Article
Physical Grounds for Causal Perspectivalism

Gerard J. Milburn [1], **Sally Shrapnel** [1,*] and **Peter W. Evans** [2]

1   Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St. Lucia, QLD 4072, Australia; g.milburn@uq.edu.au
2   School of Historical and Philosophical Inquiry, The University of Queensland, St. Lucia, QLD 4072, Australia; p.evans@uq.edu.au
*   Correspondence: s.shrapnel@uq.edu.au; Tel.: +61-7-33656931

Abstract: We ground the asymmetry of causal relations in the internal physical states of a special kind of open and irreversible physical system, a causal agent. A causal agent is an autonomous physical system, maintained in a steady state, far from thermal equilibrium, with special subsystems: sensors, actuators, and learning machines. Using feedback, the learning machine, driven purely by thermodynamic constraints, changes its internal states to learn probabilistic functional relations inherent in correlations between sensor and actuator records. We argue that these functional relations just are causal relations learned by the agent, and so such causal relations are simply relations between the internal physical states of a causal agent. We show that learning is driven by a thermodynamic principle: the error rate is minimised when the dissipated power is minimised. While the internal states of a causal agent are necessarily stochastic, the learned causal relations are shared by all machines with the same hardware embedded in the same environment. We argue that this dependence of causal relations on such 'hardware' is a novel demonstration of causal perspectivalism.

Keywords: agents; causality; control; thermodynamics; learning

1. Introduction

Is causation in the external, physical world or in our heads? Russell [1] famously denied the former, while the latter seems unacceptably subjective. The interventionist account of causation [2,3], especially when interpreted along perspectival lines [4], seems to be somewhere in between, "an irenic third way", in the words of Price [5]. We demonstrate here a plausible physical schema within which causal claims depend for their truth on the internal physical states of a special kind of machine, a causal agent. We take this to be an exemplification of a perspectival view of causation that is not anthropocentric and is dependent on the laws of physics, especially thermodynamics. Our objective is to give empirical support to Price's "causal viewpoint" as "a distinctive mix of knowledge, ignorance and practical ability that a creature must apparently exemplify, if it is to be capable of employing causal concepts" [5]. In Cartwright's terms, we physically "ground the distinction between effective strategies and ineffective ones" [6]. Or, as Ismael writes, "Causes appear in mature science not as necessary connections written into the fabric of nature, but robust pathways that can be utilised as strategic routes to bringing about ends" [7]. We seek these "robust pathways" in the physical structure of learning machines.

We take it as given that causal concepts find no application in the fundamental physics of closed systems. This is, we think, precisely what Russell [1] had in mind when denying the physical reality of causation. A system is closed if it cannot interact with anything outside itself. The physical dynamical laws of a closed system are reversible in time. In contrast, how we define an open system depends on where we draw the boundary. We call a system 'open' if its internal interactions are much stronger than any interactions with systems external to it. Of course, such a characterisation is highly contextual, and is dependent on, say, the nature of our experimental access to the system and the time scale

of such access: in the long run, even small interactions will matter. The dynamical laws that describe open systems are irreversible and stochastic. It is our contention that casual relations can only be understood in terms of open systems.

We describe a special kind of open system—a causal agent (CA)—which we take to be maintained in a *non-thermal-equilibrium* steady state. A CA contains specialised subsystems: sensors, actuators, and learning machines. The external world does work on the CA through special subsystems called *sensors*. The CA does work on the world through special subsystems called *actuators*. There is an essential thermodynamic asymmetry to sensors and actuators: in both cases, the operation of the subsystem dissipates energy. The learning machine is an irreversible physical system that exploits functional relations inherent in correlations between sensor and actuator states in order to optimise the use of thermodynamic resources. These 'causal relations' become embodied in the bias settings of the learning machine. As above, learning machines are necessarily dissipative and noisy [8,9].

We make a distinction between two ways one can describe the irreversible dynamics of an open system, which we call the *Langevin* and *Einstein* descriptions. The Langevin description of an open system is a fine-grained description that provides a complete specification of the values of the stochastic physical states of a single system as a function of time. The Einstein description of an open system is a coarse-grained description in terms of the probability distribution of internal physical states or, equivalently, average values of physical quantities. While the probability distributions of the Einstein description can be stationary, the Langevin description of physical states is stochastic. For example, a single two-level atom interacting with a thermal radiation field at fixed temperature may be described in terms of its stationary Boltzman distribution—the Einstein description—or in terms of a stochastic switching between its internal states at rates determined by the temperature of the environment—the Langevin description (Figure 1).
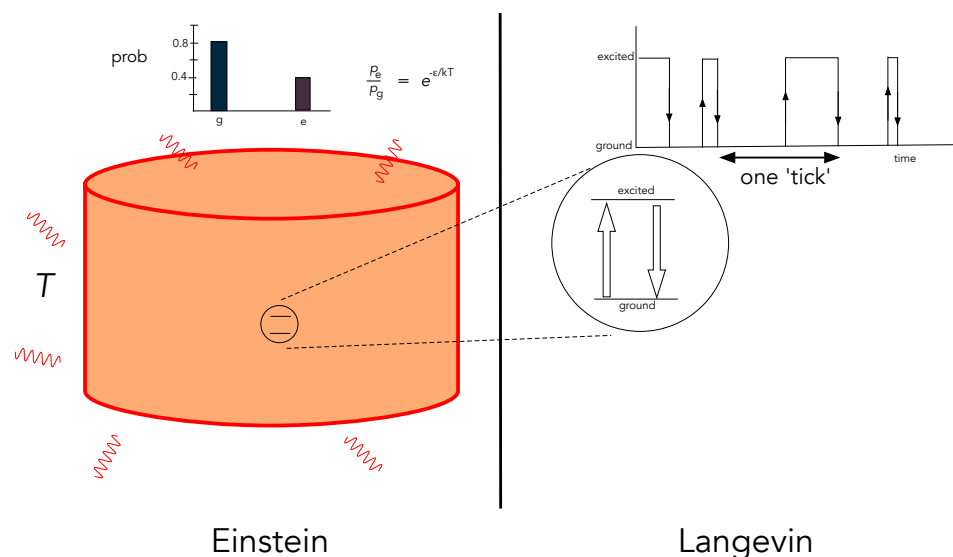

Figure 1. A depiction of the distinction between the Einstein and Langevin descriptions of an open system: a two-level atom interacting with a thermal radiation field inside a cavity at temperature T. On the left is the Einstein description in terms of the stationary probability distribution to find the atom in the ground state (g) or excited state (e). On the right is the Langevin description in terms of the stochastic record of the continuously observed state of a single atom as a function of time.

The two descriptions we highlight are based on the two equivalent ways of describing diffusion given by Langevin and Einstein. Importantly, the Langevin and Einstein descriptions both refer to objective physical states as seen from a 'third person' perspective with different levels of grain and different epistemic access to those states.

One of our primary goals in this work is to show how causal relations can be defined in terms of the physical states of the learning machine of the CA. In terms of the Langevin description, the stochastic record of sensations and actions of the CA are coupled by feedback to the irreversible process of the learning machine. The effect is to drive the learning machine stochastically to a new state in which various internal settings are fixed by implicit correlations between sensor and actuator records. The internal settings (weights and biases) of the learning machine continue to fluctuate once the leaning is complete; nonetheless, they enable prediction of the sensations that follow an action with little probability of error. On this view, the causal relations *just are* the internal settings of the learning machine. (We might imagine that these internal settings are simply the modelling parameters of the causal model constructed by the CA).

In the Einstein description, the probability distribution over the internal settings of the learning machine evolves by feedback to a new stationary distribution centred on particular domains in the state space of the learning machine. While the probability distributions over sensor and actuator records can remain completely random, the probability distribution of the settings of the learning machine are 'cooled' to particular steady-state distributions that correspond to the learned causal relation implicit in the correlations between sensor and actuator records. We use the term 'cooled' to emphasise that the dynamics of the learning machine are dissipative. It lowers its entropy at the expense of increasing the entropy of its environment.

We wish to note at this point that our construction of a causal agent has considerable overlap with similar constructs by other authors. The importance of sensors and actuators for artificial agents is a staple of textbooks on artificial intelligence [10] that provide the raw data for learning algorithms. Briegel [11] also stresses the importance of sensors and actuators for embodied agents. His novel concept of 'projective simulations' plays the role of a learning machine in our model. He emphasises the role of stochasticity for creative learning agents and possible quantum enhancements. In elucidating his concept of action-based semantics, Floridi [12] describes a two-machine artificial agent (AA). The 'two machines' of his scheme roughly correspond to the actuator/sensor machines and the learning machine in our model. Our model of a prediction/correction learning machine is close to the agent model introduced in a biological setting by Friston [13] and subsequent developments [14]. Lastly, Freeman [15] and Llinas [16] link causation directly to learning and motor processes in central nervous systems. We take our model to be in the spirit of these approaches to agency and learning.

Our argument proceeds as follows. In Section 2, we discuss a very simple classical dynamical model of a causal agent and illustrate the distinction between the Langevin and Einstein descriptions of the CA. In Section 3, we discuss the thermodynamics of sensors and actuators and the fundamental asymmetry between them. In Section 4, we introduce the concept of a learning machine (not a machine learning algorithm!) as an irreversible physical system coupled by feedback to the sensors and actuators. We give a simple example of how a learning system can embody causal concepts in its physical steady states. In Section 5, we discuss the thermodynamic constraints on learning machines, and in Section 6, we discuss how learning machines based on prediction and correction feedback can learn causal relations. In Section 7, we make explicit how our model of learning in a CA is an exemplification of causal perspectivalism.

2. A Simple Physical Model

In this section, we begin to develop our model of a causal agent (CA) using a simple example. The objective of the example is to introduce the three essential components of a CA—sensors, actuators, and learning machine—in terms of a simple physical system. We first give the Einstein description of each component followed by the Langevin description. This highlights key thermodynamic processes and irreversibility. The learning machine component can be described both in terms of a physical feedback process and in terms of learning a causal functional relation.

Consider the following simple classical dynamical system (Figure 2). A local system, the 'source', can emit particles of variable kinetic energy. It is driven by an external power supply towards a nonequilibrium steady state by dissipating power into a local environment at temperature T_s. The emitted particles travel towards a small potential hill. If they have enough kinetic energy, they can surmount the hill and never return to the source. If they do not have sufficient kinetic energy they will be reflected from the potential hill and return to the source. We assume that the motion of the particles once they leave the source is entirely conservative; that is to say, the particles move without friction. As particles that surmount the barrier are lost, we assume that the source is supplied by a particle reservoir such that the average particle number of the source is constant in time. But to make sure that particles with sufficient energy to surmount the barrier do not encounter a yet higher barrier, we add a particle absorber to the right of the potential hill. Overall, this entire physical arrangement is an irreversible system.
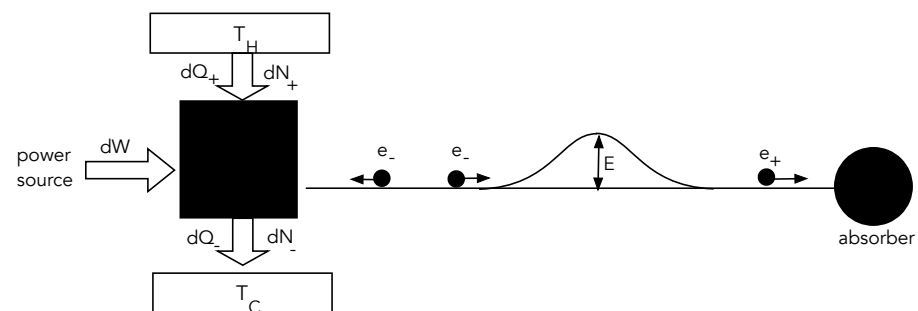

Figure 2. The system on the left, the source, is connected to a source of work (increment dW) and also two thermal/particle reservoirs at different temperatures. Higher energy particles are added to the system from the high-temperature reservoir and lower energy particles are emitted from the source into a low-temperature reservoir (increments $dQ_\pm, dN_\pm$). Doing work on the source will bias it to emit predominantly low-energy particles, thus cooling it and reducing the heating rate of the absorber.

Suppose the source initially emits particles which can take one of two possible values of kinetic energy (e_-, e_+), with equal probability, such that $e_- < E$ and $e_+ > E$. All particles that have energy e_+ are absorbed at the right, thereby raising the internal energy of the absorber. All particles that have energy $e_- < E$ are returned to the system and absorbed. The system has access to an environment that emits/absorbs particles locally to keep the average particle number and average energy constant. The entropy of the source is one bit in natural units and its average energy $(e_- + e_+)/2$. The average energy of the absorber is steadily increasing as it absorbs particles from the source. This energy is heat-extracted from the source's environment.

Assume that the CA operates in time steps of duration T. At the nth time step, it emits an e_+ particle with probability $p_+(n)$ and an e_- particle with probability $p_-(n)$. For simplicity, we assume that these are the mutually exclusive events that can happen in each time step so that $p_+(n) + p_-(n) = 1$. This is clearly a Bernoulli process with probabilities that depend on time.

2.1. Einstein Description

According to the Einstein description of the system, the source is simply a source of heat and particles. Since the laws of thermodynamics tell us that there are no perfect absorbers, we know that eventually the absorber will start emitting something—even if ultimately it becomes so massive that it undergoes gravitational collapse to form a black hole and emit Hawking radiation. We can, however, make the general assumption that the absorber is at a lower temperature than the source, and so heat transferred from the source to the absorber raises the overall entropy. As a result, the total system is not in thermal equilibrium. The Einstein description can thus be completely specified in terms of the total

average number of particles $\bar{N}(t)$ available at any time and the average number of particles of each species as distinguished by energy $\bar{n}_\pm(t)$ corresponding to energies $e_\pm$.

At the nth time step, it emits an e_+ particle with probability $p_+(n)$ and an e_- particle with probability $p_-(n)$ and $p_+(n) + p_-(n) = 1$. In Figure 3, we plot sample trajectories with constant $p_\pm$ for the actuator and sensor records, where $a(n) = 1$ if an e_+ particle is emitted and $a(n) = 0$ if an e_- particle is emitted. For the sensor, we only need that $s(n) = 1$ if a particle is returned; otherwise, it is zero.
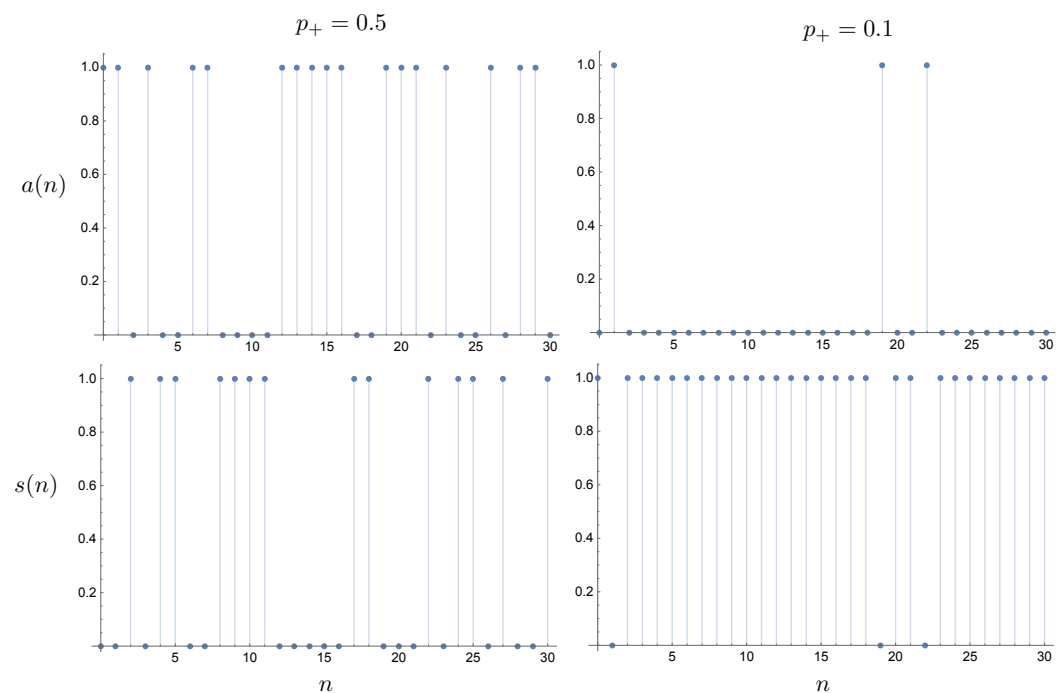

Figure 3. Examples of the stochastic record of a sensor and actuator as a function of time step n for two values of p_+.

It is clear that energy is lost whenever an e_+ particle is emitted, so clearly a mechanism needs to be implemented such that this becomes increasing unlikely based on past experience of the CA. In the Einstein description, this requires changing the discrete stochastic process governing particle emission as a function of time. This turns the Bernoulli process into a nonlinear discrete process. We assume that the fractional change in $p_+(n)$ from one time step to the next is decreasing,

$$\frac{p_+(n+1) - p_+(n)}{p_+(n)} = -\lambda p_+(n) \tag{1}$$

where $0 \le \lambda < 0$ is a constant that depends on the feedback circuit. This may be written as

$$p_+(n+1) = p_+(n)(1 - \lambda p_+(n)) . \tag{2}$$

This is related to the logistic map and has an attractor at $p_+ = 0$ given the restrictions on λ, at which point $p_- = 1$, exactly as required.

2.2. Langevin Description

The Langevin description is given in terms of a history of actions and sensations. We first define a binary record for actions, $a(n)$, and sensations, $s(n)$. In each time step, the actuator record tracks the energy of each emitted particle: $a(n) = \mathbf{0}$ if an e_- particle was emitted and $a(n) = \mathbf{1}$ if an e_+ particle was emitted. In each time step, the sensor record tracks whether a particle was received back: $s(n) = \mathbf{0}$ if no particle is received back and $s(n) = \mathbf{1}$ if a particle is received back. The logical relation of the sensations as a function of

actions is binary NOT. One way that we can imagine the operation of the machine is that it is required to simulate the NOT function with small probability of error.

The energy lost at each time step is a random variable given by

$$\Delta E(n) = a(n)e_+ \,, \tag{3}$$

as every e_- particle is returned and counted by the sensor. In Figure 4 we plot $\Delta E(n)$ for three sample trajectories.
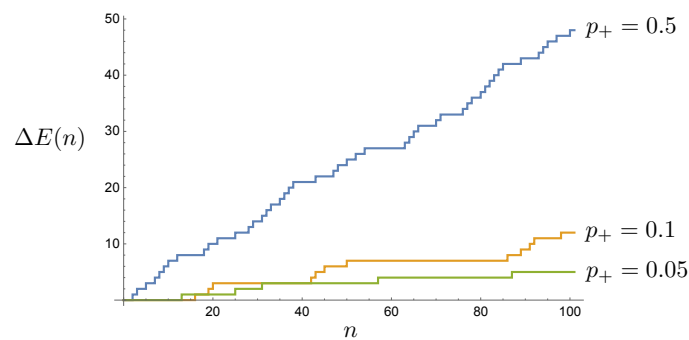

Figure 4. Two sample trajectories of the stochastic energy dissipated by the CA as a function of time step n for three values of p_+.

It is clear now what needs to happen in order to reduce the energy dissipated, thereby learning the correlation between actions and sensations. The machine needs to reduce the probability that it emits a high-energy particle; that is to say, p_+ is a decreasing function of n. The simplest way to do this is to add a feedback process from sensor to actuator record at each time step conditioned on the sensor record at that step, as the sensor only records a **1** if a particle is returned, and that only happens for low-energy particles. A simple feedback protocol is

$$
\begin{aligned}
p_+(n+1|s(n)) &= p_+(n)(1 - \lambda(1 - s(n))) & (4)\\
p_-(n+1|s(n)) &= p_-(n) + \lambda(1 - s(n))p_+(n) \,, & (5)
\end{aligned}
$$

where $s(n) \in \{0,1\}$ is the sensor record at the nth time step, and $0 < \lambda < 1$ is a feedback parameter, and we have explicitly acknowledged that these probabilities are conditioned on the sensor records. These equations ensure that $p_+(n) + p_-(n) = 1$. Keep in mind that $s(n)$ is a random variable, so these equations are intrinsically stochastic. As we have defined $s(n) = 1 - a(n)$, we can write these equations as

$$
\begin{aligned}
p_+(n+1|a(n)) &= p_+(n)(1 - \lambda a(n))) & (6)\\
p_-(n+1|a(n)) &= p_-(n) + \lambda a(n)p_+(n) \,. & (7)
\end{aligned}
$$

Averaging over an ensemble of actuator histories, this reduces to the nonlinear stochastic map of the Einstein description.

We assume the initial conditions $p_+(0) = p_-(0) = 1/2$. The solution to these equations for $\lambda << 1$ is given by

$$
\begin{aligned}
p_+(n|N) &= \frac{1}{2}e^{-\lambda N(n)} & (8)\\
p_-(n|N) &= \frac{1}{2}e^{-\lambda N(n)} + \left(1 - e^{-\lambda N(n)}\right) \,, & (9)
\end{aligned}
$$

where

$$N(n) = \sum_{j=1}^{n} a(j) \tag{10}$$

is the total count for the actuator over this time interval and is also a random variable. Clearly, p_+ goes to zero and p_- goes to one. It is important to note that the conditional probabilities entering into Equation (8) are conditioned on the entire stochastic history, $N(n)$, which is a particular kind of non-Markovian behaviour. This is analogous to the Bellman equation [17] in stochastic control theory.

The energy lost per step is the given by

$$\Delta \bar{E}(n) = \frac{e_+}{2} \mathbb{E}[e^{-\lambda N(n)}] \,, \tag{11}$$

where $\mathbb{E}$ indicates an ensemble average over the random variable $N(n)$. Evaluating this ensemble average is equivalent to deriving a stochastic fluctuation identity for this model [18]. We see that the average energy dissipated per step decreases exponentially at a rate that depends on the strength of the feedback λ. This can be regarded as the learning rate.

In what sense is this kind of feedback described as learning? The answer takes us to the connection between learning and thermodynamics, or between information and physics. The objective of the machine is to minimise energy cost operating in its environment. At each time step, the sensor record as a function of the action record can be described as a binary NOT. But the machine knows nothing about Boolean functions. All it has access to are its sensor and actuator records and an imperative to minimise the energy dissipated in its interactions with the world. If it can find a way to do this, it will have learned correlations implicit in its actuator and sensor records. We return to this question in Section 5.

In reality, both actuators and sensors have a small error probability, so that the record does not exactly match what actually happened in the actuator and sensor devices. Due to such errors, it is impossible to reach a state in which the sensor record is composed entirely of **1** s. That is an unphysical state of zero entropy. All that is required is that the probability of finding a zero in the sensor record is very small.

This example is very simple, so the causal law it discovers is also simple: if it emits a particle with low energy, it is most likely to be returned, otherwise not. This is the 'law of physics' from the perspective of this CA. Emitting a particle is an intervention. The sensor responds accordingly, and the agent learns the causal relation as a result of its intervention. The internal records of the CA are completely contingent. As the actuator/sensor records are random binary strings, every CA will, almost certainly, have different records. The physical internal state of each CA is unique but, according to the Einstein description, the internal records are unknown (by definition), and every CA of this type behaves as an equivalent thermodynamic machine minimising average thermodynamic cost and thereby lowering its average entropy.

3. Thermodynamic Constraints on Sensors and Actuators

In this section, we formulate general principles that capture key features of the simple example introduced in the previous section. There is a thermodynamic asymmetry between sensors and actuators. Actuators do work on the world, whereas the world does work on sensors. Work done on/by a system is constrained by the change in the free energy (Helmholtz or Gibbs). The change in the free energy constrains the work done on/by the sensors and actuators of a CA. The average work done by a CA must be less than the decrease in free energy. Physical changes to the sensors *increase* the free energy of the CA, while actuators *decrease* the free energy of the CA. These thermodynamic asymmetries must be built into the physical construction of sensors and actuators and are a defining feature of a causal agent.

Average work, and corresponding changes in free energy, are part of the Einstein description of a CA. To provide the Langevin stochastic description of the process, we make use of the Jarzynski equality [18]. On this account, work is a random variable conditioned on contingent physical processes inside the agent. The Jarzynski equality relates the statistics of this stochastic process to changes in free energy.

A mechanical example of the thermodynamic asymmetry of actuators and agents is given in Figure 5. The objective of the sensor is to detect a collision of a large 'signal' particle, while the objective of the actuator is to eject a large 'probe' particle. The dissipation and noise is represented by a large number of much smaller particles colliding with the mechanism.
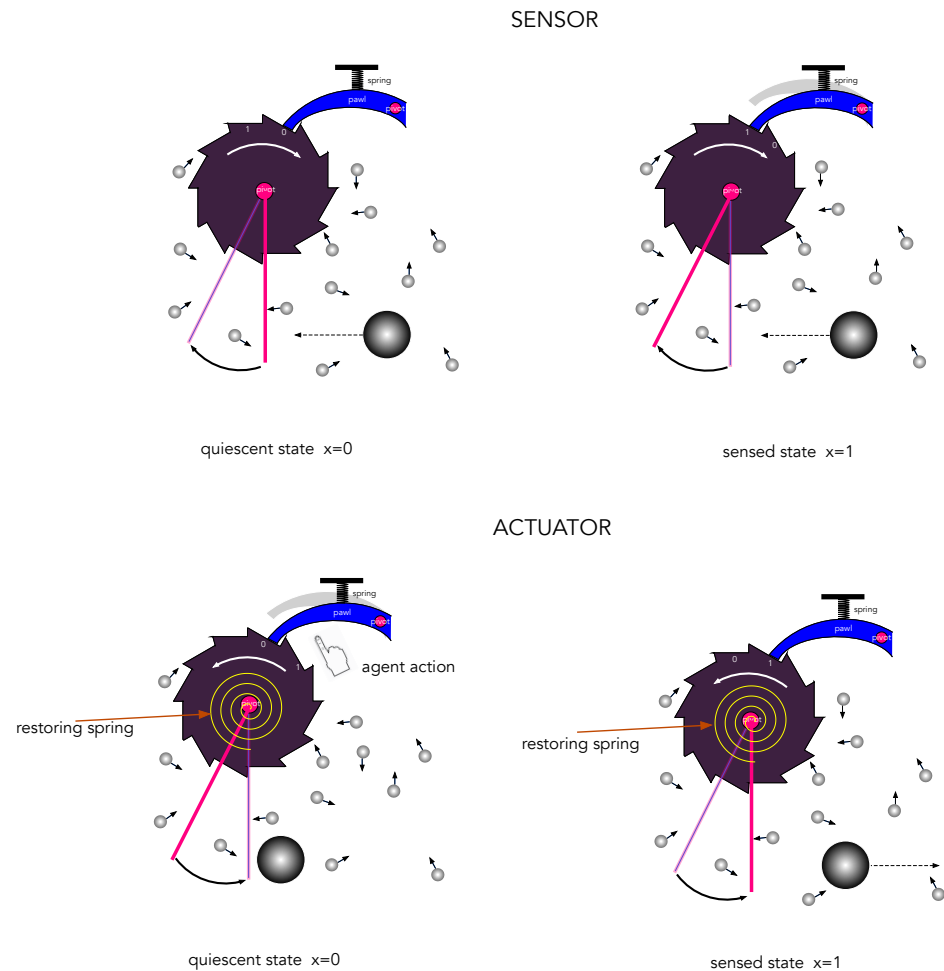

Figure 5. Example of a mechanical sensor and actuator based on a ratchet and pawl. The state of each device is represented by the binary numbers written on the ratchet. The collision of the 'signal' particle is inelastic and leads to an irreversible transfer of energy to the device through a combination of elastic compression of the pawl spring and friction. The signal particle does work on the sensor. In the case of the sensor, gravity applies a continuous force to maintain the quiescent state of the pendulum arm. The continuous collision of the background particles slowly restores the quiescent state. In the case of the actuator, a coiled spring in the ratchet (constantly rewound by the agent) applies a constant torque opposing gravity to maintain the quiescent locked state of the ratchet. The agent 'acts' by pushing the pawl off the ratchet, the pendulum falls to eject a 'probe' particle, increasing its kinetic energy and thus extracting work from gravity. The continuous collision of the background particles slowly restores the quiescent state due to the constant force applied by the restoring coiled spring inside the ratchet mechanism. The background collisions with the noise particles occasionally switches the device, corresponding to an error.

Let us assume for simplicity that both actuator and sensors have only two physically distinguishable states labelled by a binary variable $x = \{0, 1\}$ and that the energies of each state E_x are such that $E_0 < E_1$. We assume that in the absence of actions and sensations, each system is highly likely to be found in an initial 'ready' state. In the case of a sensor, the ready state is the lower energy state, and in the case of an actuator, the ready state is

the high-energy state. In the Einstein description, the system is described by a probability distribution $p_x(t)$. In the Langevin description, the system is described by the value of a binary stochastic variable $x(t)$.

We use a two-state birth–death master equation model to give the Einstein description of sensors and actuators. In the absence of interactions between the agent and the external world, the occupation probability for each state is the stationary solution to the two-state Markov process of the form

$$\frac{dp_0}{dt} = -\gamma_1 p_0 + \gamma_0 p_1 = -\frac{dp_1}{dt} \, , \tag{12}$$

where γ_1 corresponds to a transition $0 \to 1$ and γ_0 corresponds to the transition $1 \to 0$. The corresponding stationary distributions are then given by

$$\frac{p_1(\infty)}{p_0(\infty)} = \frac{\gamma_1}{\gamma_0} \, . \tag{13}$$

The conditions that distinguish the quiescent state of sensors and actuators (the 'ready' state for sensations and actions) are

$$\gamma_1 \quad < \quad \gamma_0 \qquad \text{(sensor)} \tag{14}$$
$$\gamma_1 \quad > \quad \gamma_0 \qquad \text{(actuator)} \, . \tag{15}$$

In the case of a sensor, prior to a sensation, it is more likely to be found in the lower energy state $x = 0$ than the higher energy state $x = 1$. In the case of an actuator, prior to an action, it is more likely to be found in the higher energy state $x = 1$ than the low-energy state $x = 0$.

It is important to stress, however, that neither the sensor nor the actuator are in thermal equilibrium with their environment. They are actually in nonequilibrium steady states due to external driving of a dissipative system, which in this case is the causal agent itself. In classical physics, the rates $\gamma_{0,1}$ go to zero as the temperature goes to zero. In the quantum case, they may not go to zero at zero temperature due to dissipative quantum tunnelling (as in optical bistability, for example [19]).

The stationary distributions give the time-independent Einstein description, but an individual system is certainly not stationary in the Langevin description; it is switching between the two states at rates determined by γ_0, γ_1. The two descriptions are connected by time averaging. In a long time average, the ratio of the times spent in each state is given by the ratio of the transition probabilities

$$\frac{\tau_1}{\tau_0} = \frac{\gamma_1}{\gamma_0} \, , \tag{16}$$

in the limit that the total time $\tau_1 + \tau_0 \to \infty$. Thus, prior to actions and sensations, the sensor spends more time in the lower energy state $x = 0$, while the actuator spends more time in the higher energy state $x = 1$.

On the stochastic Langevin account, we are interested in describing the energy of a single system. In the current example, $x(t)$ is a stochastic process (a random telegraph signal). We define two Poisson processes $dN(t)_x$ ($x = \{0, 1\}$) that can take the values $0, 1$ in a small time interval t to $t + dt$. In most small time intervals, the state of the system does not change, but every now and then, the system can jump from one state to the other. If a jump does happen, one or the other of $dN(t)_x = 1$ in the infinitesimal time interval $t \to t + dt$. The probability to take the value 1 in time interval dt is then simply

$$P(dN_x(t) = 1; t + dt, t) = \mathcal{E}(dN_x) = \gamma_x dt \, , \tag{17}$$

where $\mathcal{E}$ defines the average. These equations imply that the continuous record of the state label $x(t)$ satisfies the stochastic differential equation,

$$dx(t) = (1 - x(t))dN_+(t) + x(t)dN_-(t) .\qquad(18)$$

The internal states of other components in the agent are responding to these fluctuating signals at all times. The agent is said to be in a ready state, or *quiescent state*, if the time average of the signal corresponds to the stationary states in Equation (13).

We now describe how these devices respond to internal (actuator) and external (sensor) inputs. In both cases, bias forces act on sensors and actuators in such a way as to make the transition rates time-dependent. This is how work is done on/by the system, during which time heat is dissipated as they are pushed away from their quiescent steady states. We refer to these inputs as the control functions. To show this, first we define

$$\gamma_0 = \gamma e^{-\epsilon/2}\qquad(19)$$
$$\gamma_1 = \gamma e^{\epsilon/2} .\qquad(20)$$

where we refer to ϵ as the bias. If ϵ is a constant, the steady-state average energy is

$$\bar{E} = E(1 + e^{-\epsilon})^{-1} ,\qquad(21)$$

This function is plotted in Figure 6. To set the devices to their quiescent state, a particular bias value ϵ_0 is chosen. In the case of a sensor, $\epsilon_0 < 0$ and $\bar{E}$ is small, while in the case of an actuator, $\epsilon_0 > 0$ and $\bar{E}$ approaches E.
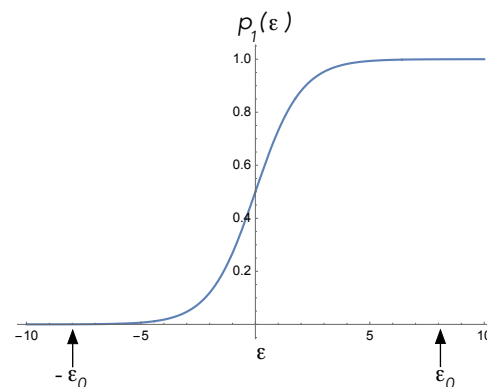

Figure 6. The probability for the sensor and actuator to be found in the high-energy state as a function of a bias parameter ϵ. The bias values for the quiescent states of each are indicated by $\epsilon_0 < 0$ for a sensor and $\epsilon_0 > 0$ for an actuator. When internal and external control functions act, the bias value is changed from ϵ_0 to $\epsilon_f = -\epsilon_0$.

Physical forces inside (actions) or outside (sensations) the CA change the value of ϵ from the quiescent state bias, ϵ_0, at time t_i to a final value, ϵ_f, at time t_f. We assume that the irreversible dynamics of the devices is sufficiently fast enough that at each time step they rapidly relax to new steady states. The occupation probabilities then adiabatically change from initial values to final values, $p_{x,i} \to p_{x,f}$. In the case of a sensor, the bias forces are external to the CA, while for the case of an actuator, the bias forces are internal to the CA.

According to the Einstein description, the average energy and average entropy of a sensor and actuator will change over this time. The change in average energy is $\bar{E} = E(p_{1,f} - p_{1,i})$. For example, if $\epsilon_f = -\epsilon_0$, the change in average energy is

$$\bar{E} = -E \tanh \epsilon_0/2 .\qquad(22)$$

In the case of a sensor, $\epsilon_0 < 0$ and the average energy increases, while in the case of an actuator, $\epsilon_0 > 0$ and the average energy decreases. In this example, the average entropy

does not change as the probabilities for each state are exchanged. Thus, the change in the Helmholtz free energy is equal to the change in the average energy. This implies that work is done on the sensor while the actuator does work on the world.

According to the Langevin description of the system, the internal state is a stochastic variable. The sensor spends most of its time in the low-energy state and the actuator in a high-energy state. As the bias forces act, this will change, and the most likely state will switch. The time for the system to switch is a random variable: some change their state early in the control pulse and some change their state late in the control pulse. Some may not change at all. This implies that the work done by/on the external world, w, is in fact a random variable according to the Langevin description.

This scenario is typical of problems addressed in the field of stochastic thermodynamics [20]. As bias forces change, and control pulses act over some time, the switching probabilities change. Some of the key results in stochastic thermodynamics relate the probability distribution for the work done to the changes between the initial and final stationary occupation probabilities. The surprising fact is that these relations can be independent of how the bias forces vary in time. Some examples for finite-state Markov models are given in [21].

As an example, suppose that at the start of a control pulse, the sensor is in the state of zero energy and the actuator is in a state of high energy $E > 0$—the most likely configuration. For an actuator, the probability for the actuator to change its state and be found in the low-energy state at the end of the control pulse is $p_{0,f}$, and the change in the internal energy is $-E$. The probability that the device remains in its high-energy state is the error probability $p_e = p_{1,f}$. In that case, the change in the internal energy is 0. So, there are two possible values for the change in the energy of the system, 0 and $-E$, with probability distributions, $Pr(0) = p_e$ and $Pr(-E) = 1 - p_e$. The work w done by the system on the outside world is one of these two values, $0, -E$, fluctuating between the two values from trial to trial. Similar statements can be made for the sensor. The error probability is now $p_e = p_{0,1}$, and the the work done on the system is a random variable $w \in 0, E$.

The Jarzynski equality [18] is a relation between the ensemble averaged values of w over many trials and the change in free energy corresponding to the two distributions p_x, p'_x, before and after the control, for systems in contact with a heat bath. The equality thus relates the Langevin (stochastic) description to the Einstein (thermodynamic) description. It is given by

$$\mathcal{E}[e^{-\beta w}] = e^{-\beta \Delta F},\tag{23}$$

where $\beta^{-1} = k_B T$ and $\Delta F = \Delta E - \beta \Delta S$, E is the average energy, and S the average entropy (in natural units) for each of the distributions p_x, p'_x. In our presentation, there is no requirement for the sensors and actuators to be thermal systems: they are maintained in arbitrary nonequilibrium steady states. Nonetheless, a Jarzynski-type equality holds (see [20]).

Even given this analysis of sensors and actuators, these components alone are not sufficient to define a causal agent. There needs additionally to be an interaction between the states of sensors and actuators and an internal learning machine. This is a physical, irreversible process coupling the fluctuating energy states of the sensor/actuator, on the basis of time spent in the high/low energy state, to physical states of the learning machine. Let us now describe this process in more detail.

4. Learning in a Causal Agent

In our model, the only data a CA has access to is the content of its sensor and actuator records. In order to learn causal relations, we grant the machine some additional systems that can implement learning based on this data. In this section, we describe a model for this process.

Before we do so, however, let us first consider a simple but thermodynamically expensive way in which a CA could learn, employing a device that uses "memory plus look-up" [10]. In this model, the CA simply keeps its entire record of sensor and actuator

data. In order to produce a particular sensation, the CA scans the data until it finds the appropriate subset of actions to produce a sensation with high probability. This requires storing an immense amount of information and scanning it quickly. The average number of bits to store that data is given by the Shannon information of the record. If the functional relation to be learned is simple enough (a few-bit Boolean function, for example), this could be an effective strategy. In general, however, the function may require many real-valued inputs, $f(x_1, x_2, \ldots x_n)$. In that case, a very large number of trials will need to be stored: the number grows exponentially with n, which is known as the "curse of dimensionality" in machine learning. The point of learning is to compress this into a much smaller set of functional relations $f_w(x)$, where the long-time values of the weights label the functions that are learned. We now describe how this can be done in an autonomous machine.

We need to make a distinction between machine learning algorithms and machines that learn. The latter are computer programs that change the parameterisation of a set of functions of an extremely large number of variables according to trial and error training. While the function itself is unknown, a sampled set of known values of the function $y_k = f(\vec{x}_k)$ is known. The inputs, $\vec{x}_k \in \mathbb{R}^n$, are the training data, and y_k is called the true label (for example, in binary classification, $y_k \in \{0, 1\}$). In the first trial, the algorithm computes a different function, parameterised by a set of real numbers $\vec{w} \in \mathbb{R}^M$, $\hat{y}_k = f_{\vec{w}_k}(\vec{x}_k)$ on the training data. The value $\hat{y}_k$ is compared with the true value y_k, and if they are not the same, the parameters $\vec{w}$ are changed according to some fixed algorithm (for example, back propagation) that ensures that the values converge probabilistically to the true labels. In neural networks, the functions are generated by nesting a large number of elementary functions (for example, sigmoidal functions). These are called activation functions. In Figure 7, we give an example of the steps in the algorithm.
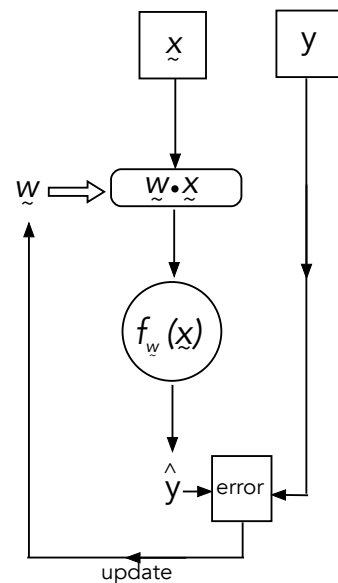

Figure 7. A very schematic representation of a machine learning algorithm. An element of the training data $\vec{x}$ is first summed over a set of weight parameters using a scalar product. The result is then fed to a binary-valued nonlinear function to give the trial result $\hat{y}$. This is compared with the correct label y, and the values of the weight parameters are adjusted accordingly by an update algorithm. The algorithm then repeats on the next element of training data.

A learning machine is not an algorithm but an open irreversible dynamical system. We assume for now that it is described by a discrete dynamical map. The input is a set of signals $\vec{x}$ and y, which we call the 'training signal' and the 'label signal', respectively. The signals are physical quantities, such as a voltage or a current. In each time step, the set of input training signals, $\vec{x}$, is fed into a nonlinear dissipative device that is biased using a set of physical parameters, the weights $\vec{w}$ (for instance, the bias voltages in an analogue circuit). The output signal, here taken to be a single binary-valued physical quantity $\hat{y}$, is

then compared with the label signal, and an error signal is used to feed back and control the weight settings according to a specific control scheme, which we can call 'cooling', before the process repeats in the next time step.

Initially, the weights are randomly distributed but, as the machine evolves in time, the distribution of weights converges to a narrow distribution on specific values that represent the unknown function $f_{\vec{w}}(\vec{x})$ inherent in the correlation between training signal and label signal.

We wish to make clear here that the essential difference between a machine learning (ML) algorithm and a learning machine (LM) is that an ML algorithm must be programmed into a suitable computational device by choosing a particular cost function, whereas an LM leverages the dynamics of its systems to learn—nature does all the work, and the cost function is a physical thermodynamic cost. In particular, (i) the ML algorithm processes numbers, while the LM processes physical input signals; (ii) the ML algorithm acts on numbers through a conventional digital computational process running on a universal computer, while the LM is a physical machine with feedback control of irreversible dynamics; and (iii) the role of data and labels in an ML algorithm is played by the values of physical signals input to the LM, and model parameters (such as weights) are played by physical parameters that control the operation of the LM's internal dynamics.

Our concept of a learning machine has more in common with analogue computers, such as the differential analyser from the middle of the last century, than it does with the standard von Neumann paradigm for numerical computation. However, unlike a differential analyser, the friction and noise that accompany it are essential for the operation of a learning machine. Considering learning machines along these lines as a model of learning offers us a richer, and more biologically relevant, route to the design of machines that learn. It also enables us to ground causal claims as relations between the physical states of learning machines. We now describe how a machine can learn through a natural process to efficiently exploit thermodynamic resources.

5. Thermodynamic Constraints on Learning Machines

Our goal in this section is to explicitly establish the link between learning in a machine and thermodynamics. To motivate the discussion, we begin with a general argument that suggests why learning machines necessarily dissipate heat and generate entropy.

Suppose we desire a learning machine that will distinguish images of sheep from images of goats (Figure 8). In the training phase, a well-labelled image of a sheep or a goat is selected with equal probability and input to the machine. The output has two channels corresponding to the true labels of the input images. At the beginning, the image comes out in either channel with equal probability: roughly 50% of the outputs match the true label and come out in the correct channel, but about the same percentage of outputs do not. These are errors.

Whenever an error occurs, feedback conditionally changes the internal biases in the learning machine to try to decrease the error probability $Pr(error) = \epsilon < 1$. Eventually, the machine gets the true labels right, and images almost always come out in the right channel. But, it can never be perfect. To see this, let us look at the entropy of the records at input and output.

The entropy at the input is clearly $\ln 2$ in natural units. Initially, the output has the same entropy but, at the end of training, the entropy of the output records is reduced to the order of $\epsilon << \ln 2$. This must be paid for by an overall increase in the entropy of the machine's environment through heat dissipation that arises every time work is done in the feedback steps. How does this arise in a physical learning machine?

At the most elementary level, neural network algorithms make use of a threshold, or *activation function*. In learning machines, the activation function becomes an *activation switch*: a dissipative, stochastic system with a nonlinear relationship between input signals and output signals (Figure 9). The operation of the switch depends on setting bias parameters, similar to sensors and actuators. The biases represent physical states of the switch. The

switch's bias parameters, which we refer to as weights, are changed by feedback in the process of learning.
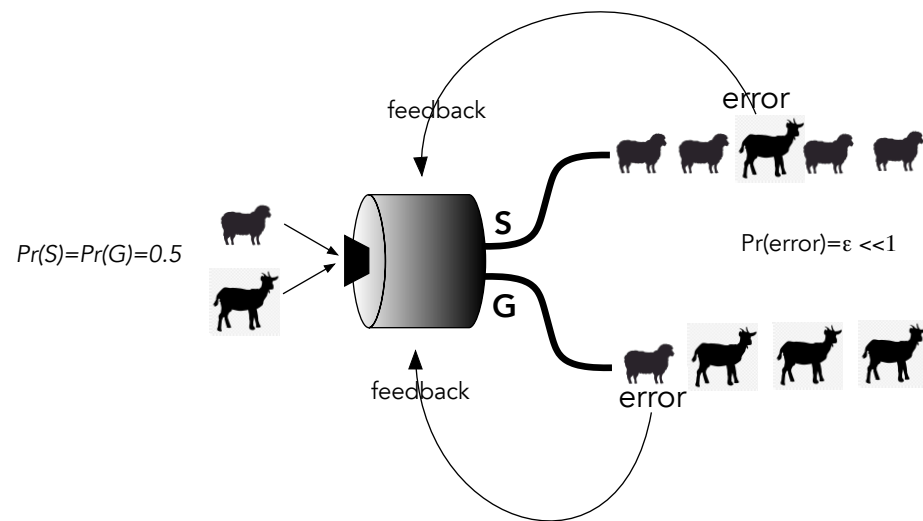

Figure 8. A learning machine is trained by feedback to classify images of goats and sheep. A correctly labelled prediction means the input image, chosen at random, comes out in the correct output. At the end of a successful training run, the probability of error, that is, the probability that an image comes out in the wrong channel, is very small $Pr(error) = \epsilon << 1$. This is achieved by actively feeding back to the machine parameters every time a mistake is made.
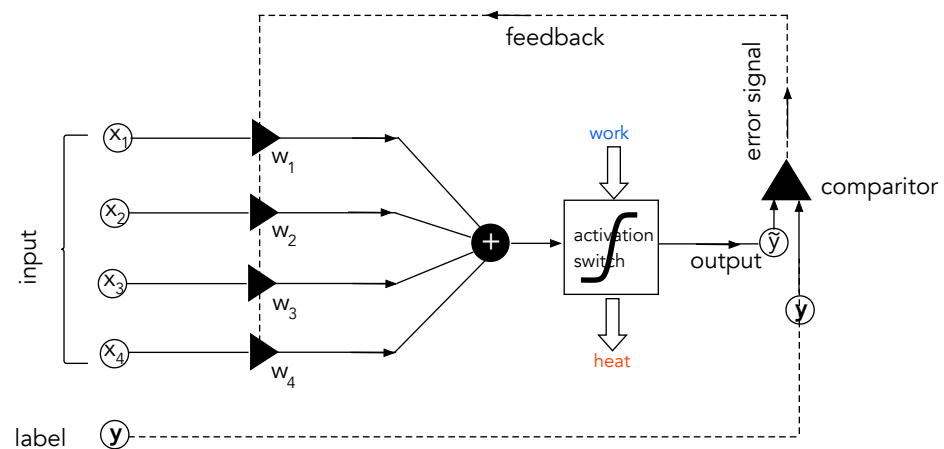

Figure 9. Schematic representation of an elementary learning machine, the perceptron, in training mode. A set of training data x_n (real numbers) together with the correct label for that data **y** (binary number) are the input. The data x_n are first multiplied by a weight factor w_n (real numbers) before being summed to give $w.x = x_1w_1 + x_2w_2 + \ldots w_nx_n$ and passed to a nonlinear device, the 'activation switch', that produces an output signal $y = f(w.x)$, where f is a binary-valued nonlinear function. This binary number is compared with the true signal label **y**. If they are the same, nothing is done, but if they are different, an error signal is sent to change the set of weights and repeat. Unavoidable physical noise means that the output is always subject to error. The cycle continues until this error probability is as small as possible. Outside of training mode, the dashed control lines are removed. The device implementing the activation function is a work-driven and dissipative nonlinear system. It must relax quickly to a steady-state output (a fixed point or a limit cycle) when the input is changed.

An example of an activation switch is shown in Figure 10. The potential function is $V(x) = x^4 - ax^2 + bx$. Including frictional damping and accompanying noise, the

displacement $x(t)$ and momentum $y(t)$ obey the Ito stochastic differential equations ([22] p. 197),

$$dx = ydt \tag{24}$$
$$dy = -(4x - 2ax + b)dt - \gamma ydt + \sigma dW, \tag{25}$$

where γ is a the frictional damping rate, $\sigma = \sqrt{2\gamma k_B T}$ is the diffusion rate for momentum at temperature T, and dW is the Wiener increment. These equations constitute the Langevin description. The Einstein description is given in terms of the stationary probability density. For a one-dimensional system of this kind, the stationary distribution is given by [22]

$$P_s(x) = \exp\left[-2\gamma \frac{V(x)}{\sigma^2}\right]. \tag{26}$$

Clearly, the steady state is peaked where the potential is at a minimum, with broadening determined by the temperature.
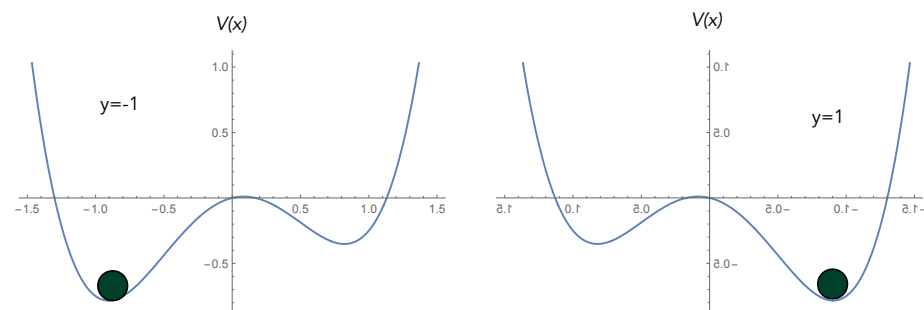

Figure 10. An example of an activation switch based on a particle moving with high friction in a double-well potential. The output variable y is the sign of the displacement $x(t)$ of the particle. Application of a constant bias force changes the shape of the potential so that the lowest energy state changes from negative to positive average displacement. This requires work to be done on the particle that is dissipated as heat.

In a deep learning algorithm, weights parameterise functions and are changed according to the back propagation algorithm. In the case of a physical neural network, this becomes a physical feedback of signals from the output to the bias conditions. When we change the bias condition of a switch, we change the probability for the output to switch from one state to another. Changing the bias condition does work on the switch, and that work is dissipated as heat. As for actuators and sensors, the work done and the heat dissipated in each trial is a random variable. Averages over these random variables are constrained by the fluctuation theorems of stochastic thermodynamics [20].

Clearly, such a system has strongly dissipative nonlinear dynamics and, by the fluctuation–dissipation theorem, must necessarily be intrinsically noisy. We must thus distinguish the ensemble average behaviour of the output—as per the Einstein description—from the single-trial output—as per the Langevin description. This important distinction is explained in Figures 10–12 with reference to a biased double well. In a single trial, the output of an activation switch is a random variable. This means that in some cases, the output will not switch when it should, corresponding to an error (Figure 12). The objective of learning is to minimise this error by changing the biases/weights of the activation switch.
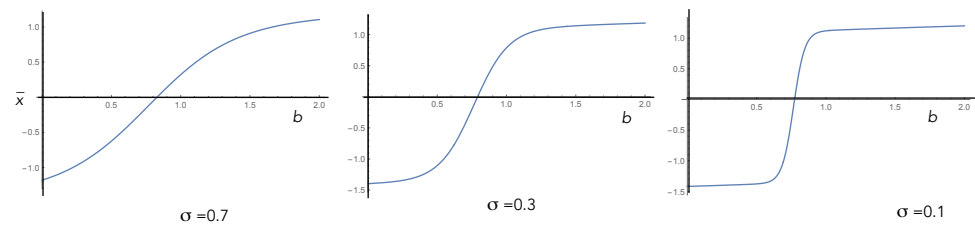
Figure 11. Einstein description: The steady-state ensemble average of the mean displacement of the particle in the double well versus the bias with varying temperatures. On the left, the temperature is high and the activation-like nature of the relationship between output and input is unclear. As the temperature is decreased, a more definite switch is seen as a function of the bias of the potential. However, the first-passage time increases as well.
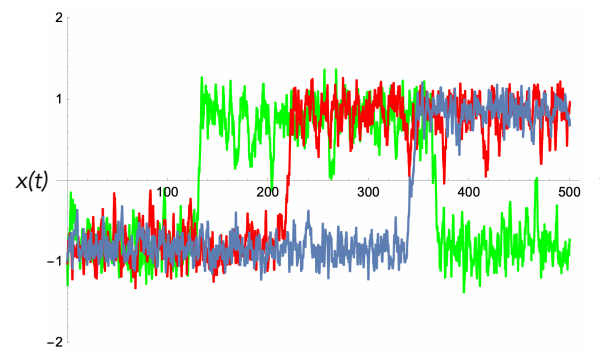

Figure 12. Langevin description: Three sample trajectories of the displacement of the particle in the double-well activation switch as a function of time as the temperature is decreased. The bias force was varied linearly in time, $b = rt$ from 0 to a constant maximum value. When the temperature is large (green, $\sigma = 1.5$), the output switches quickly but the stochastic fluctuations are large. In that case, the device switches back at high bias, indicating an error. When the temperature is small (blue, $\sigma = 1.0$), the device switches late, and the switch is slower. The time to switch for the first time is called the first-passage time. This increases as the temperature decreases.

We contend that a further key feature that distinguishes an ML algorithm from a LM is the essential relation in an LM between the minimisation of error and the minimisation of the thermodynamic cost of learning. Moreover, this relation connects an information-theoretical concept—error—to a thermodynamic quantity—heat dissipated. We set out a justification of this claim across the remainder of this section, where, following Goldt and Seifert [9], we base our discussion on the perceptron learning machine. We begin by describing a device that can take any number of binary input signals but produces a single binary signal at output. Our model is based on the two-state birth–death master equation system discussed in Section 3 and is thus intrinsically stochastic.

Consider a single system with one output signal labelled by $n \in \{-1, 1\}$. This is a binary signal. Suppose that there are K binary input signals described by a data vector, $\vec{x} = (x_1, x_2, \ldots, x_k)$, where $x_k \in \{-1, 1\}$. We set the LM the objective to learn a binary-valued function of the K input signals, $f(\vec{x}) = n_T$, where n_T is the true signal label. The training data thus consist of the inputs $(\vec{x}, n_T)$. As we discuss in more detail below, in the context of a CA, the inputs $\vec{x}$ to the learning machine are supplied as internal actuator records in the CA, and the true signal labels, n_T, are supplied by the world via the CA sensors.

There are 4^K such functions [23], and there are many supervised machine learning algorithms that can achieve this, for example, Valiant's PAC algorithm [24]. However, our interest is in designing a single machine that is able to learn any such function depending on the choice of n_T. To keep it simple, we discuss the case of $K = 1$. There are four such functions: two constant functions, the identity function, and the NOT function. Let us

consider the NOT function. The example discussed in Figure 2 can easily be framed in terms of learning a binary function of a single binary variable.

The two output states are described by the Markov process

$$\frac{dp_1}{dt} = \mu(1 - p_1) - \nu p_1 \,, \tag{27}$$

and $p_1 + p_{-1} = 1$. The transition rates, μ, ν, are nonlinear functions of a weighted sum $\vec{w}.\vec{x}$ of the components of the training vector, $\vec{x}$, and the weight vectors $\vec{w} = (w_1, w_2, \ldots w_k) \in \mathbb{R}^K$. The training of the device is performed by changing the transition rates between states by changing the weights $\vec{w}$ at each trial. If the two states represent, say, a coarse graining of an underlying double-well potential, the transition probabilities reflect thermal activation over a barrier depending on the bias forces applied [25]. We assume that the dynamical properties of the switch are such that whenever the weights change the transition probabilities, it rapidly relaxes to the new steady-state probabilities given by

$$\frac{p_1}{p_{-1}} = \frac{\mu}{\nu} \,. \tag{28}$$

We choose $p_1(\vec{w})$ to model a particular kind of activation switch described by a function called a sigmoidal perceptron [10],

$$p_1(\vec{w}) = \frac{1}{1 + e^{-\beta E_0 \vec{w}.\vec{x}}} \,, \tag{29}$$

where β is a constant fixed by experimental design. In a thermally activated device, it would be given by the temperature as $\beta^{-1} = k_B T$. In a quantum tunnelling device, it is some function of tunnelling rates. Note that $\vec{w}.\vec{x} \in \mathbb{R}$ and can be positive or negative.

At this point, we pause to make note of an important difference between a physical sigmoidal perceptron and a sigmoidal function used in machine learning algorithms. In the latter, the output is a real number, while in the physical device, the output is always a binary number, ± 1. It is the probability distribution over these values that is sigmoidal. However, we can make a direct connection by evaluating the average output $\bar{n}(\vec{w}) = p_1(\vec{w}) - p_{-1}(\vec{w})$. Normalisation then implies that $\bar{n}(\vec{w}) = 2p_1(\vec{w}) - 1$, or $p_1(\vec{w}) = (1 + \bar{n}(\vec{w}))/2$. Thus, we can sample $p_1(\vec{w})$ by running many trials with the same value of $\vec{w}.\vec{x}$ and recording the average output $\bar{n}(\vec{w})$. This set of trials is called an epoch. It is a time average over many identical trials. Another way to think of the averaging process is to imagine that we replace the single physical perceptron with a large number of identical perceptrons and only look at the average of n over all of them. Whichever way we look at it, we assume that the feedback control is based on computing the average output signal over many identical trials.

Before we can proceed with a description of the feedback process used to change the weights to implement optimal learning, we consider the thermodynamics of the perceptron activation switch. This enables us to define a thermodynamic cost function that controls the rate of learning.

When the weights change from one epoch to the next, what is the average work done on/by the device and what is the average heat dissipated? Let $\vec{w}_j, \vec{x}_j$ denote the weights and input signals at the j-epoch. We assume that the energies of each state of the two-state device are $e_n = E_0 n$. This is equivalent to the simple model in Figure 2 with a shift in the baseline energy. The change in the average internal energy between two successive epochs is [8]

$$\Delta \bar{E} = 2E_0 \Delta \bar{p}_1(\vec{w}) \,, \tag{30}$$

where $\Delta \bar{p}_1(\vec{w}) = \bar{p}_1(\vec{w} + \Delta \vec{w}) - p_1(\vec{w})$, and $\Delta \vec{w}$ is the change in the weight vector between two successive epochs. We now impose the thermodynamic constraint that the objective of feedback is to change the weights in such a way as to minimise the average energy change per epoch.

As it stands, this does not easily compare with the usual way of implementing learning, in which the focus is on minimising a cost function (for example, the average error between the output $\bar{n}(\vec{w})$ and the true label n_T). A simple way to relate the two approaches is presented in [8].

We define the error per trial as

$$\epsilon = \frac{1}{4}(n_T - n)^2 = \frac{1}{2}(1 - n_T n) , \tag{31}$$

where $n = \pm 1$ is the value of the random variable at the output of that trial. Averaging this over an epoch gives

$$\bar{\epsilon}(\vec{w}) = \frac{1}{2}(1 - n_T \bar{n}(\vec{w})) . \tag{32}$$

The change in the average error due to a variation in the weights is given by

$$\Delta \bar{\epsilon} = \Delta \vec{w} \cdot \vec{\nabla}_w \, \bar{\epsilon} . \tag{33}$$

We need this to decrease as much as possible per trial, so we set

$$\Delta \vec{w} = -\eta \vec{\nabla}_{\vec{w}} \, \bar{\epsilon} , \tag{34}$$

where η is a positive scaling constant. Thus,

$$\Delta \bar{\epsilon} = -\eta |\vec{\nabla}_{\vec{w}} \, \bar{\epsilon}|^2 . \tag{35}$$

Hence, the feedback rule is

$$\Delta \vec{w} = \eta n_T \vec{\nabla}_w p_1(\vec{w}) . \tag{36}$$

With the choice of Equation (29), we find that

$$\vec{\nabla}_w \, p_1 = \beta p_1 (1 - p_1) x \tag{37}$$

where x is the training data, and

$$\Delta \vec{w} = \eta \frac{n_T \beta (1 - \bar{n}^2)}{4} x . \tag{38}$$

The change in the weights depends on both the training data and the corresponding true labels. Note that this goes to zero as $\bar{n} \to \pm 1$, corresponding to learning the required function. With this choice, we find

$$\Delta \bar{\epsilon} = -\eta \beta^2 (1 - \bar{n}^2)^2 / 16 . \tag{39}$$

This is always negative, and has a maximum when $\bar{n} = 0, (p_1 = 1/2)$. At the start of training, this is typically the case, and the change in the average error is large. As training proceeds, it decreases. Using Equation (30), we find

$$\Delta \bar{E} = -2\eta n_T E_0 \Delta \bar{\epsilon} . \tag{40}$$

We see that the average change in energy per trial will be a minimum when the average change in error per trial is a minimum. Thus, the cost function is equivalent to minimising a thermodynamic cost. This relates an information-theoretical quantity, average error in learning, to a physical thermodynamic quantity, heat dissipation. This is similar to Landauer's principle [26] that relates the energy cost to an information-theoretical quantity: decrease in entropy by erasure of information. A different thermodynamic constraint on learning is given by Goldt and Seifert [9]. They show that the mutual information between the true labels and the learned labels is constrained by heat dissipated and entropy production.

The discrete dynamical process induced on the space of weights by feedback is stochastic, because in each epoch, the output probability distribution needs to be sampled (Equation (38)). However, this is a highly nonlinear stochastic process. When learning is complete, the weights reach a new stationary distribution on weight space that has much less entropy than the distribution used at the start of learning.

If we initialise the weights by a random vector, then the average output over an epoch, $\bar{n}(\vec{w})$, is close to zero, and the average error is large. The steady-state distribution of the perceptron is uniform, and its entropy is a maximum. The learning proceeds by feeding back onto the $\vec{w}$ until $\bar{n}(\vec{w})^2 \approx 1$. We can also relate the change in average entropy of the perceptron per epoch to the change in average error per epoch [8]. The result is that the decrease in average entropy per step is also proportional to the change in average error per step.

Stepping back, it is clear that a learning machine is a dissipative, driven, nonequilibrium system with highly nonlinear dynamics and many variables. Betti and Gori [27] have made a similar claim. Like all such systems, learning machines have nonequilibrium steady states and corresponding basins of attraction in weight space [20]. Initial high-entropy distributions over weight space are 'cooled' into the basins of attraction that characterise the learned function. The learning is driven by an imperative to optimise the use of thermodynamic resources. Where does that imperative originate? It might be the case that in an open, far-from-equilibrium system of sufficient complexity, the evolution of learning machines is a direct consequence of thermodynamics and evolution [28–30].

The function that the machine learns is labelled by the weights that it converges to in training. These are not deterministic but, after training, they are sharply peaked around particular values that label the function learned, $f_{\vec{w}}$. These labels are physical variables that define the bias parameters of the learning machine. This is how learning machines come to represent functions in terms of physical variables.

6. Learning and Causal Relations

We now return to the discussion of a causal agent that incorporates sensors, actuators, and learning machines. The model, which we refer to here as the emulator model (EM), is summarised in Figure 13. This model is inspired by models now common in neuroscience that treat the brain primarily as a prediction machine [31–33]. The core idea originated in the corollary discharge model of Sperry [34], but it is a general model for a learning machine.

The goal of the learning machine is to take actuator records and predict sensor records. The predictions are then compared with actual sensor records and feedback is used to modify the settings of the learning machine dynamics so as to minimise the error probability, that is to say, the probability that predictions do not match actual sensor records. As we have already seen, this abstract informational goal can be made equivalent to a physical goal: minimise thermodynamic cost as measured by power dissipated by the CA as it interacts with its environment.

There are no doubt many ways to engineer a learning machine along these general lines. We use the physical neural network (PNN) model described in the previous section. In terms of that model, the actual sensor records coming from the external world play the role of true labels in supervised learning. We can think of nature as a function oracle to which actuators pass arguments, to use the PAC language of Valiant [24]. The input to the PNN emulator is the actuator records that record sequences of interventions taken by the machine on the external world. The outputs of the PNN are predicted sensor records, and these are compared with the true labels provided by the actual sensors acted on by the world. The previous discussion shows that minimising error predictions is equivalent to minimising the thermodynamic cost of the CA interacting with the world.
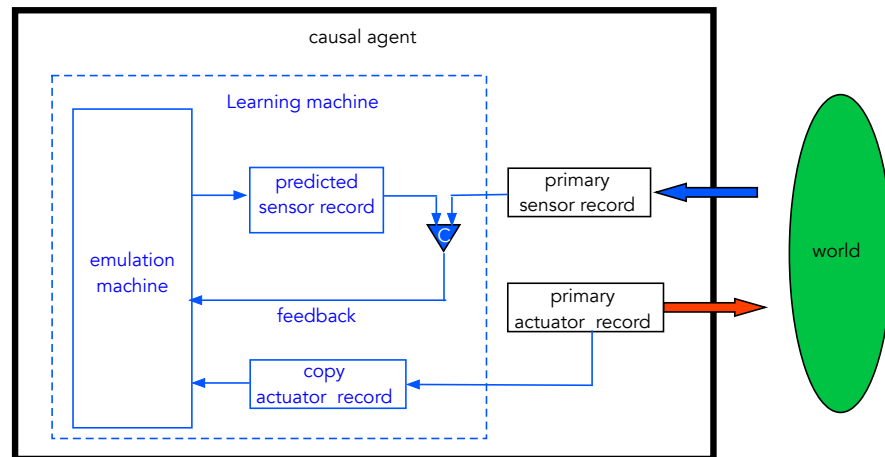
Figure 13. A schematic of a learning machine based on a physical emulator with feedback. A single round of learning proceeds as follows. The primary actuator record registers what action is taken on the external world, while the primary sensor record indicates what sensation is received from the external world. The primary actuator record is copied and sent to an emulation engine to produce a predicted sensor record. This is compared with the primary sensor record by a comparator (C), and the result is fed back to the emulation engine, which then updates the 'weights'. The process repeats until the prediction error is a minimum. This model for learning is similar to the concept of predictive processing recently developed in the philosophy of neuroscience (see [35] and references therein).

In this model, we take the correlations between sensor and actuator records learned by the CA to be causal relations. Thus, a causal relation is a learned function, $f_w(a)$, mapping actions to predicted sensations and parameterised by the physical quantities that represent the weights, w, inside the learning machine. As an example, let us return to the model in Figure 2. It is clear that the function to be learned is a one-bit Boolean function. Two of these are constant functions that always give either 1 or 0. These give the same result regardless of the input. In terms of the physical model, this would correspond to a bump that is high enough to reflect every particle or low enough to transmit every particle. The other two are the identity function and its complement, the NOT function. In both of these, the output depends on the input. Up to a relabelling of the states of the source, these are the same physical process. In particular, they are the only ones that can be the basis of an effective strategy (to use Cartwright's term [6]) as far as the CA is concerned. A one-bit Boolean function can be learned by a single perceptron [8]. The probability function is given by a single weight w and single bias b,

$$p_1(w, b) = \frac{1}{1 + e^{-\beta E_0(w.x+b)}} \ . \tag{41}$$

The NOT function corresponds to $w < 0, b = 0$, and the identity function to $w > 0, b = 0$. The two constant functions correspond to $w = 0, b \neq 0$. In Figure 14, we schematically indicate the four possible distributions for the weights (w) and bias (b) after learning has reached minimum error. The initial distribution of the w, b is very broad. After learning, they are localised in the space as four distinct distributions, corresponding to the functions learned.
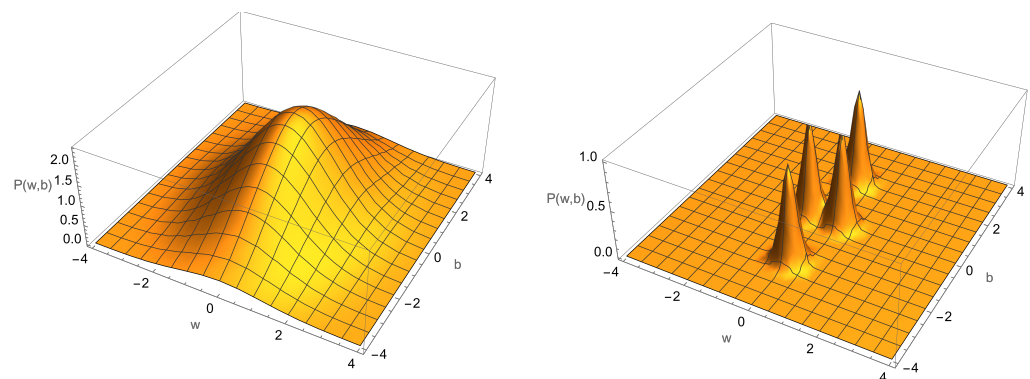
Figure 14. Learning a one-bit gate with a single physical perceptron. Left: the initial distribution of the weight and bias variables. Right: the four possible distributions after training corresponding to the four different labels for the four possible one-bit functions. Which one the machine learns depends on the training data labels. For example, if the training data have a fixed label $n_T = 1$ for all possible values of the input, x, the perceptron will stochastically evolve to $w = 0, b > 0$ with high probability.

It is important to keep in mind that the physical machine 'knows' no more about the causal relations than the settings of its weights and bias. These label the function learned in a probabilistic sense: there is not a unique label, only a region of likely labels for a given function. Can one build a learning machine that learns which function the machine has learned?

To answer this question, we need to introduce the idea of a nested hierarchy of learning machines, wherein each level learns some feature of the functions learned at the lower level by learning some feature of the weights learned at the lower level (for example, which of the four regions in Figure 14 have been reached).

Suppose we have a learning machine with two actuators and one sensor. This could learn an arbitrary two-bit function with a single binary output. There are 16 such functions. These functions could be classified by a property, and that property can be learned. For example, suppose the classification is simply the answer to the question "is the learned function balanced or constant?" A balanced function is one that takes the same value on half of the input domain and zero on the other half. An example of this is XOR.

The 16 possible functions can be labelled by a binary string of length 4. There are 256 binary functions with 4 inputs, so the higher learning machine will need more perceptrons or more layers, or both. It is know that a binary-valued function of n binary variables can be learned with a single hidden layer of 2^n perceptrons.

Thermodynamic constraints will continue to apply when learning machines are nested in this way. At each level, optimal learning will correspond to optimal thermodynamic efficiency given a particular cost function. What sets the cost function? To answer this question, we need to ask: Why learn anything at all?

One answer that is relevant for an interventionist characterisation of causation [2,3] is the following: a machine learns causal correlations between sensations and actions so as to act effectively in the world. An effective learning strategy is one that enables effective control of some system external to the CA. This insight shifts the question somewhat: What is an effective control strategy? This question is addressed in the field of optimal stochastic control [17], which explicitly addresses the question of what an optimal control *policy* is, that is to say, how best to act.

The simplest example of an optimal policy would seek to minimise the energy used to change the state of a physical system [17]. It seems likely that, in an evolutionary setting, such a cost function will be selected spontaneously. An old result of cybernetics is the *good controller theorem* [36], according to which the best stochastic controller is a replica of the system to be controlled. In our model of emulation learning, the learning machine does indeed seek to replicate the dynamics of the external world in a simpler device, so

far as it can 'see' it through actions and sensations. As we discussed in Section 5, the free energy cost per learning step is minimised as the error change per step is minimised. This enables a physical grounding of the free energy principle of Friston [13]. Learning machines are thermodynamically optimal controllers of the external world, and the cost function is ultimately thermodynamic in origin. It is not imposed from the outside. Effective strategies are grounded in thermodynamics.

7. Discussion and Conclusions

Let us finally return to the question of causal asymmetry in the context of our approach. A response to this question crucially depends upon whether causation is an objective feature of the world, or whether it is grounded in the perspective of certain special kinds of physical systems, such as ourselves, that learn and act. One of our primary goals in this work has been to demonstrate how causal relations can be defined in terms of the physical states of a special kind of physical machine, a casual agent. The sole purpose of such physical states is to enable prediction of the sensations that follow an action, within some error bound. Thus, we take these physical states, which are characterised by the internal weights and biases of the learning machine of the CA, to encode a custom-built causal model of the CA's environment (where the internal weights and biases are simply the modelling parameters of the causal model).

On this view, the causal relations *just are* the internal settings of the physical states of the learning machine, which are no less physical than those in the world outside. As such, we have argued that causal relations are nothing more than learned relations between sensor and actuator records inside the learning machine of the CA. What is more, we have demonstrated the close connection between the operation of a physical learning machine, and so also the causal agent, and the laws of thermodynamics and energy dissipation. We argue that efficient learning, in the sense of minimising error, is equivalent to thermodynamic efficiency, in the sense of minimising power dissipation. Along with the thermodynamics of sensors and actuators, this renders any such causal agent inherently 'directed', and we take this fact to underpin the asymmetry of causation.

We treat the consequences of our approach in greater philosophical depth in [37]. Briefly, however, let us outline the way in which we take the above considerations to exemplify a perspectival approach to the interventionist account of causation [4,5].

Our account is explicitly interventionist due to the necessity of actuators in a CA. These are physical subsystems in a CA that do work on the external world. Sensors alone are not enough to build a CA. Certainly, one could easily build a CA with an algorithm to find patterns in its sensor records (a Bayesian network, say), especially with time series data from multiple sensor types. It is easy to see that correlations could be found between records from different types of sensors. But would any extant correlation in the records indicate a causal relation? Such a claim would be open to Hume's objection: patterns or 'regularities' do not ground causal claims. It is only through intervening on the world that agents can confidently discriminate cause from effect.

Moreover, our model of a CA demonstrates how the asymmetry between exogenous and endogenous variables in the interventionist account of causation has its origin in the thermodynamic asymmetry of the CA itself. Since the causal model that the CA learns is underpinned by its own mechanisms of intervention, detection, and learning, then the thermodynamic directedness of its actuators, sensors, and learning machine dictate that its causal model must also be 'directed' in the same way. That is, the CA must be modelling its environment with itself embedded at the heart of the model, such that effects in the world must be thermodynamically downstream from its actuators and thermodynamically upstream from its sensors, and any modelled causal relations must be directed from variables manipulable by the actuator (exogenous) and detectable by the sensor (endogenous). Thus, in a sense, this directedness constrains the agent to be able to act only towards the 'thermodynamic future' and to gain knowledge only of the 'thermodynamic past'.

We take this to be a clear demonstration of what Price [5] calls a causal perspective: agents who are situated and embedded 'in time' in this way are constrained with respect to the actions they can perform and the functional relations they can exploit. Moreover, it is clear that this perspective is inescapable by the CA—it is a function of its own internal constitution. What is more, a CA is also constrained to model its environment exclusively in terms of the set of dynamical variables that are manipulable-and-detectable by the actuator and sensor system. As such, the causal model learned by the CA will only contain functional relations that are exploitable according to the physical constitution of its sensors and actuators. This is then a further sense in which the CA is situated and embedded in an inescapable environment, again as a function of the physical constitution of its own network of actuators, sensors, and learning machines. There is an obvious connection between this latter situatedness and Kirchoff's [35] notion of an agent's own "Markov blanket" as an environment of sorts with which its internal states must contend. Once again, this situatedness defines a causal perspective.

As a final speculative suggestion, we wish to note that CAs that inescapably share a functionally similar constitution of actuators, sensors, and learning machines will share a 'perspective' with respect to the causal models they learn. Thus, a shared physical constitution ensures the stability of a casual perspective across a class of CAs, and this then defines an equivalence class of agents that shares a perspective. It is in this way, then, that learned causal relations are shared by all CAs with the same hardware embedded in the same environment. We think the ramifications of this sort of multiagent learning are ripe for further investigation.

Our approach here shows that we can define causal relations independently of human agency, giving a perspectival interventionist account that avoids the charge of anthropocentrism. Instead, we define causal relations as learned relations between internal physical states of a special kind of open system, one with special physical subsystems—actuators, sensors, and learning machines—operating in an environment with access to a large amount of free energy. Even though such open systems need not necessarily be human, it is their ability to model the causal relations in their environment, and exploit such relations as effective strategies, that makes them causal agents.

References

1. Russell, B. On the Notion of Cause. *Proc. Aristot. Soc.* **1913**, *13*, 1–26. [CrossRef]
2. Pearl, J. *Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2000.
3. Woodward, J. *Making Things Happen: A Theory of Causal Explanation*; Oxford University Press: New York, NY, USA, 2003.
4. Ismael, J. How do causes depend on us? The many faces of perspectivalism. *Synthese* **2016**, *193*, 245–267. [CrossRef]
5. Price, H. Causal perspectivalism. In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*; Price, H., Corry, R., Eds.; Oxford University Press: Oxford, UK, 2007; pp. 250–292.
6. Cartwight, N. Causal Laws and Effective Strategies. *Noûs* **1979**, *13*, 419–437. [CrossRef]
7. Ismael, J. *How Physics Makes Us Free*; Oxford University Press: Oxford, UK, 2016.
8. Milburn, G.J.; Basiri-Esfahani, S. The physics of learning machines. *Contemp. Phys.* **2022**, *63*, 34–60. [CrossRef]
9. Goldt, S.; Seifert, U. Stochastic Thermodynamics of Learning. *Phys. Rev. Letts.* **2017**, *118*, 010601. [CrossRef] [PubMed]

10. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Prentice Hall: Hoboken, NJ, USA, 2010.
11. Briegel, H.J.; De Las Cuevas, G. Projective simulation for classical and quantum learning agents. *Sci. Rep.* **2012**, *2*, 400. [CrossRef]
12. Floridi, L. *The Philosophy of Information*; Oxford University Press: Oxford, UK, 2011.
13. Friston, K.; Stephan, K.E. Free-energy and the brain. *Synthese* **2007**, *159*, 417–458. [CrossRef]
14. Bruineberg, J.; Kiverstein, J.; Rietveld, E. The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese* **2018**, *195*, 2417–2444. [CrossRef]
15. Freeman, W.J. Perception of time and causation through the kinesthesia of intentional action. *Integr. Psychol. Behav. Sci.* **2008**, *42*, 137–143. [CrossRef]
16. Llinàs, R.R. *I of the Vortex, from Neurons to Self*; MIT Press: Boston, MA, USA, 2002.
17. Stengel, R.F. *Optimal Control and Estimation*; Dover: New York, NY, USA, 1994.
18. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690. [CrossRef]
19. Carmichael, H.J. *Statistical Methods in Quantum Optics, VII*; Springer: Berlin/Heidelberg, Germany, 2008.
20. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **2012**, *75*, 126001. [CrossRef]
21. Rao, R.; Esposito, M. Detailed Fluctuation Theorems: A Unifying Perspective. *Entropy* **2018**, *20*, 635. [CrossRef] [PubMed]
22. Gardiner, C.W. *Handbook of Stochastic Processes for Physics, Chemistry and the Natural Sciences*; Springer: Berlin/Heidelberg, Germany, 1983.
23. Anthony, M. *Learning Boolean Functions*; CDAM-LSE-2005-24; Centre for Discrete and Applicable Mathematics, LSE: Coventry, UK, 2005.
24. Valiant, L. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*; Basic Books: New York, NY, USA, 2013.
25. McNamara, B.; Wiesenfeld, K. Theory of stochastic resonance. *Phys. Rev. A* **1989**, *39*,4854. [CrossRef]
26. Landauer, R. Computation and physics: Wheeler's meaning circuit? *Found. Phys.* **1986**, *16*, 551–564. [CrossRef]
27. Betti, A.; Gori, M. The principle of least cognitive action. *Theor. Comput. Sci.* **2016**, *633*, 83–99. [CrossRef]
28. Rovelli, C. Meaning = Information + Evolution. *arXiv* **2012**, arXiv:1611.02420v1.
29. Jeffery, K.; Pollack, R.; Rovelli, C. On the statistical mechanics of life: Schrödinger revisited. *Entropy* **2019**, *21*, 1211.
30. England, J.L. Dissipative adaptation in driven self-assembly. *Nat. Nanotechnol.* **2015**, *10*, 919–923. [CrossRef] [PubMed]
31. Clark, A. Whatever next? *Behav. Brain Sci.* **2013**, *36*, 181–253. [CrossRef]
32. Seth, A. *Being You: A New Science of Consciousness*; Faber & Faber: London, UK, 2022.
33. Buzsaki, G. *The Brain from Inside Out*; Oxford University Press: Oxford, UK, 2019.
34. Sperry, R.W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* **1950**, *43*, 482–489. [CrossRef]
35. Kirchhoff, M.D. Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philos. Stud.* **2018**, *175*, 751–767. [CrossRef]
36. Conant, R.C.; Ashby, W.R. Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.* **1970**, *1*, 89. [CrossRef]
37. Evans, P.W.; Shrapnel, S.; Milburn, G.J. Causal Asymmetry from the Perspective of a Causal Agent. 2021. Available online: http://philsci-archive.pitt.edu/18844/ (accessed on 24 March 2021).