debates on scientific models. However, a few points seem to be well-understood. First, the formal models studied in computability theory are first and foremost abstract entities that do not necessarily refer to any physical computation. Initially, formal models, such as Turing machines, were developed to solve problems in mathematical logic and were not used to describe real-world target systems. Thus, they cannot be considered scientific models without additional interpretation. Second, computational models, i.e. models that use formal computational structures to model non-computing systems, are *not* to be considered models of computation (Miłkowski, 2014; Anderson and Piccinini, 2024, Ch. 8). The Drossel-Schwabel forest fire model is not a model of computation even though it uses a cellular automaton as its mathematical structure, since it does not describe a computing system. In the following, I outline an account that fulfills these requirements.

A model of computation, as I use the term, specifies a certain type of physical computation. More precisely, it describes a *computational architecture*.[16] What distinguishes computations from other physical processes is that they are essentially *medium-independent*.[17] Whether my computer has a hard disk drive and uses magnets to store data or whether it has a solid-state drive and uses trapped electrons as memory cells does not matter with respect to the computations it performs. As computations are medium-independent, *models of computation are models that can be fully specified without reference to any specific physical measurable properties and processes.* Consequently, the IVP and the ITR of a model of computation differ from those of other scientific models. In the non-computing case, the IVP links each variable and parameter of the mathematical structure to one *specific* measurable property, and the ITR states one *specific* (set of) physical processes that relate those variables and parameters. For example, the $n \times n$ variables of the forest fire model relate to the $n \times n$ measurable properties of healthy, burning, and burned trees. In the case of models of computation, however, many different types of physical processes can serve as the ITR and the measurable properties that are involved in this process can be considered the IVP. The IVP links each variable and parameter of the mathematical structure to *any* measurable property, which changes its values according to the mathematical structure, and the ITR states *any* (set of) physical processes that explain their evolution. However, this is not to say that structural properties are irrelevant in models of computation. Most models specify precisely how their fundamental components are connected. For example, not any physical process that is isomorphic to a 2D cellular automaton could serve as its ITR, as cellular automata put structural constraints of adjacency on the parts of the computing system.

A simple example of a model of computation is that of a logic gate. The following equations describe the behavior of an AND gate:

$$f(x, y) = \begin{cases} 1 & \text{if } x = 1 \text{ and } y = 1 \\ 0 & \text{otherwise} \end{cases} .$$

Whenever both inputs are one, the output itself is one, and it is zero in all the other cases. The model of the AND gate then pairs this mathematical structure not with a specific measurable property and not with one specific causal process. Rather, the parameters and variable can be interpreted with *any* measurable property that changes their values accordingly (IVP), and the transition rules can be interpreted by *any* physical process that relates those magnitudes in the right way (ITR). Most

[16]The notion of computational architecture is closely linked to the mechanistic account of computation set out by Piccinini (2015). A computational architecture comprises a system's computing components and how they are organized and function together to determine the system's overall computational abilities. Architectures are medium-independent, meaning the computing components can be implemented in various hardware systems without altering the architecture itself. Examples include register machine architectures, as well as ANNs. This concept is most fully developed in Piccinini (2025), as well as Fuentes and Piccinini (2025).

[17]This thesis has been contested. For a recent defense of the medium-independence of computation see Fuentes (2024) and Drayson (2025)

logic gates are built from electronic circuits, and the inputs and outputs are represented as voltages:

$$V_{out} = \begin{cases} V_H & \text{if } V_x \geq V_{th} \text{ and } V_y \geq V_{th} \\ V_L & \text{otherwise} \end{cases}.$$

However, this is not prescribed by the model of the AND gate, as different measurable properties might be used instead. For example, some AND gates use magnetization to represent ones and zeros:

$$M_{\text{out}} = \begin{cases} M_\downarrow & \text{if } M_x = M_\downarrow \text{ and } M_y = M_\downarrow \\ M_\uparrow & \text{otherwise} \end{cases}.$$

Both physical systems are correctly described by the model of the AND gate.

This notion of a model of computation has direct implications for what is considered a simulation and a duplicate of a given computation. A simulation of a computation can be defined in exactly the same way as a simulation of any other physical process as a system whose states map onto the mathematical structure of the model of computation. However, duplication goes beyond simulation in that not every physical process that maps onto the mathematical structure is considered a duplicate. Duplicates must fulfill further criteria: First of all, a subtle difference between simulations and duplicates that was of minor importance in the context of non-computational duplication, gains importance in the case of computational duplicates. Whereas the measurable properties of a simulation merely *map onto* the mathematical states, the measurable properties of a duplicate must be *isomorphic* to it. This means they must be *structure preserving*. Figuratively speaking, this means that an increase/decrease of the mathematical state by a certain proportion must be mirrored in an equivalent increase/decrease in the value of the corresponding measurable property. This is due to the fact that the predictions of the model must align with the results of potential measurements. For example, if a model predicts that a given forest fire expands by fifty percent during the course of a day, then the real forest fire has to expand by exactly that proportion. Otherwise the model would be flawed. Similarly, if a model of computation prescribes the increase of one of its variables by fifty percent, then the value of the corresponding measurable property needs to increase by that proportion.

Second, compared to models of non-computing processes, the *IVP criterion* needs to be relaxed to account for the medium independence of computation. A model of computation only specifies that there is *a* measurable property for each variable and parameter. For example, a magnetic AND gate duplicates an electronic AND gate with respect to the model of the AND gate.[18] On the face of it, this requirement seems rather weak. However, note that not all properties are measurable properties. For example, the capacitor arrays in the ANN simulation have the property of representing floating point numbers.[19] Those floating point numbers *are* isomorphic to the mathematical structure. However, this is not a property that can be measured using an appropriate device. Rather, it is a quality that we attribute to it based on our interpretation. More on this below.

[18] The magnetic AND gate, however, does not duplicate the electronic AND gate with respect to an electro-dynamic model. This is due to the fact that computationally speaking both gates are equivalent whereas they differ in their physical makeup. This illustrates the medium-independence of models of computation.

[19] A common encoding scheme is the IEEE 754 standard: binary32 floating point format that uses one voltage level to represent the sign of a number, eight levels to represent the exponent, and 23 levels to store the significand. The

Third, the *ITR criterion* needs to be relaxed as well. Given their medium-independence, models of computation are agnostic to which specific physical processes they describe. However, they do delineate certain structural requirements that the target system must satisfy regarding fundamental aspects of its computational behavior, such as whether it operates globally or locally, in a parallel or serial manner, and whether its processing is integrated or not. For example, a (physical) one tape Turing machine might be isomorphic to a (physical) two-tape Turing machine. Nevertheless, the former does not duplicate the latter because, despite their medium-independence, models of two-tape Turing machines put structural constraints on the physical processes they model. Namely, these systems must consist of two separate tapes.[20] Models of computation remain agnostic regarding the physical medium, not specifying which measurable properties or processes they describe. However, they do impose structural constraints on the way their targets compute.

6 Duplicating Neural Computations

Next, let us consider computations in the brain and the question of which criteria must be fulfilled for a system to duplicate the brain with respect to models of neural computation. Arguably, the most prominent family of models of neural computation are ANNs (cf. Doerig et al., 2023). ANNs are models composed of interconnected neurons, where each connection has an associated weight, and each neuron applies an activation function to process inputs and generate outputs. In most ANNs, the activation of a neuron is computed by applying an activation function to the weighted sum of its inputs (plus a bias)[21]:

$$a_j = \phi_j \left(\sum_i x_i w_{ji} \right), \tag{7}$$

where ϕ_j is the activation function[22], x_i is the activation of the incoming neurons, and $w_{j,i}$ is the synaptic weight of the incoming neurons.

A system S is considered a simulation of a given ANN if its measurable properties $\{\{s_1, s_2, ..., s_n, t_s\}_i\}$ map onto the states of the ANN $\{\{x_1, x_2, ..., x_m, w_{1,1}, w_{1,2}, ..., w_{m,o}, t_{ANN}\}_i\}$. Most ANNs are simulated in this sense, as the RAM states of a digital computer map onto the weights and activations of the ANN. Every weight and activation level is stored as a floating-point number in RAM during every program run. This means that the measurable properties of usually 32 charged and discharged capacitors map onto the weights and activations of the mathematical structure of the ANN.

value can then be computed as value $= (-1)^{\text{sign}} \times 2^{(E-127)} \times \left(1 + \sum_{i=1}^{23} b_{23-i} 2^{-i}\right)$, where E is the exponent in the binary representation derived from interpreting the levels of high and low voltages as one and zero, sign is the sign bit, and b_i are the different significand bits.

[20]It goes without saying that tapes do not have to be made of paper; they could be made from any other suitable medium as well.

[21]For reasons of simplicity, I omit the bias in the following.

[22]Most ANNs use a ReLU activation function.

According to the definition provided above, duplication requires an isomorphism between measurable properties and mathematical structure as well as fulfilling the medium-independent IVP and ITR. For this, the duplicating system D must have a set of measurable properties that is isomorphic to the parameters and variables of the ANN (IVP).[23] Furthermore, they must be governed by a physical process that adheres to the structural constraints of the ANN model. Most notably, the artificial neurons—i.e., objects with a measurable property representing a certain activation level—must interact locally with adjacent artificial synapses—i.e., objects with a measurable property representing associated weights (ITR). A very illustrative example of a duplicate of an (extremely simple) neural structure according to the ANN model is the Mechanical Neural Network as described by Schaffland and Schöning (2023). It consists of wooden levers, whose x-displacement are isomorphic to the activation levels, connected via strings and pulleys, whose positions are isomorphic to the synaptic weights.

Before we turn to neuromorphic computers, let me explain why a simulation of an ANN on standard digital hardware does *not* meet the criteria for duplication. A genuine duplication of a physical computation requires an isomorphism between the measurable properties of the original system and those of its candidate duplicate (IVP). Specifically, if a model variable representing a neural measurable property increases or decreases, the corresponding measurable property in the duplicating system must also systematically increase or decrease by the same amount. In digital simulations, however, variables such as activation levels or connection strengths are encoded in binary form, typically as patterns of charged or discharged capacitors within memory cells. When a simulated variable changes by a certain percentage, the voltages across these capacitors do not change proportionally or in a manner structurally isomorphic to the original neural process. Instead, the changes in capacitor states merely encode new values symbolically, without corresponding to any continuous or proportionate physical change. Thus, the measurable properties within a digital computer do not physically increase or decrease as the simulated neural variables do. It follows that digital brain simulations, while capable of formally representing neural processes, do not physically duplicate the underlying computations performed by biological brains. Figuratively speaking, it is not possible to find any measurable properties in the simulation that can be substituted for the variables and parameters of the equations described in section 4.[24]

Now, I elaborate on whether neuromorphic computers can duplicate neural computations according to a member of the ANN family. First, which requirements do neuromorphic chips have to fulfill in order for them to meet the ITR criterion? According to the present account, they must represent neural activation levels with measurable properties that change their values in accordance with the mathematical structure. As most ANNs have real-valued weights and connections, they

[23]Note that, in contrast to non-computational scientific models, these measurable properties are not further specified by the ANN model as, for example, membrane potentials and synaptic resistance. This is due to the medium-independent nature of computation.

[24]Note that not all systems that are computationally equivalent duplicate each other. Computational equivalence is usually spelled out in terms of simulation: Two computational systems O and P are computationally equivalent, if and only if O is capable of simulating P, and vice versa, i.e. if the states of O map one-to-one and onto the states of P. However, O only duplicates P if they are correctly described by the same model of computation, i.e. if they satisfy the medium-independent IVP and ITR. ANNs and digital computers are computationally equivalent; however, no digital computer can duplicate neural structures according to an ANN model.

must therefore be *analog*.[25] Neuromorphic chips sometimes use continuous voltages, currents, or spike frequencies to represent neural activation levels. These measurable properties *are* isomorphic to the variables they represent; that is to say, they increase and decrease in proportion to the activation level they implement. For instance, a twofold increase in activation is reflected by an equivalent increase in voltage across the implementing capacitor. Furthermore, components of a duplicate must directly represent the synaptic weights $w_{i,j}$ with measurable properties. Good candidates in this context are chips that use memristors as artificial synapses because they encode synaptic weights in varying resistances. A change in the synaptic weight of the ANN by a certain proportion is reflected by an equivalent change in the memristor's resistance. Therefore, the real-valued synaptic weights comprising the ANN are isomorphic to the real-valued resistances of the memristors that implement them. Which additional constraints must be fulfilled to meet the ITR criterion? The ITR criterion requires the system to satisfy the structural constraints imposed by the ANN model. For example, a given ANN describes which nodes are connected to which other nodes. Neuromorphic computers might be built to respect that structure by connecting artificial neurons and synapses in precisely the way that the ANN describes. Although the number of hardware components in neuromorphic computers is usually much smaller than the number of nodes and connections in state-of-the-art ANNs, neuromorphic chips that directly couple artificial hardware neurons with artificial hardware synapses could meet the ITR criterion for small-scale models of neural computation.

7 Objections and Replies

7.1 Duplication is Model Dependent

In this section, I test my account of duplication by addressing five potential criticisms and providing responses to them. An obvious objection could be formulated as follows: ”Your account is explicitly model-dependent. This seems unreasonable, given that duplication in most cases does not involve modeling in any obvious way. Why would I need to create a model of a target before I can duplicate it?”

In response, I want to stress that a model doesn't necessarily need to be explicitly formulated for it to mediate between a target and its duplicate. For example, the matches-on-chessboard system described in section 3 duplicates the forest fire with respect to the Drossel-Schwabl model irrespective of whether it is built by a conscientious student to study the model or by a lazy student who is just playing around with the matches. Either way, the system is correctly described by the Drossel-Schwabl model, and therefore counts as a duplicate. Furthermore, I provide three reasons for why it is reasonable to understand duplication in a model-dependent way:

First, simulations are inherently model relative, as highlighted by Hartmann (1996). When simulat-

[25]This use of the term 'analog' follows the common distinction between chips that are digital, analog, or mixed signal. It is not meant to imply that analog computation is essentially computation via primitive physical magnitudes (Lewis, 1971; Maley, 2018), rather than continuous representations (Papayannopoulos, 2020).

ing a target system, a model of the system is often developed first to capture its essential dynamics and interactions of its components. A given simulation of a target is only a simulation with respect to a certain model. For example, a forest fire can be simulated according to the Drossel and Schwabl (1992) model or according to the model in Rothermel (1972). Both yield different simulations of the same target. Duplicates are similar to simulations in many regards. Simulation simpliciter is not possible, so we shouldn't expect duplication simpliciter.

Second, as previously mentioned, duplication cannot involve replicating all causal powers, as it would result in a perfect copy. Instead, duplicating a target requires abstracting away from irrelevant details and determining which properties are relevant to a given phenomenon. "When talk is of "relevant" and "irrelevant" details, we must always ask, "to whom?" and "for what purposes" (Chirimuuta, 2024, p. 7)? This form of abstraction is at the core of scientific practice, making it natural to rely on scientific models when addressing the concept of duplication.

Third, the model-dependent nature of duplication aligns well with our intuition. For example, total artificial hearts are prime examples of duplicates. They can be implanted to replace the biological heart. Also, 3D-printed in vitro hearts made from human tissue (Noor et al., 2019) are archetypes of heart duplicates. They are made from the same type of cells and are anatomically similar to the real heart. However, total artificial hearts and in vitro hearts are wholly different from each other and both are individually different from real hearts. The former is made of ceramics and metals, while the latter cannot function as a heart. The intuition that different systems can duplicate the same target can be straightforwardly explained by assuming a model-dependent notion of duplication. Total artificial hearts duplicate the human heart with respect to fluid dynamical models, whereas in vitro hearts duplicate it with respect to immunological and anatomical models.

7.2 How Similar is Similar Enough?

A further worry might be the following: "In nature, no two structures are ever exactly isomorphic: There are always fluctuations, noise, and small differences in measurement. Taken at face value, your account thus implies that duplication is impossible for natural systems, since there will never be a perfect isomorphism between target and duplicate."

Strictly speaking, the account does not require that target and duplicate be directly isomorphic to one another. Rather, it requires that both be isomorphic to a common mathematical model. Given the transitivity of isomorphism, however, target and duplicate are thereby isomorphic as well. But the underlying worry is not specific to my notion of duplication; it is the familiar problem that no scientific model is ever perfectly isomorphic to a real-world system. The standard way to address this in the philosophy of modeling is to introduce a fidelity or tolerance condition. Weisberg (2013, p. 41-42), for example, speaks of "dynamic fidelity" in this context, understood as the requirement that the measured values of the target system stay within an accepted margin of error of the model's predicted trajectories. On this view, a model does not specify only a mathematical structure and an interpretation thereof; it also implicitly or explicitly specifies how close the model's variables must be to the target's measured properties. This strategy directly carries over to the case of duplication.

In practice, duplication does not require that every measured value in the duplicate exactly match the corresponding measured value in the target. Rather, for those measurable properties and dynamical relations that the model deems relevant, the measurements in the duplicate must fall within an appropriate margin of error of those in the target, where the relevant error bounds are given by the same sort of fidelity criteria that determine when the model itself is an adequate representation of its target.

7.3 Which Details to Ignore

Furthermore, a critic of the proposed account might argue: "On your account, a system D duplicates a target T with respect to a model M iff both D and T are correctly represented by M. Intuitively, this means D duplicates T if D matches T in all and only those respects that M deems relevant. But then it follows that if D and T are duplicates in this sense, simply adding extra detail that is irrelevant according to M should not break the duplication relation. So, for every D and T we can add as much irrelevant detail in which they are different and they would still count as duplicates. To see that this is problematic, consider the following example: A particular forest fire in the north of Germany duplicates another forest fire in the west of Portugal with respect to the Drossel-Schwabel model. Most features of Portugal, be it the railroad system, the cities' architecture, the height of the terrain, etc. are irrelevant according to the Drossel-Schwabel model. The same holds for the railroad system, the cities' architecture, and the height of the terrain in Germany. Therefore, it follows that not just the forest fire in Germany duplicates that in Portugal but that Germany itself duplicates Portugal. This consequence is absurd."

My reply is that this consequence does not follow on the proposed account, because the Drossel–Schwabl model is not a model of Portugal as a whole or Germany as a whole, but a model of a forest fire. D duplicates T with respect to M only if both D and T are correctly represented by M. In the forest-fire case, this condition is satisfied only when T is taken to be the forest-fire subsystem. Neither the country of Portugal nor Germany as a whole is correctly represented by the Drossel–Schwabl model; hence, they are not even candidates for duplication with respect to that model.

A natural rejoinder to this reply might be the following: "Why does the Drossel–Schwabl model not count as a (very inaccurate) model of Portugal? After all, models can be highly idealized and omit many details. Moving from the forest fire to Portugal just adds more detail that the model happens to leave out; why not treat that as an extreme case of inaccuracy and keep the duplication claim? After all, the Drossel-Schwabel model is by itself highly idealized and leaves out many properties of real forest fires."

However, for a model to (accurately or inaccurately) represent a target T, it is generally assumed that it needs to fulfill what Frigg and Nguyen (2021, p. 3- 5) call the 'Surrogative Reasoning Condition'. This means that M counts as a scientific model of T only if one can use M to conduct surrogative reasoning that yields claims about T. Even highly idealized models must still be suitable vehicles for drawing (possibly inaccurate) inferences about their intended targets. Applied to the present example, the Drossel–Schwabl model does satisfy the Surrogative Reasoning Condition when T is taken to be the forest-fire. It can be used to reason for example about the spread and percolation in that fire. But it does not satisfy that condition when T is taken to be Portugal as a whole (or Germany as a whole). All claims that can be derived from the model are claims about a forest fire, not claims about a country.

7.4 Emergent Properties

When it comes to duplicating neural computations with respect to an ANN model, the following objection might be raised: "Even if the low-level physical states of a digital computer (for example, voltages across individual capacitors) are not themselves isomorphic to the mathematical description of the neural computation, there might nonetheless be higher-level, emergent states or patterns that are. On this view, the relevant physical states of a digital simulation are not the microphysical configurations themselves, but the structured patterns that represent the computational states. For instance, in a simulated neural network, specific configurations of capacitors in my laptop represent the activation levels of neurons. One can then point to a property that is isomorphic to the states of the target computation, namely, the property of representing a particular numerical value (e.g., a floating-point number corresponding to a membrane potential). This is analogous to the thermodynamic notion of temperature: Temperature is not identified with any single molecule's state, but with an emergent high-level property that is rulefully determined by the distribution of microstates. In this sense, the floating-point convention is analogous to statistical mechanics in that it maps low-level states to high-level emergent properties. As the example of temperature shows, emergent properties can be the object of scientific investigation just as much as low-level properties."

In reply, I want to stress that emergent properties are not problematic for my account, nor do I deny that they can be scientifically modeled. The crucial issue is not micro- vs. macro-level, but whether there is an objective, convention-independent measurable property that is isomorphic to the model variables. From this perspective, the thermodynamic analogy does not hold up. Temperature is indeed emergent from molecular motion, but it is a measurable macro-quantity: There is a non-arbitrary ordering from colder to hotter, grounded in systematic differences in (for example) average kinetic energy. Different temperature scales (Celsius, Kelvin, Fahrenheit) are merely different notations for the same underlying ordered property; they preserve the same physical ordering and ratios (cf. Tal, 2020, Sec. 3). By contrast, in the floating-point case, the property "representing value 0.93" or "representing value 3.34" is not a measurable property with an intrinsic ordering, but rather assigned to a particular bit pattern by a chosen coding scheme. There is no convention-independent metric or order on these memory states. Changing the representational standard can lead to the very same physical states representing different numbers, such that a state that previously represented a smaller number now represents a larger number.

Furthermore, note that even if representing a certain floating point number was a measurable property – which it is not- the ITR criterion would still be violated as the capacitors that represent the activation levels of two connected neurons are in no more causal contact than those representing activation levels of neurons that are not connected (cf. Kohár, 2024). The medium-independent structural constraints are thus not fulfilled in the simulation.

7.5 Virtual Duplicates

A further objection might be raised as follows: "Even if the physical hardware of a digital computer does not itself instantiate the right kind of measurable properties, a computer simulation of a target might nevertheless be sufficient for duplication—at least in the computational case—because it gives rise to *virtual* objects with *virtual* properties that can be measured *virtually*. For example, Chalmers (2017) famously defends a robust realism about virtual reality, claiming, for example, that "(1) Virtual objects really exist. (2) Events in virtual reality really take place. (3) Experiences in virtual reality are non-illusory" (Chalmers, 2017, p. 310). If we assume that computer simulations are like virtual reality and that the simulated objects really exist, then it is valid to assume that reading a value from a computer simulation counts as measuring an existing virtual property. And if the simulation is accurate, those virtually measurable properties are isomorphic to the mathematical structure. For instance, a brain simulation contains virtual neurons and synapses that possess virtual properties, such as membrane potentials and firing rates, which evolve according to the mathematical structure of the model. Therefore, computer simulations are virtual duplicates of their targets."

In response to this, first note that this move cannot rescue non-computational duplication. Even Chalmers concedes that virtual properties are not identical to the corresponding physical properties they mirror (Chalmers, 2017, p. 321). Virtual voltages are not voltages in the usual physical sense. Non-computational duplication, as I understand it, requires that the same type of property be instantiated. Substituting a virtual analogue therefore does not meet the requirement. One might still hope that the move works for computational duplication: In this case it is not important that the very same property be present, but only that there is some property—e.g. "having virtual potential"—that evolves according to the mathematical structure of the model. I also reject this claim, for two reasons. *First*, strong realism about virtual reality is itself contentious, in particular with respect to virtual causation. Kohár (2024), for instance, raises serious difficulties for treating virtual events as straightforwardly causally efficacious in the way that physical events are. This has direct consequences for the notion of duplication, as any indeterminacy regarding the causal status of the virtual events carries over to the causal processes that figure in the ITR. *Second*, even if the virtual realist's position works for immersive virtual reality, it is less plausible in the context of computer simulations. Beisbart (2024) develops a compelling case against importing virtual realism into the context of computer simulations. He identifies an underdeterminacy problem and argues that the computational structure of the computer simulation is insufficient to determine which virtual objects and virtual properties it gives rise to. As discussed earlier, the same computational structure might be a simulation of a forest fire, giving rise to virtual trees that are virtually burning, *and* a simulation of the spread of a disease giving rise to virtual patients that are virtually infected. "A main attraction of virtual realism is that it allows us to take talk about simulated cells, cats, etc. at face value, but the problem is that the underlying computer calculations are too poor to determine what the objects really are" (Beisbart, 2024, p. 44). So even if we grant that virtual realism is plausible for certain kinds of virtual reality experiences, this does not entail that computer simulations give rise to virtual duplicates.

8 Outlook and Conclusion

This paper has aimed to elaborate on the often-blurred distinction between simulation and duplication and to make these notions precise enough to apply them to existing and future neuromorphic computers. For this, I first presented an account of scientific models that construes them as mathematical structures, together with an interpretation of variables and parameters, as well as an interpretation of transition rules. In section 3, I defined simulations as systems that share a mathematical structure with the target. More strictly speaking, a system S simulates a target T if and only if the measurable properties of S map onto the mathematical states of the model of T. I defined duplicates of a target T as those systems that the model of T is a model of. Relating back to the thoughts in section 2, this means that a duplicate of a target T not only shares a mathematical structure with it but also has the same relevant measurable properties and operates in accordance with the same causal processes as T. The general notion of duplication was applied to brain models in section 4 to derive criteria for brain duplication. Six criteria for a system to count as a brain duplicate have been derived: (1) It must consist of (vast amounts of) objects with the measurable properties of a neuron, i.e. an appropriate capacitance, potential and current. It must also comprise (vast amounts of) objects with the same measurable properties as synapses according to the model. (2) The neuron-like objects must have spiking potentials. (3) Dynamics must be based on ion transport mechanisms. (4) The synapse-like objects must be dopamine- and enzyme-mediated and (5) exhibit both short- and long-term plasticity. (6) Neuron- and synapse-like objects must be appropriately interconnected. Section 5 and 6 addressed the question of whether neuromorphic computers can duplicate neural computations. The considerations showed that analog neuromorphic computers that use memristive technology are candidates for duplicating neural structures with respect to simple models of neural computation. In section 7, I discussed five possible objections to my account of duplication.

It is interesting to note that each criterion for brain duplication, with the possible exception of criterion (6), appears to be met by existing technologies. Systems such as SpiNNaker2 (Mayr et al., 2019) consist of huge numbers of hardware neurons. More bio-realistic neuromorphic devices, such as those in Mahowald and Douglas (1991) have spiking potentials very similar to those of biological neurons (see figure 1). The system described in Agarwal et al. (2017) uses ion transport mechanisms. Keene et al. (2020) describe a dopamine-mediated organic neuromorphic device. And for example, memristive devices are capable of exhibiting neural plasticity. Nevertheless, no currently existing neuromorphic computer meets all the criteria for brain duplication, nor does any come close to doing so. Furthermore, it is likely that additional criteria have yet to be identified. However, duplicating the biology of the brain may not be necessary to achieve many cognitive faculties; duplicating *neural computations* may suffice. As discussed in section 6, existing neuromorphic computers duplicate certain aspects of neural computations. With big tech companies and leading research institutions investing heavily in its development, neuromorphic computing is one of the most promising candidates for replacing the von Neumann paradigm in the future. In anticipation of this, philosophers should consider the implications of creating brain-like computers. I have taken a first step in this direction by exploring the extent to which neuromorphic computers duplicate the brain.

Acknowledgments

I am especially grateful to Nikola Kompa for the many constructive discussions and her practical advice during the development of the manuscript, as well as to Gualtiero Piccinini for his extensive feedback on earlier drafts of this paper. I would also like to thank Gregor Hörzer and Christian Nimtz, as well as the participants of the Oberseminar Philosophy Osnabrück and the MüBiOs colloquium, for their valuable comments and suggestions. My thanks further go to three anonymous referees for their helpful and constructive remarks. I would also like to thank the German Research Foundation DFG which supported this project in the context of funding the Research Training Group "Situated Cognition" (GRK 274877981).

Bibliography

Abu-Hassan, K., Taylor, J., Morris, P., Donati, E., Bortolotto, Z., Indiveri, G., and Nogaret, A. (2019). Optimal solid state neurons. *Nature Communications*, 10(1):5309.

Agarwal, S. et al. (2017). Achieving ideal accuracies in analog neuromorphic computing using periodic carry. In *2017 Symposium on VLSI Technology T13.2*. IEEE.

Anderson, N. G. and Piccinini, G. (2024). *The Physical Signature of Computation: A Robust Mapping Account*. Oxford University Press.

Beisbart, C. (2024). Do computer simulations include digital artifacts? *Metaphysics*, 7(1).

Chalmers, D. J. (1995). Absent qualia, fading qualia, dancing qualia. *Conscious experience*, pages 309–328.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chalmers, D. J. (2017). The virtual and the real. *Disputatio: International Journal of Philosophy*, 9(46).

Chirimuuta, M. (2024). *The Brain Abstracted: Simplification in the history and philosophy of neuroscience*. MIT Press.

Chua, L. (1971). Memristor-the missing circuit element. *IEEE Transactions on Circuit Theory*, 18(5):507–519.

Davies, M., Srinivasa, N., Lin, T., Chinya, G., Cao, Y., Choday, S., and Wang, H. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., and Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7):431–450.

Drayson, Z. (2025). Defending the medium-independence of computation. *Minds and Language*.

Drossel, B. and Schwabl, F. (1992). Self-organized criticality in a forest-fire model. *Physica A: Statistical Mechanics and its Applications*, 191(1-4):47–50.

Ellis, B. (2010). Causal powers and categorical properties. In *The Metaphysics of Powers*, pages 141–150. Routledge.

Frigg, R. and Hartmann, S. (2020). Models in science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, summer edition.

Frigg, R. and Nguyen, J. (2021). Scientific representation. Stanford Encyclopedia of Philosophy (Winter 2021 Edition). Accessed 2026-01-23.

Fuentes, J. and Piccinini, G. (2025). A mechanistic account of physical computation. Unpublished manuscript.

Fuentes, J. I. (2024). Computational systems as higher-order mechanisms. *Synthese*, 203(2):55.

Guala, F. (2002). Models, simulations, and experiments. In *Model-based reasoning: Science, technology, values*, pages 59–74. Springer US.

Guala, F. (2005). *The Methodology of Experimental Economics*. Cambridge University Press, Cambridge, UK.

Hartmann, S. (1996). The world as a process: Simulations in the natural and social sciences. In *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, pages 77–100. Springer Netherlands, Dordrecht.

Hinsch, K. and Zupanc, G. K. H. (2007). Generation and long-term persistence of new neurons in the adult zebrafish brain: a quantitative analysis. *Neuroscience*, 146(2):679–696.

Indiveri, G. (2021). Introducing 'neuromorphic computing and engineering'. *Neuromorphic Computing and Engineering*, 1(1):010401.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral Cortex*, 17(10):2443–2452.

Izhikevich, E. M. and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, 105(9):3593–3598.

Keene, S. T., Lubrano, C., Kazemzadeh, S., Melianas, A., Tuchman, Y., Polino, G., and Santoro, F. (2020). A biohybrid synapse with neurotransmitter-mediated plasticity. *Nature Materials*, 19(9):969–973.

Kohár, M. (2024). Illusionism about virtual causation. *Inquiry*, pages 1–26.

Lewis, D. (1971). Analog and digital. *Nous*, 3:321–327.

Mahowald, M. and Douglas, R. (1991). A silicon neuron. *Nature*, 354(6354):515–518.

Maley, C. (2018). Toward analog neural computation. *Minds and Machines*, 28:77–91.

Mayr, C., Hoeppner, S., and Furber, S. (2019). Spinnaker 2: A 10 million core processor system for brain simulation and machine learning. *arXiv preprint arXiv:1911.02385*.

Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–1636.

Meng, J., Song, J., Fang, Y., Wang, T., Zhu, H., Ji, L., and Chen, L. (2024). Ionic diffusive nanomemristors with dendritic competition and cooperation functions for ultralow voltage neuromorphic computing. *ACS Nano*, 18(12):9150–9159.

Merolla, P., Arthur, J., Alvarez-Icaza, R., Cassidy, A., Sawada, J., Akopyan, F., and Modha, D. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673.

Miłkowski, M. (2014). Computational mechanisms and models of computation. *Philosophia Scientiae*, 183(3):215–228.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology*, 12(2):317–329.

Neckar, A., Fok, S., Benjamin, B. V., Stewart, T. C., Oza, N. N., Voelker, A. R., and Boahen, K. (2018). Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model. *Proceedings of the IEEE*, 107(1):144–164.

Nieuwenhuizen, T. M. and Liska, M. T. (2015). Simulation of the hydrogen ground state in stochastic electrodynamics. *Physica Scripta*, 2015(T165):014006.

Noor, N., Shapira, A., Edri, R., Gal, I., Wertheim, L., and Dvir, T. (2019). 3d printing of personalized thick and perfusable cardiac patches and hearts. *Advanced Science*, 6(11):1900344.

Papayannopoulos, P. (2020). Computing and modelling: Analog vs. analogue. *Studies in History and Philosophy of Science Part A*, 83:103–120.

Parker, W. S. (2009). Does matter really matter? computer simulations, experiments, and materiality. *Synthese*, 169:483–496.

Pehle, C., Billaudelle, S., Cramer, B., Kaiser, J., Schreiber, K., Stradmann, Y., Weis, J., Leibfried, A., Müller, E., and Schemmel, J. (2022). The brainscales-2 accelerated neuromorphic system with hybrid plasticity. *Front*.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. OUP Oxford.

Piccinini, G. (2025). Neural hardware for the language of thought: New rules for an old game. *arXiv preprint arXiv:2510.10251*.

Rothermel, R. C. (1972). A mathematical model for predicting fire spread in wildland fuels. Technical report, Intermountain Forest & Range Experiment Station, Forest Service, US Department of Agriculture.

Roush, S. (2018). The epistemic superiority of experiment to simulation. *Synthese*, 195(11):4883–4906.

Schaffland, A. and Schöning, J. (2023). Mechanical neural network: Making ai comprehensible for everyone. In *2023 IEEE 2nd German Education Conference (GECon)*, pages 1–6. IEEE.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457.

Serrano-Gotarredona, T., Masquelier, T., Prodromakis, T., Indiveri, G., and Linares-Barranco, B. (2013). Stdp and stdp variations with memristors for spiking neuromorphic learning systems. *Frontiers in Neuroscience*, 7:2.

Tal, E. (2020). Measurement in science. Stanford Encyclopedia of Philosophy (Fall 2020 Edition). Accessed 2026-01-23.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press, New York, NY.