frontiers | Frontiers in Neurorobotics

TYPE Hypothesis and Theory
PUBLISHED 30 September 2022
DOI 10.3389/fnbot.2022.846982

OPEN ACCESS

EDITED BY
Inês Hipólito,
Humboldt University of Berlin,
Germany

REVIEWED BY
Tetsushi Nonaka,
Kobe University, Japan
Riccardo Manzotti,
Università IULM, Italy

*CORRESPONDENCE
Vadim Weinstein
vadim.weinstein@oulu.fi
Steven M. LaValle
steven.lavalle@oulu.fi

RECEIVED 31 December 2021
ACCEPTED 28 July 2022
PUBLISHED 30 September 2022

CITATION
Weinstein V, Sakcak B and LaValle SM (2022) An enactivist-inspired mathematical model of cognition. *Front. Neurorobot.* 16:846982. doi: 10.3389/fnbot.2022.846982

An enactivist-inspired mathematical model of cognition

Vadim Weinstein*, Basak Sakcak and Steven M. LaValle*

Center for Ubiquitous Computing, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

In this paper we start from the philosophical position in cognitive science known as enactivism. We formulate five basic enactivist tenets that we have carefully identified in the relevant literature as the main underlying principles of that philosophy. We then develop a mathematical framework to talk about cognitive systems (both artificial and natural) which complies with these enactivist tenets. In particular we pay attention that our mathematical modeling does not attribute contentful symbolic representations to the agents, and that the agent's nervous system or brain, body and environment are modeled in a way that makes them an inseparable part of a greater totality. The long-term purpose for which this article sets the stage is to create a mathematical foundation for cognition which is in line with enactivism. We see two main benefits of doing so: (1) It enables enactivist ideas to be more accessible for computer scientists, AI researchers, roboticists, cognitive scientists, and psychologists, and (2) it gives the philosophers a mathematical tool which can be used to clarify their notions and help with their debates. Our main notion is that of a sensorimotor system which is a special case of a well studied notion of a transition system. We also consider related notions such as labeled transition systems and deterministic automata. We analyze a notion called *sufficiency* and show that it is a very good candidate for a foundational notion in the "mathematics of cognition from an enactivist perspective." We demonstrate its importance by proving a uniqueness theorem about the minimal sufficient refinements (which correspond in some sense to an optimal attunement of an organism to its environment) and by showing that sufficiency corresponds to known notions such as sufficient history information spaces. In the end, we tie it all back to the enactivist tenets.

KEYWORDS

enactivism, transition systems, automaton, cognitive modeling, information spaces, robotics

1. Introduction: Mathematizing enactivism

The premise of this paper is to lay down a logical framework for analyzing agency in a novel way, inspired by enactivism. Classically, mathematical and logical models of cognition are in line with the cognitivist paradigm in that they rely on the notion of symbolic representation and do not emphasize embodiment or enactment (Newell and Simon, 1972; Fodor, 2008; Gallistel and King, 2009; Rescorla, 2016).

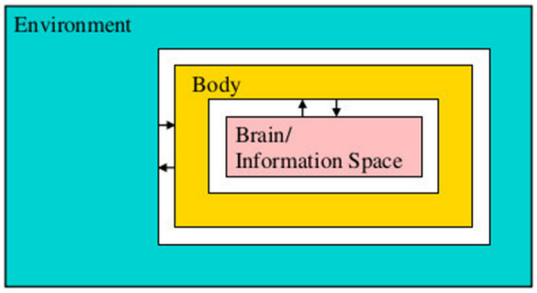

FIGURE 1
The environment, body, and nervous system (or brain) will be modeled as inseparable parts of a coupled transition system.

Cognitivism presumes that the world possesses objective structure and the contentful information of this structure is acquired and represented by the cognitive agent. This aligns well with the classical model-theoretic paradigm. In this paradigm a formal language is describing a static model (such as when sentences in the language of rings describe algebraic structures—such as rings).

In the cognitivist analogy, the agent possesses ("in its head") formulas of the language and the model is the world or the environment of the agent. If the formulas possessed by the agent hold in the model, then the agent's representation of the world is correct; otherwise, it is incorrect. Such view of cognitive agency is rejected by the enactivists either weakly or strongly depending on the branch of enactivism. For example, radical enactivism (Hutto and Myin, 2012, 2017) rejects this view strongly. Our question for this paper is: What would the mathematical logic of cognition look like, if even the radical enactivists were to accept it?

We do not take part in the cognitivist-enactivist, or the representationalist-antirepresentationalist debate (Pezzulo et al., 2011; O'Regan and Block, 2012; Gallagher, 2018; Fuchs, 2020). Rather, we take a somewhat extreme enactivist and antirepresentational view as our axiomatic starting point and as a theoretical explanatory target. Then we develop a mathematical theory that attempts to account for cognition in a way congruent with this view. Even though most forms of enactivism (even radical ones) have room for representation, it is not our main goal at the moment to bridge the gap between "basic minds" and "scaffolded minds," to use terminology of (Hutto and Myin, 2017). Thus, in this terminology, we are going to explore a mathematical (only) of *basic minds*.

The following "axioms" we take as fundamentals for our work:

(EA1) Embodiment. "From a third-person perspective the organism-environment is taken as the explanatory unit" (Gallagher, 2017). The environment, the body, and the nervous system are inseparable parts of the system which they form by coupling; see Figure 1. They cannot be meaningfully understood in isolation from each other. "Mentality is in all cases concretely constituted by, and thus literally consists of, the extensive ways in which organisms interact with their environments, where the relevant ways of interacting involve, but are not exclusively restricted to, what goes on in brains" (Embodiment Thesis Hutto and Myin, 2012).

(EA2) Groundedness. The brain does not "acquire" or "possess" contentful states, representations, or manipulate semantic information in any other way. "Mentality-constituting interactions are grounded in, shaped by, and explained by nothing more, or other, than the history of an organism's previous interactions. Nothing other than its history of active engaging structures or explains an organism's current interactive tendencies." [Developmental-Explanatory Thesis (Hutto and Myin, 2012)].

(EA3) Emergence. The crucial properties of the brain-body-environment system from the point of view of cognition emerge from the embodiment, the brain-body-environment coupling, the situatedness, and the skills of the agent. The agent's and the environment's prior structure come together to facilitate new structure which emerges through the sensorimotor engagement. "[T]he mind and world arise together in enaction, [but] their manner of arising is not arbitrary" (i.e. it is structured) (Varela et al., 1992).

(EA4) Attunement. Agents differ in their ways of attunement and adaptation to their environments, and in the skills they have. A *skill* is a potential possibility to engage *reliably* in complex sensorimotor interactions with the environment (Gallagher, 2017).

(EA5) Perception. Sensing and perceiving are not the same thing. Perception arises from skillful sensorimotor activity. To perceive is to become better attuned to the environment. O'Regan and Noë (2001) and Noë (2004) "Perception and action, sensorium and motorium, are linked together as successively emergent and mutually selecting patterns." Varela et al. (1992).

The mathematics we use to capture those ideas is a mixture of known and new concepts from theoretical robotics, (non-)deterministic automata and transition systems theory, and dynamical systems (Goranko and Otto, 2007). It will also build upon the *information spaces* framework, introduced in LaValle (2006) as a unified way to model sensing, actuation, and planning in robotics; the framework itself builds upon earlier ideas such as dynamic games with imperfect information (von Neumann and Morgenstern, 1944; Başar and Olsder, 1995), control with imperfect state information (Kumar and Varaiya, 1986; Bertsekas, 2001), knowledge states (Lozano-Pérez et al., 1984; Erdmann, 1993), perceptual equivalence classes (Donald

and Jennings, 1991; Donald, 1995), maze and graph-exploring automata (Shannon, 1952; Blum and Kozen, 1978; Fraigniaud et al., 2005), and belief spaces (Kaelbling et al., 1998; Roy and Gordon, 2003).

Although information spaces refer to "information," they are not directly related to Shannon's *information theory* (Shannon, 1948), which came later than von Neumann's use of information in the context of sequential game theory. Neither does "information" here refer to content-bearing information. One important intuition behind the information in information spaces is that more information corresponds to narrowing down the space of possibilities (for example of future sensorimotor interactions).

The main mathematical concept of this paper is a *sensorimotor system* (SM-system), which is a special case of a transition system. Sensorimotor systems can describe the body-brain system, the body-environment system as well as other parts of the brain-body-environment system. Given two SM-systems they can be *coupled* to produce another (third) SM-system. Mathematically, the coupling operation is akin to a direct product. We introduce several notions that describe the coupling of the agent and the environment from an outside perspective (not from the perspective of the agent or the environment). The main notion is that of *sufficiency*. In some sense it guarantees that the coupling is of "high fidelity." It does not compare "internal" models of the agent to "external" states of affairs. Rather it asks whether the way in which the agent engages in sensorimotor patterns is well structured. The notion of *sufficiency* compares the sensorimiotor capacity of the agent *to itself* by asking whether the past sensorimotor patterns (in a given environment) determine reliably the future sensorimotor patterns. We then introduce several related notions. The *degree of insufficiency* is a measure by which various agents can be compared in their coupling versatility (Def 4.11). *Minimal sufficient refinement* is a concept that can be used in the most vivid ways to illustrate how the sensorimotor interaction "enacts" properties of the brain-body-environment system. The notion of minimal sufficient refinement ties together mathematics of sensorimotor systems and the philosophical ideas of emergence, structural coupling and enactment of the "world we inhabit" (cf. Varela et al., 1992); see Example 4.25. We prove the uniqueness of minimal sufficient refinements (Theorem 4.19) and point out their connection to the notions of bisimulation and sufficient information mappings. *Strategic sufficiency* is a mathematically more challenging concept, but has appealing properties in the philosophical and practical sense. A sensor mapping is strategically sufficient *for* some subset of the state space G, if that sensor can (in principle) be used by the agent to reach G [1]. Again, any sensor mapping has minimal strategic refinements, but this time they are not unique. Different minimal refinements in this case can be thought of as different adaptations to the same environmental demands.

Mathematically, sufficiency is a relative concept to some known notions in theoretical computer science and robotics: that of *bisimulation* in automata and Kripke model theory (Goranko and Otto, 2007), and *sufficient information mappings* in information spaces theory (LaValle, 2006).

Minimal sufficient refinements lead to unique classifications of agent-environment states that "emerge" from the way in which the agent is coupled to the environment, not merely from the way the environment is structured on its own. Thus, the world is simultaneously objectively existing (from the global "god" perspective), but also "brought about" by the agent.

This should be enough to answer the two questions that, according to Paolo (2018), any embodied theory of cognition should be able to provide precise answers to: What is its conception of bodies? What central role do bodies play in this theory different from the roles they play in traditional computationalism?

Section 2 introduces the basics of transition and SM-systems, their coupling, and other mathematical constructs such as quotients. Section 3 illustrates the introduced notions with detailed examples. Section 4 introduces the notion of sufficiency, sufficient refinements, and minimal sufficient refinements. We will prove the uniqueness theorem for the latter and illustrate the notions in a computational setting. We will explore the importance of sufficiency and related notions for the enactivist way of looking at cognitive organization. Finally, Section 5 ties the mathematics back to the philosophical premises.

2. Transition systems and sensorimotor systems

At the most abstract level, the central concept for our mathematical theory is that of a *transition system*. This is a standard definition from automata theory (for instance Goranko and Otto, 2007):

Definition 2.1. A *transition system* is a triple (X, U, T) where X is the *state space* (mathematically it is just a set), U is the set of names for outgoing transitions (another set), and $T \subseteq X \times U \times X$ is a ternary relation.

The intuitive interpretation of (X, U, T) is that it is possible to transition from the state $x_1 \in X$ to the state $x_2 \in X$ via $u \in U$ iff $(x_1, u, x_2) \in T$. We use the notation $x_1 \xrightarrow{u} x_2$ to mean that $(x_1, u, x_2) \in T$. Our notion of transition system is often called a *labeled* transition system in the literature, because each potential transition has a name or label, $u \in U$. However, we drop the term "labeled" because in Section 2.5 we will introduce a version of transition systems in which the states are relabeled, thereby introducing a new kind of labeling. Note that when working with such transition

1 This idea of a reachable set G is the simplest way to formalize affordances.

systems as modeling agency, we are safely within the realm of the Developmental-Explanatory Thesis (EA2). The following definitions are standard (although we do not restrict X to be finite):

Definition 2.2. Let $\mathcal{X} = (X, U, T)$ and $\mathcal{X}' = (X', U', T')$ be transition systems. An *isomorphism* is a bijective function $f: X \to X'$ such that for all $x_1, x_2 \in X$ and $u \in U$ we have $(x_1, u, x_2) \in T \iff (f(x_1), u, f(x_2)) \in T'$. A *simulation* is a relation $R \subseteq X \times X'$ such that for all $(x_1, x_1') \in R$, all $u \in U$ and all $x_2 \in X$, we have that if $(x_1, u, x_2) \in T$, then there exists $x_2' \in X'$ with $(x_1', u, x_2') \in T'$ and $(x_2, x_2') \in R$. A *bisimulation* is a relation R such that both R and $R^T = \{(y, x) : (x, y) \in R\}$ are simulations.

The notation $\mathcal{X} \cong \mathcal{X}'$ means that $\mathcal{X}$, $\mathcal{X}'$ are isomorphic, and $\mathcal{X} \sim \mathcal{X}'$ means that there is a bisimulation R such that $X = \mathrm{dom}(R)$ and $X' = \mathrm{ran}(R)$. We speak of *automorphism* and *autobisimulation*, if $\mathcal{X} = \mathcal{X}'$.

We are ready to make the first observation:

Proposition 2.3. *If $\mathcal{X} \cong \mathcal{X}'$, then $\mathcal{X} \sim \mathcal{X}'$.*

Proof: Let f be an isomorphism $f: X \to X'$. Then $R = \{(x_1, x_2) \in X \times X' \mid x_2 = f(x_1)\}$ is a bisimulation. □

2.1. Transition systems as a unifying concept

There are several ways in which transition systems and their relatives appear in the literature relevant to us.

Examples 2.4. Let (X, U, T) be a transition system.

1. Let $x_0 \in X$ and $F \subseteq X$. Let $\hat{T}: X \times U \to \mathcal{P}(X)$ be defined by $\hat{T}(x, u) = \{x_2 \in X \mid x_1 \xrightarrow{u} x_2\}$. Then $(X, U, \hat{T}, x_0, F)$ is a *nondeterministic automaton*. If in addition X and U are finite, then it is a *nondeterministic finite automaton* (NFA).
2. Let $\tilde{T}: X \times X \to \mathcal{P}(U)$ be the function $\tilde{T}(x_1, x_2) = \{u \in U \mid x_1 \xrightarrow{u} x_2\}$. Then $\tilde{T}(x_1, x_2)$ is the set of all u that take x_1 to x_2. Then, $(X, \tilde{T})$ is a labeled directed graph in which the labels are subsets of U. Another way to think of it is as a labeled directed multigraph: the multiplicity of the edge from x_1 to x_2 is $n = |\tilde{T}(x_1, x_2)|$ and these n edges are labeled by the labels from the set $\tilde{T}(x_1, x_2)$.
3. If for all $x_1 \in X$ and $u \in U$ there is a unique $x_2 \in X$ with $x_1 \xrightarrow{u} x_2$, let $\tau: X \times U \to X$ be the function defined such that $\tau(x_1, u) = x_2$ iff $x_1 \xrightarrow{u} x_2$. Let $x_0 \in X$ and $F \subseteq X$. Then (X, U, τ, x_0, F) is a *deterministic automaton*, and if X and U are finite, then it is a *deterministic finite automaton* (DFA). Without F, (X, U, τ, x_0) also satisfies the definition of the *temporal filter* of LaValle (2012, 4.2.3). In this case X is the *information space* or the *I-space* (usually denoted by $\mathcal{I}$ instead of X), and U is the observation space (usually denoted by Y instead of U).

2.2. Information spaces and filters

We can reformulate the notion of a *history information space* introduced by LaValle (2006) as follows. In this context, X is an external state space that characterizes the robot's configuration, velocity, and environment, U is an action space, f is a state transition mapping that produces a next state from a current state and action, h is a sensor mapping that maps states to observations, and Y is a sensor observation space. As in LaValle (2006), for each $x \in X$, let $\Psi(x)$ be a finite set of "nature sensing actions" and for each $x \in X$ and $u \in U$ let $\Theta(x, u)$ be a finite set of "nature actions." Let $X_\Psi = \{(x, \psi) \mid \psi \in \Psi(x)\}$ and let $h: X_\Psi \to Y$ be a "sensor mapping" where Y is a set called the "observation space." Let $X_\Theta = \{(x, u, \theta) \mid \theta \in \Theta(x, u)\}$ and let $f: X_\Theta \to X$ be the "transition function." The following definition is an adaptation from LaValle (2006).

Definition 2.5. A *valid history I-state for* X, Ψ, Θ, f is a sequence $(u_0, y_0, \ldots, u_{k-1}, y_{k-1})$ of length $2k$ for which there exist $\bar{x} = (x_0, \ldots, x_{k-1})$, $\bar{\psi} = (\psi_0, \ldots, \psi_{k-1})$ and $\bar{\theta} = (\theta_0, \ldots, \theta_{k-2})$ such that for all $i < k$ we have

1. $\theta_i \in \Theta(x_i, u_i)$,
2. if $i < k - 1$, then $x_{i+1} = f(x_i, u_i, \theta_i)$,
3. $\psi_i \in \Psi(x_i)$,
4. $y_i = h(x_i, \psi_i)$.

In this case we say that $(u_0, y_0, \ldots, u_{k-1}, y_{k-1})$ is *witnessed* by $\bar{x}$, $\bar{\psi}$ and $\bar{\theta}$.

Now let $\mathcal{I}$ be the set of all valid history I-states for X, Ψ, Θ, f. For all $k \in \mathbb{N}$, all $\bar{x} \in X^{k-1}$, all $\bar{\psi} = (\psi_0, \ldots, \psi_{k-1})$ and all $\bar{\theta} = (\theta_0, \ldots, \theta_{k-2})$, let $\mathcal{I}^k(\bar{x}, \bar{\psi}, \bar{\theta})$ be the set of all valid paths $(u_0, y_0, \ldots, u_{k-1}, y_{k-1})$ witnessed by $\bar{x}$, $\bar{\psi}$, and $\bar{\theta}$. Now let $T \subseteq \mathcal{I} \times (U \times Y) \times \mathcal{I}$ be defined by

$$T = \Big\{(\eta, (u, y), \eta') \mid \text{ there exist } k \in \mathbb{N},$$
$$\bar{x} = (x_0, \ldots, x_{k-1}), \bar{\psi} = (\psi_0, \ldots, \psi_{k-1}),$$
$$\bar{\theta} = (\theta_0, \ldots, \theta_{k-2}), \theta \in \Theta(x_{k-1}, u) \text{ and } \psi \in \Psi(f(x_{k-1}, u, \theta))$$
$$\text{such that}$$
$$\eta \in \mathcal{I}^k(\bar{x}, \bar{\psi}, \bar{\theta}) \wedge \eta' \in \mathcal{I}^{k+1}(\bar{x}', \bar{\psi}', \bar{\theta}'),$$
$$\text{where } \bar{x}' = \bar{x}^\frown(f(x_{k-1}, u, \theta)),\ \bar{\psi}' = \psi^\frown(\psi),\ \text{and } \bar{\theta}' = \theta^\frown(\theta)\Big\}.$$

Here, $x^\frown y$ is the concatenation of sequences x and y. Then $(\mathcal{I}, U \times Y, T)$ is the *history I-space transition system*.

Suppose for each $x, y \in X$ there is at most one $u \in U$ with $x \xrightarrow{u} y$. Let

$$E_T = \{(x, y) \in X^2 \mid \exists u \in U(x \xrightarrow{u} y)\},$$

and let $l: E_T \to U$ be defined so that $l((x, y))$ is the unique u such that $x \xrightarrow{u} y$. Then (X, E_T, l, x_0) with $x_0 \in X$ is a passive I-state graph as in O'Kane and Shell (2017, Def 1).

The following definition is more of a notational than mathematical value.

Definition 2.6. Let $\mathcal{X} = (X, U, T)$ be a transition system. If for all $(x, u) \in X \times U$ there is a unique $y \in X$ with $(x, u, y) \in T$, then we denote the function $(x, u) \mapsto y$ by τ, and write (X, U, τ) instead of (X, U, T). In this case we call $\mathcal{X}$ an *automaton*. Note that usually in computer science literature an automaton is finite and also has an initial state and a set of accepting states, but we do not have those in our definition.

For automata we also use the notation $x * u = \tau(x, u)$ and if $\bar{u} = (u_0, \ldots, u_{k-1})$, then $x * \bar{u}$ is defined by induction for $k > 1$ as follows: $x * (u_0, \ldots, u_{k-1}) = (x * (u_0, \ldots, u_{k-2})) * u_{k-1}$.

Examples 2.7. Automata and transition systems can model agent-environment and related dynamics.

1. If $(X, \cdot)$ is a group, $U \subseteq X$ is a set of generators, and $\tau(x, u) = x \cdot u$, then (X, U, τ) is an automaton. For example, consider the situation in which $X = \mathbb{Z} \times \mathbb{Z}$ and $U = \{a, b, a^{-1}, b^{-1}\}$ in which $a = (1, 0)$ and $b = (0, 1)$. Thus, X is presented with generators a, b, and relation $a \cdot b = b \cdot a$. This models an agent moving without rotation in an infinite 2D-grid and the agent can move left, right, up and down. There are no obstacles. The standard Cayley graph is equivalent to the graph based representation of the automaton.
2. Let U^* be the set of all finite sequences ("strings") of elements of U. If $\bar{u} = (u_0, \ldots, u_{k-1}) \in U^*$ and $u_k \in U$, we denote by $\bar{u}^\frown u_k$ the *concatenation* $(u_0, \ldots, u_{k-1}, u_k)$. If $\bar{u}_0, \bar{u}_1 \in U^*$, then $\bar{u}_0{}^\frown \bar{u}_1$ is similarly the concatenation of two strings. The operation of concatenation turns U^* into a monoid. Suppose $\tau : X \times U^* \to X$ is an action of the monoid U^* on X meaning that it satisfies $\tau(\tau(x, \bar{u}), \bar{u}') = \tau(x, \bar{u}^\frown \bar{u}')$ and $\tau(x, \varnothing) = x$. Then the automaton (X, U, τ) is a discrete-time control system. A sequence of *controls* $\bar{u} = (u_0, \ldots, u_{k-1})$ produces a unique *trajectory* $(x_0, \ldots, x_k)$, given the initial state x_0 by induction: $x_{i+1} = \tau(x_i, u_i)$ for all $i < k$.
3. Consider an automaton (X, U, τ) in which U is a group, and τ is a group action of U on X. In some situations it can be natural to consider the set of motor-outputs of an agent to be a group: the neutral element is no motor-output at all, every motor-output has an "inverse" for which the effect is the opposite, or negating (say, moving right as opposed to moving left), the composition of movements is many movements applied consecutively. The action τ of U on X is then the realization of those motor-outputs in the environment. In realistic scenarios, however, this is not a good way to model the sensorimotor interaction because of the following reason. Suppose the agent has actions "left" and "right," but it is standing next to an obstacle on its left. Then moving "left" will result in staying still (because of the obstacle), but moving "right" will result in actually moving right, if there is no obstacle at the right of the agent. In this situation the sequence "left-right" results in a different position of the agent than the sequence "right-left," so if "left" and "right" are each other's inverses in G, then the axioms of group action are violated.
4. Note that if $T = \varnothing$, then (X, U, T) is a transition system.
5. Let $X = \{0, 1\}^*$ as in (2), $U = \{0\}$, and $(x, 0, y) \in T$ if and only if $|y| = |x| + 1$, then (X, U, T) is a transition system, where $|x|$ is the length of the string x.
6. If (X, U, T) is a transition system and $E \subseteq X$ an equivalence relation, then $(X/E, U, T/E)$ is a transition system, where $X/E = \{[x]_E \mid x \in X\}$ and $T/E = \{([x]_E, u, [y]_E) \mid (x, u, y) \in T\}$, and $/$ denotes a quotient space; see Definition 2.33.

2.3. Sensorimotor systems

Next, we will define a *sensorimotor system*, which is a special case of a transition system. Following the tenet (EA1) that "environment is inseparable from the body which is inseparable from the brain," our sensorimotor systems can model any part of the environment-body-brain coupling. The model that describes the environment differs from the one that describes the agent merely in the type of structure it possesess, but not in an essential mathematical way.

SM-systems can be thought of as a partial specification of (some part of) the brain-body-environment coupling. Physicalist determinism demands that under full specification[2] we are left with a deterministic system. A specification is partial when it leaves room for unknowns in some, or all, parts of the system.

Definition 2.8. A *sensorimotor system* (or *SM-system*) is a transition system (X, U, T) where $U = S \times M$ for some sets S and M, which we call in this context the *sensory set* and the *motor set*, *respectively*.

The interpretation is that if $x \xrightarrow{(s,m)} y$, then s is the sensation that either occurs at x, or along the transition to the next state y, and m the motor action which leads to the transition. We will show later how SM-systems can be connected together (Definition 2.22) to form coupled systems. Sometimes an SM-system is modeling a brain-body totality, and other times it is modeling body-environment totality. A coupling between these two will model the brain-body-environment totality. This is a flexible framework which enables enactivist-style analysis. We do not assume that the agent "knows" the effect of a given $m \in M$ or that the "meaning" of a given $s \in S$. The sets S and M are purely mathematical sets denoting the interface between the agent and the environment from the third person perspective.

In fact, the sensory and motor components can be decoupled which might be more natural from the mathematics' point of view in some cases. The following shows that we can look at it both ways.

2 This means a full specification of the environment, the agent's body, its brain, their coupling, as well as the initial states.

Definition 2.9. An *asynchronous SM-system* is a transition system (X, U, T) such that there exist partitions $U = S \cup M$ and $X = X_s \cup X_m$ such that for all $(x, u, y) \in T$ we have

1. if $x \in X_s$, then $u \in S$,
2. if $x \in X_m$, then $u \in M$, and
3. $x \in X_m \iff y \in X_s$.

Thus, the state space of a sequential SM-system contains separate *sensory states* and *motor states*.

Definition 2.10. Suppose E is an equivalence relation on a set X. We say that a map $f : X \to X$ is *E-preserving* if for all $x, y \in X$, we have $xEy \iff f(x)Ef(y)$.

There is a natural correspondence between SM-systems and their asynchronous counterpart:

Theorem 2.11. *Let* SM *and* aSM *be the classes of SM-systems and asynchronous SM-systems, respectively. There are functions* F: SM $\to$ aSM *and* G: aSM $\to$ SM *such that*

1. *F and G are isomorphism and bisimulation preserving,*
2. *restricted to finite systems, F and G are polynomial-time computable, and restricted to the infinite ones they are Borel-functions in the sense of classical descriptive set theory (Kechris, 1994).*

Proof: See Appendix B. □

Another type of a system, which is in a similar way equivalent to a special case of an SM-system, is a state-labeled transition system which we will introduce next, and prove a similar result, Lemma 2.19.

2.4. Quasifilters and quasipolicies

The amount of information specified in a given SM-system depends on which part of the brain-body-environment system we are modeling. At one extreme, we specify the environment's dynamics down to the small detail and leave the brain's dynamics completely unspecified. In this case the SM-system will have only one sensation corresponding to each state and the transition to the next state will be completely determined by knowing the motor action. This is, in a sense, the environment's perspective. At the other extreme, we specify the brain completely, but leave the environment unspecified. We "don't know" which sensation comes next, but we "know" which motor actions are we going to apply. This is in a sense the perspective of the agent. The first extreme case is the perspective often taken in robotics and other engineering fields when either specifying a planning problem (Ghallab et al., 2004; Choset et al., 2005; O'Kane and LaValle, 2008), or designing a filter (Hager, 1990; Thrun et al., 2005; LaValle, 2012; Särkkä, 2013) (also known as sensor fusion). This is why we call SM-systems of that sort *quasifilters* (Definition 2.12). The other extreme is the perspective of a policy. The policy depends on sensory input, but the motor actions are determined (by the policy). This is why we call the SM-systems of the latter sort *quasipolicy*. The "quasi-" prefix is used because both are weaker and more general notions than those that appear in the literature; see Remarks 2.20 and 2.21.

Another way to look at this is the dichotomy between virtual reality (VR), and robotics. In virtual reality, scientists are designing the (virtual) environment for an agent whereas in robotics they are typically designing an agent for an environment. In the former case the agent is partially specified: the type of embodiment is known (S and M are known) and some types of patterns of embodiment are known (eye-hand coordination). However, the specific actions to be taken by the agents are left unspecified. The job of the designer is to specify the environment down to the smallest detail, so that every sequence of motor actions of the agent yields targeted sensory feedback. The VR-designer is designing a quasifilter constrained by the partial knowledge of the agent's embodiment and internal dynamics. The case for the robot designer is the opposite. She has a partial specification of the robot's intended environment and usually works with a complete specification of the robot's mechanics. She is designing a quasipolicy. For VR-designers the agent is a black box; for roboticists the agent is a white box (Suomalainen et al., 2020) (unless the task is to reverse engineer an unknown robot design). For the environment, the roles are reversed. A similar dichotomy can be seen between biology (in which the agent is a black box) and robotics (in which it usually is a white box).

All the definitions in this section are new.

Definition 2.12. Suppose that $(X, S \times M, T)$ is an SM-system with the property that for all $x_1 \in X$ there exists $s_{x_1} \in S$ such that for all $x_2 \in X$ and all $(s, m) \in S \times M$ we have that $x_1 \xrightarrow{(s,m)} x_2$ implies $s = s_{x_1}$. Then, $(X, S \times M, T)$ is a *quasifilter*.

In a quasifilter the sensory part of the outgoing edge is unique. The dual notion (quasipolicy) is when the motor part is unique:

Definition 2.13. Suppose that $(X, S \times M, T)$ is an SM-system with the property that for all $x \in X$ there exists $m_x \in M$ such that for all $y \in X$ and all $(s, m) \in S \times M$ we have that $x \xrightarrow{(s,m)} y$ implies $m = m_x$. Then, $(X, S \times M, T)$ is a *quasipolicy*.

Before explaining the connections between quasifilter and a filter and quasipolicy and a policy, let us define projections of the sensorimotor transition relation to "motor" and to "sensory":

Definition 2.14. Given an SM-system $(X, S \times M, T)$, let

$$T_M = \{(x, m, y) \in X \times M \times X \mid \exists s \in S (x, (s, m), y) \in T\}$$
$$T_S = \{(x, s, y) \in X \times S \times X \mid \exists m \in M (x, (s, m), y) \in T\}.$$

These are called the *motor* and the *sensory projections*, respectively of the sensorimotor transition relation. They are

also called the *motor transition relation* and the *sensory transition relation*, respectively. The corresponding transition systems (X, M, T_M) and (X, S, T_S) are called the *motor* and the *sensory projection systems*.

Definition 2.15. Given a transition system (X, U, T), and $x \in X$, let $O_T(x) \subseteq U$ be defined as the set $O_T(x) = \{u \in U \mid (\exists y \in X)(x \xrightarrow{u} y)\}$. Combining this notation with the one introduced in Example 2.4(2), given $x, y \in X$, we have

$$O_T(x) = \bigcup_{y \in X} \tilde{T}(x, y).$$

For a transition relation $T \subseteq X \times (S \times M) \times X$, define its *transpose* by $T^t \subseteq X \times (S \times M) \times X$ such that $T^t = \{(x, (m, s), y) \mid (x, (s, m), y) \in T\}$. Note that $(T^t)^t = T$. For a subset of a Cartesian product $A \subseteq S \times M$, let A_1 be the projection to the first coordinate $A_1 = \{s \in S \mid (\exists m \in M)((s, m) \in A)\}$ and A_2 the projection to the second one: $A_2 = \{m \in M \mid (\exists s \in S)((s, m) \in A)\}$.

Mathematically coupling of two transition systems is symmetric [see Theorem 2.24(3)], but from the cognitive perspective there is (usually) an asymmetry between the agent and the environment (which can be evident from some specific properties of the agent and of the environment). Because of the partial symmetry, many properties of an agent can dually be held by the environment and vice versa. The following proposition highlights the duality between quasipolicy and quasifilters: reversing the roles of the environment and the agent.

Proposition 2.16. *For an SM-system $\mathcal{X} = (X, S \times M, T)$ the following are equivalent:*

1. *$\mathcal{X}$ is a quasifilter,*
2. *$\mathcal{X}^t = (X, S \times M, T^t)$ is a quasipolicy,*
3. *$O_{T_S} = (O_T(x))_2 = (O_{T^t}(x))_1$ is a singleton for each $x \in X$.*

Similarly, $\mathcal{X}$ is a quasipolicy if and only if $O_{T_M}(x) = (O_T(X))_1$ is a singleton for each $x \in X$.

Proof: A straightforward consequence of all the definitions. □

2.5. State-relabeled transition systems

It will become convenient in the coming framework to assign labels to the states. The elements x of the state space X are already named; thus, our labeling can be more properly considered as a *relabeling via* a function $h : X \to L$, in which L is an arbitrary set of *labels*. This allows partitions to be naturally induced over X by the preimages of h. Intuitively, this will allow the state space X to be characterized at different levels of "resolution" or "granularity." Thus, we have the following definition:

Definition 2.17. A *state-relabeled transition system* (or simply *labeled transition system*) is a quintuple (X, U, T, h, L) in which $h : X \to L$ is a labeling function and (X, U, T) is a transition system.

We think of *state-relabeled* to be a more descriptive term, but we shorten it in the remainder of this paper to being simply *labeled*.

Remark 2.18. In an analogy to Definition 2.6, a labeled transition system is a *labeled automaton*, if T happens to be a function; in other words, for all $(x, u) \in X \times U$ there is a unique $y \in X$ with $(x, u, y) \in T$. In this case we may denote this function by $\tau : (x, u) \mapsto y$ and work with the labeled automaton (X, U, τ, h, L). For example, the temporal filter in Section 2.1 is a labeled automaton.

The isomorphism and bisimulation relations are defined similary as for transition systems, but in a label-preserving way.

One intended application of a labeled transition system (X, U, T, h, L) is that h is a sensor mapping, L is a set of sensor observations, and U is a set of actions. Thus, actions $u \in U$ allow the agent to transition between states in X while h tells us what the agent senses in each state. We intend to show that this can be seen as a special case of an SM-system by proving a theorem similar to Theorem 2.11, but stronger, namely these corresponces preserve isomorphism:

Lemma 2.19. *Let $\mathcal{F}$ be the class of quasifilters, $\mathcal{P}$ the class of quasipolicies, and $\mathcal{L}$ the class of labeled systems. Then there are one-to-one maps*

$$\mathrm{LTS}_P : \mathcal{P} \to \mathcal{L} \text{ and } \mathrm{LTS}_F : \mathcal{F} \to \mathcal{L}$$

such that

1. *LTS_P and LTS_F are isomorphism and bisimulation preserving,*
2. *restricted to finite systems, LTS_P and LTS_F are polynomial-time computable, and restricted to the infinite ones they are Borel-functions in the sense of classical descriptive set theory.*

Proof: See Appendix B □

Remark 2.20. Let $\mathcal{X} = (X, S \times M, T)$ be a quasifilter and $\mathcal{X}' = \mathrm{LTS}_F(\mathcal{X}) = (X, M, T_M, h, S)$ as in Lemma 2.19. Suppose further that for each $x, y \in X$ there is at most one $u \in U$ with $x \xrightarrow{u} y$. Let

$$E_T = \{(x, y) \in X^2 \mid \exists u \in U(x \xrightarrow{u} y)\},$$

Then (X, M, E_T, x_0) coincides with the definition of a filter (O'Kane and Shell, 2017, Def 3). If it is also an automaton, meaning that above we replace "at most one" by "exactly one," then every sequence of motor actions $(m_0, \ldots, m_{k-1})$ determines a unique resulting state $x_{k-1} \in X$. This is analogous, and can be proved in the same way, as the fact that each sequence of sensory data determines a unique resulting state in Remark 2.21 below.

Remark 2.21. Usually, a *policy* is a function which describes how an agent chooses actions based on its own past experience. Thus, if M is the set of motor commands and S is the set of sensations,

a policy is a function $\pi : S^* \to M$ where S^* is the set of finite sequences of sensory "histories"; see for example (LaValle, 2006). Now, suppose that an SM-system $\mathcal{X} = (X, S \times M, T)$ is a quasipolicy in the sense of Definition 2.13 and let $x \mapsto m_x$ be as in that Definition. Assume further that $\mathcal{X}$ is an automaton (Section 2.1) and let $\tau : X \times (S \times M) \to X$ be the corresponding transition function so that for all $x \in X$ and $(s, m) \in S \times M$ we have $(x, (s, m), \tau(x, (s, m))) \in T$. Let $x_0 \in X$ be an initial state. We will show how the pair $(\mathcal{X}, x_0)$ defines a function $\pi : S^* \to M$ in a natural way. Let $\bar{s} = (s_0, \ldots, s_{k-1}) \in S^k$ be a sequence of sensory data. If $k = 0$, and so $\bar{s} = () = \varnothing$, let $\pi(\bar{s}) = m_{x_0}$. If $k > 0$, assume that $\pi(s_0, \ldots, s_{k-2})$ and x_{k-1} are both defined (induction hypothesis). Then let $x_k = \tau(x_{k-1}, (m_{x_{k-1}}, s_{k-1}))$ and $\pi(s_0, \ldots, s_{k-2}, s_{k-1}) = m_{x_k}$. The idea is that because of the uniqueness of m_x, a sequence of sensory data determines (given an initial state) a unique path through the automaton $\mathcal{X}$.

2.6. Couplings of transition systems

The central concept of this work pertaining to all principles (EA1)–(EA5) is the coupling of SM-systems. We define coupling, however, for general transition systems with the understanding that our most interesting applications will be for SM-systems where $U_0 = U_1 = S \times M$. The idea is that in every transition there is a sensory component and a motor component. The set S could be thought of as all possible events that trigger afferent nervous signals, or their combinations. The elements of M are those events that are triggered by efferent nervous signals. This is an abstract space and in transitioning from one state to another some subset of $S \times M$ is "active." If we know little of what kind of sensory data the agent receives during the transition, then that transition will occupy a subset of $S \times M$ whose projection to the S-coordinate is large. If, on the other hand we know a lot, and can specify the exact sensory data, then the projection to the S-coordinate is small. Vice versa, if we do not know which motor actions lead from one state to another, then the projection of the corresponding subset to the M-coordinate is large etc. This was made more precise in Section 2.4. The fact that the transition consists of pairs (s, m) where s is a sensory input and m is a motor command does not mean that the agent is equipped with the semantics of what m "means," or what it "does" in the world. The effect of m is "computed" by the environment and the agent only receives the next "s" as the feedback. It might have been more intuitive, but more cumbersome to make this definition in terms of functions that map events of the environment to sensory stimuli and internal events of the nervous system to motor actions, and further functions that map the motor actions to the actual events in the environment, etc., but from the point of view of essential mathematical structure these extra identifications wouldn't add anything qualitatively new.

Definition 2.22. Let $\mathcal{X}_0 = (X_0, U_0, T_0)$ and $\mathcal{X}_1 = (X_1, U_1, T_1)$ be two transition systems. The *coupled* system $\mathcal{X}_0 * \mathcal{X}_1$ is the transition system (X, U, T) defined as follows: $X = X_0 \times X_1$, $U = U_0 \cap U_1$, and

$$T = T_0 * T_1 = \{((x_0, x_1), u, (y_0, y_1)) \mid (x_0, u, y_0) \in T_0 \wedge (x_1, u, y_1) \in T_1\}.$$

Equivalently, for all $((x_0, x_1), (y_0, y_1)) \in (X_0 \times X_1)^2$ we have

$$\tilde{T}((x_0, y_0), (x_1, y_1)) = \tilde{T}_0(x_0, x_1) \cap \tilde{T}_1(y_0, y_1)$$

(recall the $\tilde{T}$ notation from Example 2.4(2)).

Example 2.23. A simple example of coupling is illustrated in Figure 2.

Mathematically the coupling is a product of sorts. If we think of one transition system as "the environment" and the other as "the agent," then the coupling tells us about all possible ways in which the agent can engage with the environment. The fact that the state space of the coupled system is the product of the state spaces of the two initial systems reflects the fact that the coupled system includes information of "what would happen" if the environment was in any given state while the agent is in any given ("internal") state.

We immediately prove the first theorem concerning coupling:

Theorem 2.24. *Suppose that $\mathcal{X}_i = (X_i, U_i, T_i)$ and $\mathcal{X}'_i = (X'_i, U'_i, T'_i)$ for $i \in \{0, 1\}$ are four SM-systems. Then the following hold:*

1. *If $\mathcal{X}_i \cong \mathcal{X}'_i$ for $i \in \{0, 1\}$, then $\mathcal{X}_0 * \mathcal{X}_1 \cong \mathcal{X}'_0 * \mathcal{X}'_1$.*
2. *If $\mathcal{X}_i \sim \mathcal{X}'_i$ for $i \in \{0, 1\}$, then $\mathcal{X}_0 * \mathcal{X}_1 \sim \mathcal{X}'_0 * \mathcal{X}'_1$.*
3. *$\mathcal{X}_0 * \mathcal{X}_1 \cong \mathcal{X}_1 * \mathcal{X}_0$.*

Proof: See Appendix B □

Coupling provides an interesting way to compare SM-systems from the "point of view" of other SM-systems. For example, given an SM-system $\mathcal{E}$ one can define an equivalence relation on SM-systems by saying that $\mathcal{I} \sim_E \mathcal{I}'$, if $\mathcal{E} * \mathcal{I} = \mathcal{E} * \mathcal{I}'$. If $\mathcal{E}$ is the "environment" and $\mathcal{I}, \mathcal{I}'$ are "agents," this is saying that the agents perform identically in this particular environment. Or vice versa, for a fixed $\mathcal{I}$, the relation $\mathcal{E} * \mathcal{I} = \mathcal{E}' * \mathcal{I}$ means that the environments are indistinguishable by the agent $\mathcal{I}$.

Remark 2.25. In the definition of coupling we see that the two SM-systems constrain each other. This is seen from the fact that in the definition we take intersections. For example, when an agent is coupled to an environment, it chooses certain actions from a large range of possibilities. In this way the agent structures its own world through the coupling (EA3). To make this notion further connect to enactivist paradigm, we invoke the dynamical systems

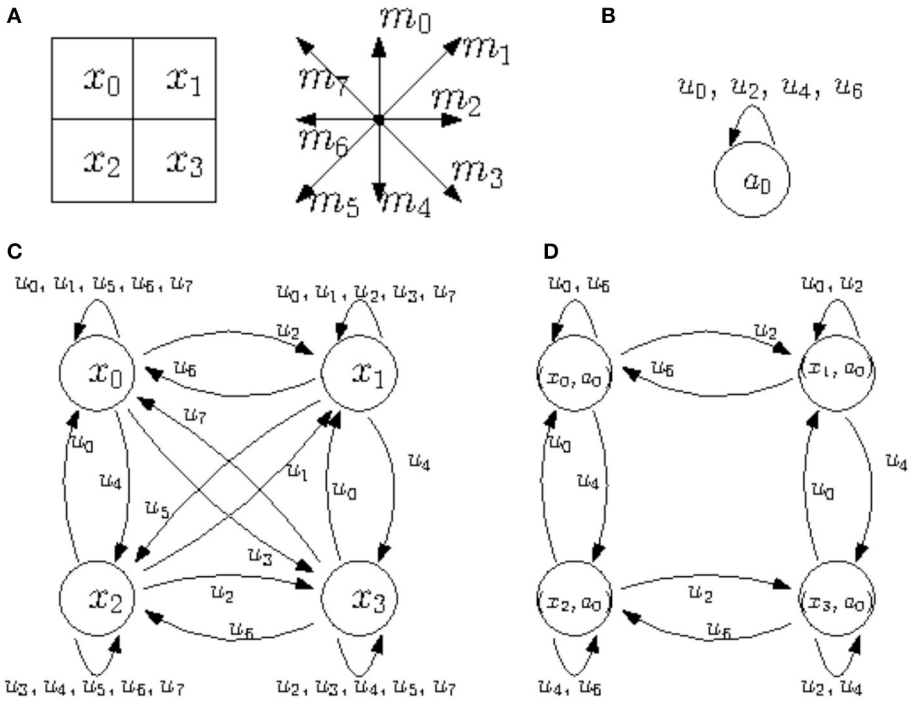

FIGURE 2
(A) States and actions for the transition system $\mathcal{X}_0$ that describes a 2-by-2 grid. 8 actions populating the set $M = \{m_0, \ldots, m_7\}$ correspond to a move (to a neighbor cell if possible) either sideways or diagonally. Suppose S is a singleton such that $S = \{s\}$. Then, in the following, u_i corresponds to the transition $u_i = (m_i, s)$ for $i = 1, \ldots, 7$. **(B)** Transition system $\mathcal{X}_1$. **(C)** Transition system $\mathcal{X}_0$. **(D)** The coupled system $\mathcal{X}_0 * \mathcal{X}_1$.

approach to cognition (Tschacher and Dauwalder, 2003). An *attractor* in a transition system $\mathcal{X} = (X, U, T)$ is a set $A \subseteq X$ with the property that for all infinite sequences

$$x_0 \overset{u_0}{\to} x_1 \overset{u_1}{\to} \cdots x_{k-1} \overset{u_{k-1}}{\to} x_k \overset{u_k}{\to} \cdots$$

there are infinitely many indices n such that $x_n \in A$. There could be other possible definitions, such as "for all large enough n, $x_n \in A$". For the present illustration purposes it is, however, irrelevant. It could be the case that $A \subseteq X$ is not an attractor of $\mathcal{X}$, but after coupling with $\mathcal{X}' = (X', U', T')$, $A \times X'$ may be an attractor of $\mathcal{X} * \mathcal{X}'$. Thus, if $\mathcal{X}$ is the environment and $\mathcal{X}'$ is the agent and A is a set of desirable environmental states, then we may say that the agent is well attuned to $\mathcal{X}$, if A was not initially an attractor, but in $\mathcal{X} * \mathcal{X}'$, then $A \times X'$ becomes one. It could also be that the agent needs to arrive to A while being in a certain type of an internal state $B \subseteq X'$, for example, if A is "food" and B is "hungry". Then it is not important that $A \times X'$ is an attractor, but it is imperative that $A \times B$ is one.

2.7. Unconstrained and fully constrained SM-systems

As we mentioned before, the information specified in an SM-system depends on which part of the brain-body-environment system we are modeling. In the extreme case we do not specify *anything*, except for the very minimal information. Consider a body of a robot for which the set of possible actions (or motor commands) is M and the set of possible sensor observations is S. Suppose that is all we know about the robot. We do not know what kind of environment it is in and we do not know what kind of "brain" (a processor or an algorithm) it is equipped with. Thus, we do not know of any constraints the robot may have in sensing or moving. We then model this robot as an *unconstrained* SM-system:

Definition 2.26. An SM-system $(X, S \times M, T)$ is called *unconstrained* iff for all $x \in X$, we have $O_T(x) = S \times M$; recall Definition 2.15.

Unconstrained systems have the role of a neutral element with respect to coupling (Proposition 2.29). We now show that given all unconstrained SM-systems with shared M and S are mutually bisimulation equivalent:

Proposition 2.27. *Suppose that* $\mathcal{X} = (X, S \times M, T)$ *and* $\mathcal{X}' = (X', S \times M, T')$ *are unconstrained systems. Then* $\mathcal{X} \sim \mathcal{X}'$.

Proof: See Appendix B □

There are many intuitions behind the above. An unconstrained system is one where anything could happen: the agent might perform any actions in any order and the environment could provide the agent with any sensory data. Such a world is reminiscent of *white noise*. Such a system is only interesting from an abstract mathematical perspective, it is in some sense "maximal". The content of Proposition 2.27 is that such systems are indistinguishable from each other. An unconstrained system has a similar role with respect to all SM-systems as the free group has to other groups, although we haven't made this universality claim precise in the present paper. Intuitively it means that every possible agent-environment combination can be found as a subsystem (or possibly a quotient) of the unconstrained one. The term "unconstrained" refers in particular to that when coupled to other systems, this system doesn't constrain them, so it acts in the same way as 0 in arithmetic addition (Proposition 2.29). The opposite is the fully constrained system (Definition 2.31, Proposition 2.32). In that case, the intuition is the opposite: in environments where nothing happens and actions do not have any effects, any agent is as good as any other and vice versa: agents that don't do anything are equivalent.

Corollary 2.28. *The SM-system* $\varepsilon = (\{0\}, \{0\} \times (S \times M) \times \{0\})$ *is the unique, up to bisimulation, unconstrained system.*

Proposition 2.29. *Let* ε *be as in Corollary 2.28 and let* $\mathcal{X} = (X, S \times M, T)$ *be any SM-system. Then* $\mathcal{X} * \varepsilon \cong \mathcal{X}$.

Corollary 2.30. *If* $\mathcal{X}$ *and* $\mathcal{X}'$ *are SM-systems and* $\mathcal{X}'$ *is unconstrained, then* $\mathcal{X} * \mathcal{X}' \sim \mathcal{X}$.

Proof: By Corollary 2.28 $\mathcal{X}' \sim \varepsilon$, So by Theorem 2.24 we have $\mathcal{X} * \mathcal{X}' \sim \mathcal{X} * \varepsilon$. However, by Proposition 2.29, $\mathcal{X} * \varepsilon \sim \mathcal{X}$; thus, $\mathcal{X} * \mathcal{X}' \sim \mathcal{X}$. □

The opposite of an unconstrained system is a fully constrained one:

Definition 2.31. An SM-system $(X, S \times M, T)$ is *fully constrained* iff $T = \varnothing$.

Proposition 2.32. *Dually to the propositions above, we have that (1) all fully constrained systems are bisimulation equivalent to each other, (2) the simplest example being* $\lambda = (\{0\}, S \times M, \varnothing)$, *and (3) if* $\mathcal{X} = (X, S \times M, T)$ *is another SM-system, then* $\mathcal{X} * \lambda \sim \lambda$.

All transition systems are in some sense between the fully constrained and the unconstrained, these being the two theoretical extremes.

2.8. Quotients of transition systems

When considering labelings and their induced equivalence relations, it will be convenient to develop a notion of quotient systems, analogous to quotient spaces in topology. Suppose $\mathcal{X} = (X, U, T)$ is a transition system and E is an equivalence relation on X. We can then form a new transition system, called the *quotient* of $\mathcal{X}$ by E in which the new states are E-equivalence classes and the transition relation is modified accordingly.

The following definition of a quotient is standard in Kripke model theory, especially bisimulation theory:

Definition 2.33. Suppose $\mathcal{X} = (X, U, T)$ and E are as above. Let $X/E = \{[x]_E \mid x \in X\}$, in which each $[x]_E$ is an equivalence class of states x under relation E, and $T/E = \{([x]_E, u, [y]_E) \mid (x, u, y) \in T\}$. Then $\mathcal{X}/E = (X/E, U, T/E)$ is the *quotient* of (X, U, T) by E.

The following definition is inspired by the idea of sensory pre-images, see LaValle (2019), but is also needed for technical reasons.

Definition 2.34. Given any function $h\colon X \to L$, denote by E^h the inverse-image equivalence: $E^h = \{(x, y) \in X^2 \mid h(x) = h(y)\}$. We will denote the equivalence classes of E^h by $[x]_h$ instead of $[x]_{E^h}$ if no confusion is possible.

The equivalence relation E^h partitions X according to the preimages of h, as considered in the sensor lattice theory of LaValle (2019). The partition of X induced by h directly yields an quotient transition system by applying the previous two definitions:

Definition 2.35. Let $\mathcal{X} = (X, U, T)$ be a transition system and $h\colon X \to L$ be any mapping. Then define $\mathcal{X}/h$ to be $\mathcal{X}/E^h$ where we combine Definitions 2.34 and 2.33.

Proposition 2.36. *If* h *is one-to-one, then* $\mathcal{X}/h \cong \mathcal{X}$.

Proof: h is one-to-one if and only if E^h is equality, in which case it is straightforward to verify that the function $x \mapsto [x]_{E^h}$ is an isomorphism. □

For $h\colon X \to L$, the transition system $(X/h, U, T/h)$ is essentially a new state space over the preimages of h. In this case $\mathcal{X}/h$ is called the *derived information space* (as used in LaValle, 2006). More precisely:

Proposition 2.37. *Let* $L' = \mathrm{ran}(h) \subseteq L$. *Define*

$$T' = \{(l, u, l') \in L' \times U \times L' \mid (h^{-1}(l), u, h^{-1}(l')) \in T/h\}$$
$$= \{(h(x), u, h(y)) \mid (x, u, y) \in T\}.$$

Then $(X/h, U, T/h)$ *is isomorphic to* (L', U, T') *via the isomorphism* $f\colon [x]_{E^h} \mapsto h(x)$.

Proof: See Appendix B □

The intuitive meaning of the quotient is the following. There is a Soviet comedy film from the 1970's where the main character ends up in an apartment in Leningrad, while he thinks that he is actually in Moscow. The apartement in Leningrad is identical to his home in Moscow and he cannot distinguish between them. He thinks for a while that he is at his home in Moscow while being in an apartment in Leningrad. Even his key from Moscow worked for the Leningrad apartment. The pun is that in Soviet times all houses were built according to the same blueprint. Now, before he realized his situation, as far as he was concerned, he *was* in Moscow. He thought he came to the same place in the evening as in the morning, while he actually didn't. The idea of the quotient captures exactly that: We identify those states that "look the same" (the label is the same) even though they are actually different states. In fact, let us look at a cognitive system on several levels of granularity: When I type on my laptop at home or in a cafeteria, my fingers experience the keyboard in (approximately) the same way. As far as my fingers (and associated motor areas) are concerned, we can identify all situations where they are pressing keys on my keyboard. On a higher level, I might be coming home after a 10 h time and experience as if I am in the same place, but we all know that the planet, on which my home is, has moved, so I actually am not in the same place, just like the main character in the movie referenced above.

3. Illustrative examples of SM-systems

We next illustrate how sensorimotor systems model body-environment, brain-body, and brain-body-environment couplings. Consider a body in a fully understood and specified deterministic environment. In this case the body-environment system will be modeled by a quasifilter, Definition 2.12. Instead of using the quasifilter definition, we work with a labeled transition system which, according to Proposition 2.19, is equivalent. According to the assumption of full specification, we will in fact work with labeled automata.

The body has a set M of possible motor actions each of which has a deterministic influence on the body-environment dynamics. Denote the set of body-environment states by E_0. Whenever a motor action $m \in M$ is applied at a body-environment state $e \in E$, a new body-environment state $A(e,m) \in E$ is achieved. At each state $e \in E$ the body senses data $\sigma(e)$. Denote the set of sensations by S. In this way, the labeled automaton $\mathcal{E}_0 = (E, M, A, \sigma, S)$ models this body-environment system. This model is ambivalent toward the agent's internal dynamics, its strategies, policies and so on, but not ambivalent toward its embodiment and its environment's structure. In fact, it characterizes them completely.

Alternatively, consider a brain in a body, and suppose that the brain is fully understood and deterministic (for example, perhaps it is designed by us), but we do not know which environment it is in. We model this by an SM-system which is a quasipolicy. Again, by the analogous considerations as above, we work directly an equivalent labeled automaton specification. Denote the set of internal states of the brain by I. The agent's internal state is a function of the sensations; therefore, let $B\colon I \times S \to I$ be a function (B stands for *brain*) that takes one internal state to another based on new sensory data. At each internal state, the agent produces a motor output which is an element of the set M; therefore, let $\mu\colon I \to M$ be a function assigning a motor output to each internal state. Now, $\mathcal{I} = (I, S, B, \mu, M)$ is a labeled transition system modeling this agent. It is ambivalent toward the type of the environment the agent is in, but it is not ambivalent toward the agent's internal dynamics, policies, strategies and so on; in fact, it determines them completely.

Now, the coupling of the environment $\mathcal{E}$ and the agent $\mathcal{A}$ is the SM-system obtained as

$$\mathrm{LTS}_F^{-1}(\mathcal{E}) * \mathrm{LTS}_P^{-1}(\mathcal{A}).$$

The sensory and motor sets S and M capture the interface between the brain and the environment because they characterize the body (but not the *embodiment*).

Example 3.1. Consider an agent that has four motor outputs, called "up" (U), "down" (D), "left" (L), and "right" (R), and there is no sensor feedback (this defines the body). In Corollary 2.28 we gave a minimal example of an unconstrained SM-system. On the other extreme one can give large examples. For instance the free monoid generated by the set $M = \{U, D, L, R\}$.

Let X be the set of all possible finite strings in the four "letter" alphabet M, let $T = \{(x, m, y) \mid x^\frown m = y\}$. "No sensor data" is equivalent to always having the same sensor data; thus, we can assume that $S = \{s_0\}$ is a singleton and the sensor mapping $h\colon X \to S$ is constant.[3] The resulting unconstrained transition system $\mathcal{U} = (X, T, M, \sigma, S)$ can be represented by an infinite quaternary tree, shown in Figure 3A.

Suppose that this body is situated in a 2×2 grid. The body can occupy one of the four grid's squares at a time, and when it applies one of the movements, it either moves correspondingly, or, if there is a wall blocking the movement, it doesn't. This defines the body-environment system. The set of states is now E and has four elements corresponding to all the possible positions of the body. The transition function $A\colon E \times M \to E$ tells where to move, and the rest is as above. The system $\mathcal{E} = (E, A, M, \sigma, S)$ is shown in Figure 3B. Let us now look at the agent. Suppose that it applies the following policy: (1) In the beginning move left; (2) if the previous move was to the left, then move right, otherwise

3 We do not mean to say that no data is always the same as some other data. We are talking here about an agent that *never* receives any data, or an agent that *always* receives the same data. Thus, it cannot rely on any "change" between having and not having any sensory input. Thus, there is no "presense in absence" paradox here.

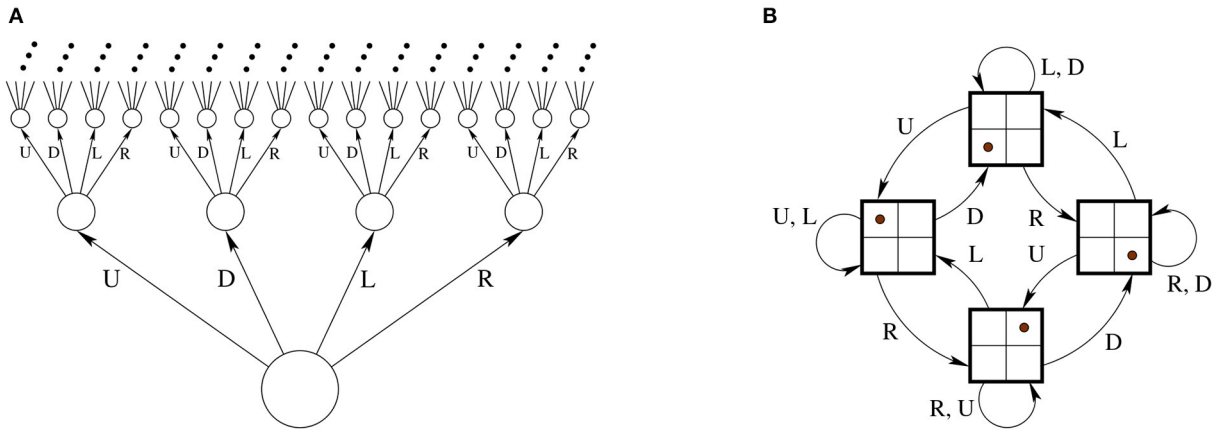

FIGURE 3
(A) Having motor commands and no sensory feedback leads to an infinite tree automaton. **(B)** Once the body is coupled with a 2 × 2 grid environment, a four-state automaton results.

move left. This can be modeled with a two-state automaton $\mathcal{I} = (I, S, B, \mu, M)$ where $I = \{L, R\}$, $S = \{s_0\}$, $B(L, s_0) = R$, $B(R, s_0) = L$, $\mu(L) = l$ and $\mu(R) = r$. Now, the coupling $\mathrm{LTS}_F^{-1}(\mathcal{E}) * \mathrm{LTS}_P^{-1}(\mathcal{I})$ is an automaton that realizes the policy in the environment, as shown in Figure 4A.

If the agent has a different embodiment in the same environment, then all of the automata will look different. Suppose that instead of the previous four actions, the agent has two: "rotate 90-degrees counterclockwise" (C),"forward one step" (F). Note that these are expressed in the local frame of the robot: It can either rotate relative to its current orientation, or it can move in the direction it is facing; the previous four actions were expressed as if in a global frame or the robot is incapable of rotation. Under the new embodiment, the unconstrained automaton with no sensor feedback is an infinite *binary tree*, with every node having two outgoing edges, labeled C and F, respectively, instead of the quaternary infinite tree depicted on Figure 3A. Instead of the four-state automaton of Figure 3, the automaton describing the environment transitions is a 16 state-automaton, because the orientation of the agent can now have four different values. See Figure 4B. Finally the automaton describing the internal mechanics of the agent $\mathcal{I}$ is a quasipolicy in these two actions, and finally, the coupling corresponds essentially to taking a path in the 16-state automaton above.

Note that there is a bisimulation between $\mathcal{U}$ and $\mathcal{E}$ which reflects the fact that from the point of view of an agent they are indistinguishable. This is natural because there is no sensory data, so from the agent's viewpoint it is unknowable whether or not it is embedded in an environment. A bisimulation is given as follows: Let $y_0 \in Y$ be the top-right corner and $x_0 \in X$ the root of the tree. Define $R \subseteq X \times Y$ be the minimal set satisfying the following conditions:

1. $(x_0, y_0) \in R$.
2. If $(x, y) \in R$ and $m \in M$, then $(T(x, m), U(y, m)) \in R$.

Example 3.2. The 16-state automaton of Example 3.1 has four automorphisms corresponding to the rotation of the environment by 90 degrees counterclockwise. Each of those automorphisms corresponds to an auto-bisimulation. Mirroring is not an automorphism because the agent's rotating action fixes the orientation of the automaton.

Example 3.3. Figure 5 shows an example of how an automaton with non-trivial sensing could look. Jumping a little bit ahead, it will be seen that the labeling provided by h in this figure is not sufficient (a notion introduced in Definition 4.2).

4. Sufficient refinements and degree of insufficiency

This section presents the concept of *sufficiency*, which will be the main glue between enactivist philosophy and mathematical understanding of cognition. In Section 4.1 we introduce the main concepts and explain its profound relevance to enactivist modeling and how it can be a precursor to the emergence of meaning from meaningless sensorimotor interactions. In Section 4.2 we introduce the notion of minimal sufficient refinements, prove a uniqueness result about them, and show how they are connected to the classical notions of bisimulation as well as derived information state spaces[4].

4 There could be an interesting relationship between this concept and the free energy principle proposed by K. Friston. A system which is attuned to its environment in a sufficient way can be interpreted by an inspector as a system that is making] predictions about its environment.

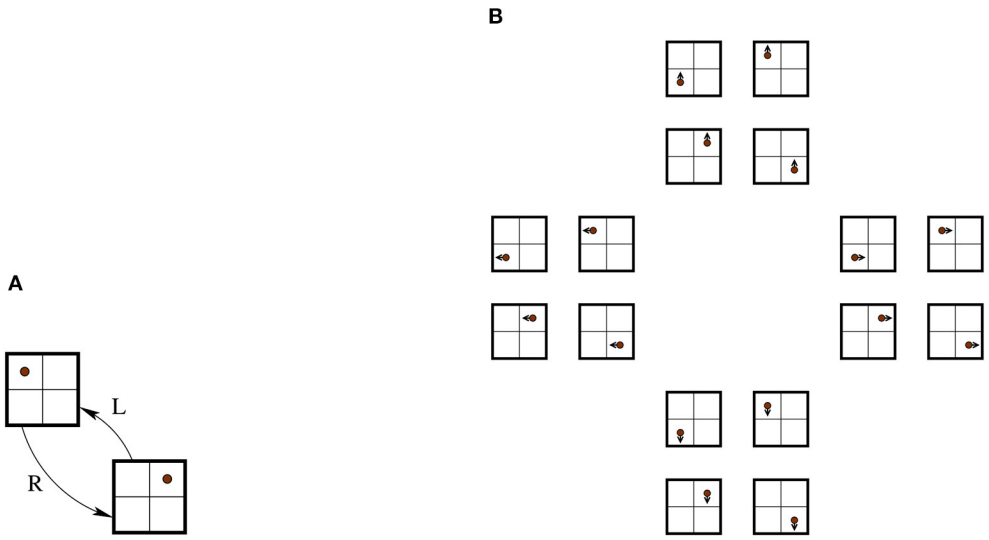

FIGURE 4
(A) A two-state automaton results from the realized policy. **(B)** If there are only two actions (rotate 90 degrees counterclockwise and going straight) then the second automaton has 16 states instead of four as in Figure 3B.

4.1. Sufficiency

The following consider the main definition of this work. It is based on the idea of sufficiency in LaValle (2006, Ch.11).

Definition 4.1. Let (X, U, T) be a transition system and $E \subseteq X \times X$ an equivalence relation. We say that E is *sufficient* or *completely sufficient*, if for all $(x, y) \in E$ and all $u \in U$, if $(x, u, x') \in T$ and $(y, u, y') \in T$, then $(x', y') \in E$.

This means that if an agent cannot distinguish between states x and y, then there are no actions it could apply to later distinguish between them. To put it differently, if the states are indistinguishable by an instant sensory reading, then they are in fact indistinguishable even through sensorimotor interaction. This is related to the equivalence relation known as Myhill-Nerode congruence in automata theory.

The equivalence relation of indistinguishability in the context of sensorimotor interactions is at its simplest the consequence of indistinguishability by sensors. Thus, we define sufficiency for labelings or sensor mappings:

Definition 4.2. A labeling $h\colon X \to L$ is called *sufficient* (or *completely sufficient*) iff for all $x, y, x', y' \in X$ and all $u \in U$, the following implication holds:

$$(h(x) = h(y) \wedge (x, u, x') \in T \wedge (y, u, y') \in T) \Rightarrow h(x') = h(y').$$

Proposition 4.3. *If (X, U, τ) is an automaton, then $h\colon X \to L$ is sufficient if and only if for all $x, y \in X$ and all $u \in U$, we have that if $h(x) = h(y)$, then $h(\tau(x, u)) = h(\tau(y, u))$.*

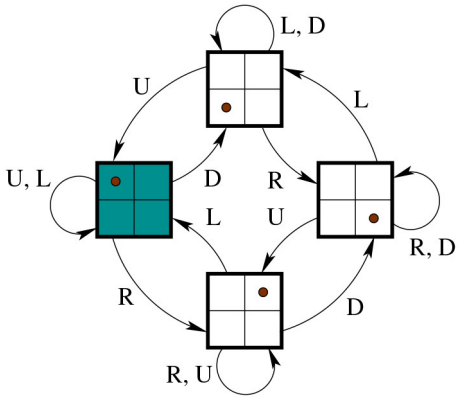

FIGURE 5
Consider the automaton $\mathcal{E}$ of Figure 3B from Example 3.1, but assume that the agent can "smell" a different scent in the top-left corner. This can be modeled by having a two-element set $S = \{0, 1\}$ instead of a singleton, and $h\colon X \to \{0, 1\}$ such that $h(x) = 0$ iff x is not the top-left corner. The state with a scent is shaded.

Proof: Checking the definitions. □

The above proposition is saying that when the sensorimotor system is deterministic, then sufficiency is equivalent to predictability.

There is a connection with the classical notion of bisimulation in classical transition systems theory (recall Definition 2.2):

Proposition 4.4. *An equivalence relation on a state space of an automaton (X, U, τ) is sufficient if and only if it is an autobisimulation.*

Proof: See Appendix B. □

The above proposition can intuitively be interpreted as saying that a sufficient relation is one where different states with the same label are not only indistinguishable on their own, but are actually indistinguishable even by their consequences. Starting from one of two states with same labels, there is no way to ever find out which one of them it was, no matter how much will the agent investigate its environment, compare to the discussion in the end of Section 2.8.

Proposition 4.5 below is an important proposition on which the idea of derived I-spaces and combinatorial filters builds upon (LaValle, 2006, 2012; O'Kane and Shell, 2017), although as far as the authors are aware, in the literature, only the "if"-direction is mentioned. We say that a transition system (X, U, T) is *full*, if for all $x_1 \in X$ and all $u \in U$ there exists at least one $x_2 \in X$ with (x_1, u, x_2).

Proposition 4.5. *Suppose $\mathcal{X} = (X, U, T)$ is a transition system. Let $h\colon X \to L$ be a labeling. Then $\mathcal{X}/h$ is an automaton if and only if $\mathcal{X}$ is full and h is sufficient.*

Proof: See Appendix B. □

The above proposition brings together the ideas of a quotient, automaton and sufficiency. The idea of the quotient is that indistinguishable states can be in some circumstances considered the same and the idea of an automaton is that it is deterministic. The above proposition says that as far as the agent is concerned, if it equalizes indistinguishable states, then the world looks deterministic from the agent's perspective if and only if the underlying labeling satsifies Definition 4.2.

The sufficiency of an information mapping was introduced in LaValle (2006, Ch 11), and is encompassed by a sufficient labeling in this paper. In the prior context, it has meant that the current sensory perception together with the next action determine the next sensory perception. The elegance with respect to our principle (EA2) is that sufficiency is *not* saying that the agent's internal state corresponds to the environment's state (as is in representational models). Nor is it saying that the agent *predicts* the next action. It is saying, rather, that the agent's current sensation together with a choice of a motor command *determine* the agent's next sensation; and this statement is true only as a statement made about the system from outside, not as a statement which would reside "in the agent." The sensation may carry no meaning at all "about" what is actually "out there." However, if the agent has found a way to be coupled to the environment in a sufficient way, then sensations *begin* to be *about* future sensation. In this way meaning emerges from sensorimotor patterns. This relates to (EA3) and somewhat touches on the topic of perception (EA5). Furthermore, the property of determining future outcomes is related to (EA4) because that is what *skill* is. There is no potential to *reliably* engage with the environment in complex sensorimotor interactions, if the sensations do not *reliably* follow certain historical patterns.

Thus, the notion of sufficiency is considered by us to be of fundamental importance for enactivist-inspired mathematical modeling of cognition. The violation of sufficiency means that the current sensation-action pair does not correlate with the future sensation, making it harder to be attuned to the environment. Having a different sensation following the same pattern can be seen as a primitive notion of a "surprise." This can be seen as aligning with the predictive coding and the free energy principle from neuroscience (Rao and Ballard, 1999; Friston and Kiebel, 2009; Friston, 2010), although our framework leaves the space to a clean non-representational interpretation while this is not obvious for these other frameworks. Does the notion of sufficient labelings capture the same ideas in a more general way? This is an open question for further research.

A generalization of sufficiency is n-sufficiency, in which the data of n previous steps is needed to determine the next label. Here, we define an n-chain.

Definition 4.6. An *n-chain* in $\mathcal{X} = (X, U, T)$ is a sequence

$$c = (x_0, u_0, \cdots, x_{n-1}, u_{n-1}, x_n) \in (X \times U)^n \times X$$

such that $x_i \xrightarrow{u} x_{i+1}$ for all $i < n$. If $n = 0$, then by convention $c = (x_n)$. Let $E \subseteq X \times X$ be an equivalence relation. Let $k < n$. We say that two n-chains $c = (x_0, u_0, \ldots, x_{n-1}, u_{n-1}, x_n)$, $c' = (x'_0, u'_0, \ldots, x'_{n-1}, u'_{n-1}, x'_n)$ are (T, E, k)-equivalent if for all $i < k$, we have $u_i = u'_i$ and $(x_i, x'_i) \in E$. An ∞-chain is defined in the same way as n-chain, except the sequences are infinite, without the "last" x_n.

Definition 4.7. For a transition system $\mathcal{X} = (X, U, T)$, an equivalence relation E on X is called *n-sufficient* if there are no two (T, E, n)-equivalent n-chains

$$c = (x_0, u_0, \ldots, x_{n-1}, u_{n-1}, x_n) \text{ and}$$
$$c' = (x'_0, u'_0, \ldots, x'_{n-1}, u'_{n-1}, x'_n)$$

such that $(x_n, x'_n) \notin E$. A labeling $h\colon X \to L$ is called *n-sufficient* if E^h is n-sufficient (Recall Definition 2.34).

Proposition 4.8. *An equivalence relation E is 0-sufficient if and only if there is only one E-equivalence class, and a labeling function h is 0-sufficient if and only if it is constant.*

Proof: See Appendix B □

Proposition 4.9. *An equivalence relation E (resp. a labeling h) is sufficient if and only if it is 1-sufficient.*

Proof: See Appendix B □

Proposition 4.10. *Suppose $n < m$ are natural numbers. If a labeling h is n-sufficient, then it is m-sufficient. The same holds for equivalence relations.*

Proof: See Appendix B □

This enables us to define the degree of insufficiency:

Definition 4.11. The *degree of insufficiency* of the labeled automaton $\mathcal{X} = (X, U, \tau, h, L)$ is defined to be the smallest n such that h is n-sufficient, if such n exists, and ∞ otherwise. Denote the degree of insufficiency of $\mathcal{X}$ by degins($\mathcal{X}$), or degins(h) if only the labeling needs to be specified and $\mathcal{X}$ is clear from the context.

The intuition is that the larger the degree of insufficiency of an environment $\mathcal{X}$, the harder it is for an agent to be attuned to it. We talk more about the connection between attunement and sufficiency in the following sections.

4.2. Minimal sufficient refinements

In this section we prove that the minimal sufficient refinements are always unique (Theorem 4.19). This will follow from a deeper result that the sufficient equivalence relations form a complete sublattice of the lattice of all equivalence relations. This does not hold for n-sufficient equivalence relations for $n > 1$ (Example 4.20). We will then explore how the minimal sufficient refinements can be thought of as an enactive perceptual construct that emerges from the body-environment, brain-body, and brain-body-environment dynamics. The idea is that a minimal sufficient refiniment corresponds to an optimal attunement of the agent to the base labeling which corresponds to some minimal information that the agent is interested in the environment, such as death or life, danger or safety information. It is "optimal" by minimality and "attunement" by sufficiency. Our Theorem 4.19 states that such attunement is mathematically unique.

Definition 4.12. An equivalence relation E is a *refinement* of equivalence relation E', if $E \subseteq E'$, also denoted $E' \leqslant_r E$. A labeling function h is a refinement of a labeling function h', if E^h is a refinement of $E^{h'}$.

An important interpretation of the concept of a refinement is that a better sensor provides the agent with more information about the environment[5]. Each sensor mapping h induces a partition of X *via* its preimages, and refinement applies in the usual set-theoretic sense to the partitions when comparing sensors mappings. If a sensor mapping h is a refinement of h', then it enables the agent to react in a more refined way to nuances in the environment. Using the partial ordering given by refinements, we obtain the *sensor lattice* (LaValle, 2019).

By a referee's request, let us give a couple of biological examples.

Example 4.13 (First biological example)**.** There is an accepted theory that primates see red color wavelength, because it enables them to distinguish ripe fruit from non-ripe. Assuming this theory is true, it is an example of a refinement which is to some extent "minimal" and to some extent "sufficient" (of course strictly speaking it is neither – in the same way as there is no ideal circle in the physical world). The minimality is seen in this example, because we perceive other things as red, even if it is completely unnecessary (certainly unnecessary to tell the ripeness of fruits). So we are not distinguishing "too much." On the other hand, perceiving red color is a refinement of ripe/non-ripe which is only detected through stomach ache after the fruit has been already consumed. And it is sufficient in the sense that it is predictive of the original "base" labeling (ripe/non-ripe).

Example 4.14 (Second biological example)**.** Where our eyes look depends on the position of our head as well as the position of our eyes. Despite this, "looking up" (or "left," "right" etc..) are not ambiguous, even though these can be achieved with virtually infinitely many different head-eye configurations. One way to understand how this invariance could emerge is through minimal sufficient refinements. Suppose at birth, every head-eye configuration is considered as a separate state, but we label them by what we see in any given (stable) situation. A minimal sufficient refinement of that labeling will never distinguish between different states in which the eyes are pointing in the same direction. So then, by learning the minimal sufficient refinements, the agent may learn eye-direction invariance.

4.3. Lattice of sufficient equivalence relations

Please refer to Appendix A in the Supplementary material for notations and definitions used in this section.

We will prove in this section that if (X, U, τ) is an automaton, the sufficient equivalence relations form a complete sublattice of $(\mathcal{E}(X), \subseteq)$. Given an automaton $\mathcal{X} = (X, U, \tau)$, denote by $\mathcal{E}_{\text{suf}}^{U,\tau}(X) \subseteq \mathcal{E}(X)$ the set of sufficient equivalence relations on X. When U and τ are clear from the context, we write just $\mathcal{E}_{\text{suf}}(X) = \mathcal{E}_{\text{suf}}^{U,\tau}(X)$.

Theorem 4.15. *Suppose (X, U, τ) is an automaton and suppose that $\mathcal{E} \subseteq \mathcal{E}_{\text{suf}}(X)$ is a set of sufficient equivalence relations. Then $\bigwedge \mathcal{E}$ and $\bigvee \mathcal{E}$ are sufficient. Thus, $(\mathcal{E}_{\text{suf}}(X), \subseteq)$ is a complete sublattice of $(\mathcal{E}(X), \subseteq)$.*

Proof: See Appendix B. □

5 Here we are not talking about contentful or semantic information, but merely about correlational information in the philosophical sense.

Suppose that a labeling h is very important for an agent. For example, h could be "death or life," or it could be relevant for a robot's task. Suppose that h is not sufficient. The robot may want to find a sufficient refinement of h. Clearly a one-to-one h' would do. However, assume that the agent has to use resources for distinguishing between states; thus, the fewer distinctions the better. This motivates the following definition. Recall Definition 4.12 of refinements.

Definition 4.16. Let (X, U, T) be a transition system and $E_0 \subseteq X \times X$ an equivalence relation. *A minimal sufficient refinement* of E_0 is a sufficient equivalence relation E which is a refinement of E such that there is no sufficient E' with $E_0 \leqslant_r E' <_r E$.

Given a labeling h_0 of a transition system (X, U, T), a *minimal sufficient refinement* of h_0 is a labeling h such that E^h is a minimal sufficient refinement of E^{h_0} (recall Definition 2.34).

Example 4.17. Let $\mathcal{X} = (X, U, \tau)$ be an automaton where $X = \{0,1\}^*$, $U = \{0,1\}$ and $\tau(x, b) = x^\frown b$ (concatenation of the binary string x with the bit b). Let $h(x) = 1$ if and only if the number of ones and the number of zeros in x are both prime; otherwise $h(x) = 0$. Then the only sufficient refinements of h are one-to-one.

Example 4.18. Let $\mathcal{X}$ be as above and let $h\colon X \to \{0,1\}$ be such that if $|x|$ is divisible by 3, then $h(x) = 1$; otherwise, $h(x) = 0$. Then h is not sufficient. Let $h'\colon x \mapsto \{0,1,2\}$ be such that

$$h'(x) \equiv |x| \mod 3.$$

Then h' is a minimal sufficient refinement of σ.

Theorem 4.19. *Consider an automaton $\mathcal{X} = (X, U, \tau)$ and let E_0 be an equivalence relation on X. Then a minimal sufficient refinement of E_0 exists and is unique.*

Proof: See Appendix B □

Theorem 4.19 fails, if "automaton" is replaced by "transition system," or if "sufficient" is replaced by "n-sufficient" for $n > 1$ (recall Definition 4.7)

Example 4.20 (Failure of uniqueness for n-sufficiency)**.** Let $X = \{0,1,2,3,4,5\}$, $U = \{u_0\}$ and

$$\tau(0,u_0) = 1, \tau(1,u_0) = 2, \tau(2,u_0) = 2,$$

and

$$\tau(3,u_0) = 4, \tau(4,u_0) = 5, \tau(5,u_0) = 5.$$

Let E_0 be an equivalence relation on X such that the equivalence classes are $\{0,1,3,4\}$, $\{2\}$ and $\{5\}$. Then this relation is not 2-sufficient, because $(0, u_0, 1, u_0, 2)$ and $(3, u_0, 4, u_0, 5)$ are $(T, E_0, 2)$-equivalent, but 2 and 5 are not E_0-equivalent. Let $E_1, E_2 \subseteq E_0$ be equivalence relations with equivalence classes as follows:

$$E_1 :\{0,1\},\{3,4\},\{2\},\{5\},$$

$$E_2 :\{0,4\},\{1,3\},\{2\},\{5\}.$$

Then E_1 and E_2 are refinements of E_0. They are both 2-sufficient, because there doesn't exist any $(T, E_1, 1)$ or $(T, E_2, 1)$ equivalent 2-chains. They are also both $\leqslant_r$-minimal with this property which can be seen from the fact that they are actually $\leqslant_r$-minimal refinements of E_0 as equivalence relations (not only as sufficient ones).

Example 4.21 (Failure of uniqueness for transition systems)**.** Let $X = \{0,1,2,3,4\}$, $U = \{u_0\}$ and $T = \{(0, u_0, 3), (2, u_0), 4\}$. Let E_0 be the equivalence relation with the equivalence classes $\{0,1,2\}$, $\{3\}$ and $\{4\}$. Then E_0 is not sufficient, because $(0,2) \in E_0$, but $(3,4) \notin E_0$. Let E_1 and E_2 be the refinements of E_0 with the following equivalence classes:

$$E_1 :\{0,1\},\{2\},\{3\},\{4\},$$

$$E_2 :\{0\},\{1,2\},\{3\},\{4\}.$$

Now it is easy to see that both E_1 and E_2 are sufficient refinements of E_0, and by a similar argument as in Example 4.20 they are both minimal. The reason why this is possible is the odd behavior of the state 2 which doesn't have out-going connections. Such odd states are the reason why the decision problem "Does there exist a sufficient refinement with k equivalence classes?" is NP-complete for finite transition systems (O'Kane and Shell, 2017).

Remark. It is worth noting that Theorems 4.15 and 4.19 do not assume anything about the cardinality of X or of U, other structure on them (such as metric or topology) nor anything about the function τ or the relation E_0. Keeping in mind potential applications in robotics, X and U could be, for instance, topological manifolds, and τ a continuous function, or X could be a closed subset of $\mathbb{R}^n$, U discrete and τ a measurable function, or any other combination of those. In each of those cases, the sublattice of sufficient equivalence relations is complete, as per Theorem 4.15, and every equivalence relation E_0 on X admits a unique minimal sufficient refinement as per Theorem 4.19.

Recall Definition 2.10 of an equivalence relation preserving function. We say that an equivalence relation E on X is *closed under* $f\colon X \to X$ if for all $x \in X$, we have $(x, f(x)) \in E$. If E is closed under f, then f is E-preserving: given $(x, x') \in E$, we have $(x, f(x)), (x', f(x')) \in E$, because E is closed under f. Now by transitivity of E we have $(f(x), f(x')) \in E$, so f is E-preserving.

Definition 4.22. Let $f\colon X \to X$ be a bijection. The induced *orbit equivalence relation* is the relation E_f on X defined by $(x, x') \in E_f \iff (\exists n \in \mathbb{Z})(f^n(x) = x')$, in which $f^n(x)$ is defined by induction as: $f^0(x) = x$, $f^{n+1}(x) = f(f^n(x))$, $f^{n-1}(x) = f^{-1}(f^n(x))$.

Theorem 4.23. *If f is an automorphism of the automaton (X, U, τ), then E_f is a sufficient equivalence relation.*

Proof: See Appendix B □

Theorem 4.24. *Let $\mathcal{X} = (X, U, \tau)$ be an automaton and E be an equivalence relation on X. Suppose $f: X \to X$ is an automorphism such that E is closed under f. Let E' is the minimal sufficient refinement of E. Then E' is closed under f and $E \leqslant_r E' \leqslant_r E_f$.*

Proof: See Appendix B □

Example 4.25. Consider the environment which is a one-dimensional lattice of length five, $E = \{-2, -1, 0, 1, 2\}$, in which the corners "smell bad"; thus, we have a sensor mapping $h: E \to S$, $S = \{0, 1\}$ defined by $h(n) = 0 \iff |n| = 2$; see Figure 6A. Consider two agents in this environment. Both are equipped with the same h sensor, but their action repertoires differ. Both have two possible actions. One has actions L = "move left one space" and R = "move right one space," and the other one has actions T = "turn 180 degrees" and F = "go forward one space." Let $M_0 = \{L, R\}$ and $M_1 = \{T, F\}$. Thus, these agents have a slight difference in embodiment. Although both of them can move to every square of the lattice in a very similar way (almost indistinguishable from the outside perspective), we will see that the differences in embodiment will be reflected in that the minimal sufficient refinements will produce non-equivalent "categorizations" of the environment. The structures that emerge from these two embodiments will be different. These agents *enact* different environments, although physically the environments are the same, as congruent with tenet (EA3).

First, we define the SM-systems that model these agents' embodiments in E. The first agent does not have orientation. It can be in one of the five states, and the state space is $X_0 = E$ (Figure 6B). For the second agent, the effect of the F action depends on the orientation of the agent (pointing left or pointing right). Thus, there are ten different states the agent can be in, yielding $X_1 = E \times \{-1, 1\}$ (Figure 6C). The effects of motor outputs are specified completely (L means moving left, and so on), whereas the agent's internal mechanisms are left completely open, so our systems will be quasifilters. According to Remark 2.18, we can work with a labeled automaton instead. Hence, let $\tau_0: X_0 \times M_0 \to X_0$ be defined by $\tau_0(x, L) = \max(x - 1, -2)$ and $\tau_0(x, R) = \min(x + 1, 2)$. For the other agent, let $\tau_1((x, b), T) = (x, -b)$ and $\tau_1((x, b), F) = (\min(\max(x + b, -2), 2), b)$. Now we have labeled automata $\mathcal{X}_0 = (X_0, M_0, \tau_0, h, S)$ and $\mathcal{X}_1 = (X_1, M_1, \tau_1, h, S)$.

It is not hard to see that the one-to-one map $h_0: X_0 \to \{-2, -1, 0, 1, 2\}$ with $h_0(x) = x$ is a sufficient refinement of h which is minimal (see Figure 7A). Thus, every state needs to be distinguished by the agent for it to be possible to determine the following sensation from the current one. The derived information space automaton $\mathcal{X}_0/h_0$ isomorphic to $\mathcal{X}_0$ (Proposition 2.36).

For the second automaton, consider the labeling $h_1: X_1 \to \{-2, -1, 0, 1, 2\}$ defined by $h_1(x, b) = b \cdot x$ (see Figure 7C).

Claim. h_1 is a minimal sufficient refinement of h in $\mathcal{X}_1$.

Proof: See Appendix B □

Both minimal sufficient labelings, h_0 and h_1 have five values; thus, they categorize the environment into five distinct state-types. However, the resulting derived information spaces are different in the sense that the quotients $\mathcal{X}_0/h_0$ and $\mathcal{X}_1/h_1$ are not isomorphic; compare Figure 7B with Figure 7D.

Example 4.26. Figure 8A shows a filtering example from Tovar et al. (2014). More complex versions have been studied more recently in O'Kane and Shell (2017), and are found through automaton minimization algorithms and some extensions. It can be shown that this example's four-state derived information space depicted on Figure 8B corresponds to the unique minimal sufficient refinement of the labeling that only distinguishes between "are in the same region" and "are not in the same region." To see this, first note that this labeling is sufficient (since it can be represented as an automaton, this follows from Theorem 4.5). It follows from Theorem 4.19 that if this labeling is not minimal, then there is a minimal one which is strictly coarser, and so can be obtained by merging the states in the automaton of Figure 8B. This is impossible: the state T cannot be merged with anything because it violates the base-labeling; if, say D_a and D_c, are merged, then transition a will lead to inconsistency as it can lead either to D_b (from D_c) or to T (from D_a). This proves that this derived information space is indeed minimal sufficient, and by Corollary 4.19 there are no others up to isomorphism.

4.4. Computing sufficient refinements

This section sketches some computational problems and presents computed examples. The problem of computing the minimal sufficient refinement in some cases reduces to classical deterministic finite automaton (DFA) minimization, and in other cases it becomes NP-hard (O'Kane and Shell, 2017).

Consider an automaton (X, M, τ) and a labeling function h_0, and the corresponding labeled automaton described using the quintuple (X, M, τ, h_0, L). Suppose that the automaton (X, M, τ) corresponds to that of an body-environment system. Hence, X corresponds to the states of this coupled system. Suppose h_0 is not sufficient and consider the problem of computing a (minimal) sufficient refinement of h_0, that is, the coarsest refinement of h_0 that is sufficient.

Despite the uniqueness of the minimal sufficient refinement of h_0 (by Corollary 4.19), we can argue that the formulation of the problem, in particular, the input, can differ based on the level at which we are addressing the problem (for example, global perspective, agent perspective or something in between). Since the labeled automaton corresponding to an agent-environment coupling is described from a global perspective, the input to an

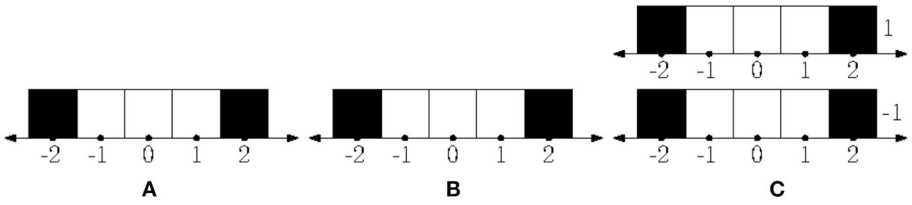

FIGURE 6
(A) One-dimensional lattice environment described in Example 4.25. **(B)** State space of the agent 0. **(C)** State space of the agent 1. The states for which the value of the sensor mapping is 0 are shown in black.

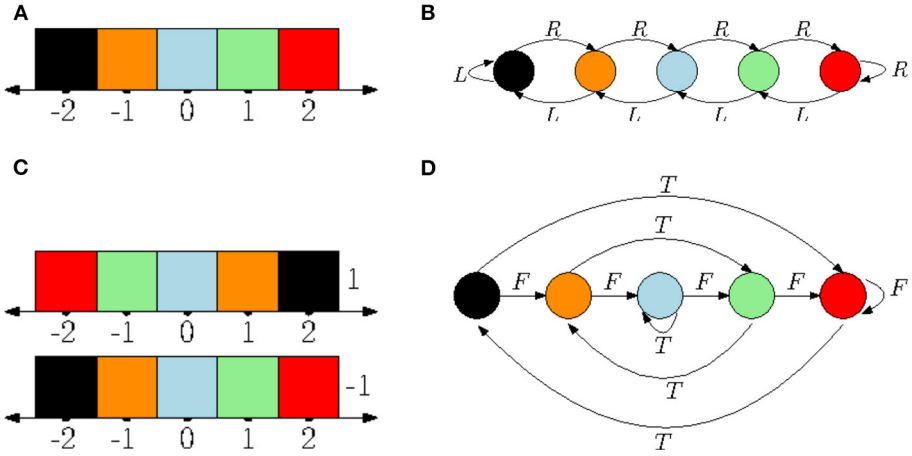

FIGURE 7
(A) State space of the agent 0 categorized by h_0, states that belong to the same class are colored with the same color. **(B)** Resulting quotient $\mathcal{X}_0/h_0$ for the agent 0. **(C)** State space of the agent 1 categorized by h_1, states that belong to the same class are colored with the same color. **(D)** Resulting quotient $\mathcal{X}_1/h_1$ for the agent 1.

algorithm that addresses the problem from this perspective is the labeled automaton $\mathcal{A} = (X, \tau, M, h_0, L)$ itself. Then, the problem is defined as given $\mathcal{A}$ compute $\mathcal{A}' = (X, M, \tau, h, L)$ such that h is the minimal sufficient refinement of h_0.

A special case of this problem from the global perspective occurs if the preimages of h_0 partition X in two classes which can be interpreted as the "accept" and "reject" states, for example, goal states at which the agent accomplishes a task and others. Furthermore, suppose that the initial state of the agent is known to be some $x_0 \in X$. Then, computing a minimal sufficient refinement becomes identical to minimization of a finite automaton, that is, given a DFA (X, M, τ, x_0, F) in which x_0 is the initial state and F is the set of accept states find (X', M, τ', x_0', F') such that no DFA with fewer states recognizes the same language. Existing algorithms, for example Hopcroft (1971), can be used to compute a minimal automaton.

Here, we also consider this problem from the agent's perspective for which the information about the environment states is obtained through its sensors, more generally, through a labeling function. Note that by agent's perspective we do not necessarily imply that the agent is the one making the computation (or any computation) but it means that no further information can be gathered regarding the environment other than the actions taken and what is sensed by the agent. At this level we address the following problem; given a set M of actions, a domain X, and a labeling function h_0 defined on X, compute the minimal sufficient refinement of h_0. The crux of the problem is that unlike the global perspective described above, the labeled automaton $\mathcal{A}$ is not given, in particular, the state transitions are not known a priory. Instead, the information regarding the state transitions can only be obtained locally by means of applying actions and observing the outcomes, that is, through sensorimotor interactions. Hence, the current body-environment state is also not observable. To show that an algorithm exists to compute a sufficient refinement of h_0 at this level, we propose an iterative algorithm (Algorithm 1) that explores X through agent's actions and sensations by keeping the history information state, that is, the history of actions and sensations (labels). We then show, by empirical results, that the

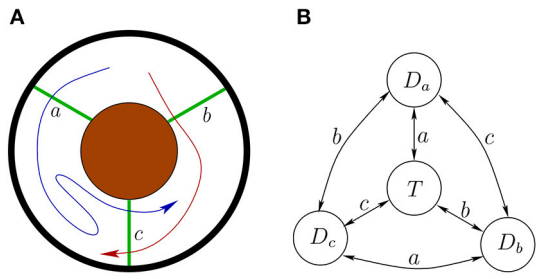

FIGURE 8
(A) Two point-sized independent bodies move along continuous paths in an annulus-shaped region in the plane. There are three sensor beams, *a*, *b*, and *c*. When each is crossed by a body, its corresponding symbol is observed. Based on receiving a string of observations, the task is to determine whether the two bodies are together in the same region, with no beam separating them. **(B)** The minimal filter as a transition system has only 4 states: T means that they are together, and D_x means that are in different regions but beam x separates them. Each transition is triggered by the observation when a body crosses a beam.

sufficient refinement computed by Algorithm 1 is minimal for the selected problem.

```
1: Input: h_0, l_0, M
2: Initialize: H ← ∅, h ← h_0, s ← s_0
3: for each step do
4:    m ← policy(s)
5:    apply action m and obtain resulting s'
6:    add (s, m, s') to H
7:    if ∃(s, m, s'') ∈ H such that s' ≠ s'' then
8:       h ← split(h, s)
9:    if there are labels that can be merged then
10:      h ← merge(h, H, h_0)
11:   s ← s'
```

Algorithm 1.

The functioning of Algorithm 1 is as follows. Starting from an initial sensation $s_0 = h(x_0)$, the agent moves by taking an action[6] given by the mapping policy$: L \rightarrow M$. Particularly, we used a fixed policy which samples an action m from a uniform distribution over M for each $s \in S$. In principle, any policy that ensures all states that are reachable from x_0 will be visited infinitely often should be enough. The history information state is implemented as a list, denoted by H, of triples (s, m, s') such that $s = h(x)$ and $s' = h(x')$ in which $x' = \tau(x, m)$. At each step, it is checked whether the current sensation is consistent with the history (Line 7). Current sensation is inconsistent with the history if there exists a triple (s, m, s'') in the history such that $s' \neq s''$. If it is not consistent then the label is split, which means that $h^{-1}(s)$ is partitioned into two parts P and Q. In particular, we apply a balanced random partitioning, that is, we select P and Q randomly from a uniform distribution over the partitions of $h^{-1}(s)$ that have two elements with balanced cardinalities. The labeling function is updated by a split operation as

$$h(x) := \begin{cases} s_Q & \text{if } x \in Q \\ s_P & \text{if } x \in P \\ h(x), & \text{otherwise.} \end{cases}$$

Recall that labels or subscripts do not carry any meaning from the agent's perspective.

Even a trivial strategy that splits the preimage of the label seen at each step would succeed computing a sufficient refinement. However, this would result in h being a one-to-one mapping. Hence, the finest possible refinement. Splitting only at the instances when an inconsistency is detected might reach a coarser refinement that is sufficient but there might be more equivalence classes than the ones induced by the minimal sufficient refinement of h_0. Therefore, a merge operation is introduced (Line 10). Let s and s' be two distinct labels for which $\exists s'' \in h_0[X]$ such that $h^{-1}(s) \subseteq h_0^{-1}(s'')$ and $h^{-1}(s') \subseteq h_0^{-1}(s'')$. Let t denote a triple in H and let t_k, $k = 1, 2, 3$, denote the k^{th} element of that triple. Suppose $s' = s$, if there are at least N number of triples in H such that for each triple t, $(t_1, t_2) = (s, m)$ and $\forall m \in M$ and $\forall t, t' \in H$ such that $(t_1, t_2) = (t'_1, t'_2) = (s, m)$ it is true that $t_3 = t'_3$ then labels s and s' are merged. The merge procedure goes through all labels and updates h as

$$h(x) := \begin{cases} s & \text{if } h(x) \in \{s, s'\} \\ h(x) & \text{otherwise.} \end{cases}$$

for each pair of labels s and s' that satisfies the aforementioned condition. Note that in principle, one can merge two labels regardless of the number of occurrences in the history. However, we noticed that this can result in oscillatory behaviour between split and merge operations especially for states that are reached less frequently. At present, we considered N as a tunable parameter and we know that it depends on the cardinality of the state space X such that larger the number of states, larger N should be. The problem of defining N as a function of the problem description remains open.

In the following, we present an illustrative example to show the practical implications of the previously introduced concepts in Section 4.2. In particular, we show how a simple algorithm like Algorithm 1 can be used by a computing unit which relies only on the sensorimotor interactions of an agent to further categorize the environment such that there are no inconsistencies in terms of the actions taken by the agent and the resulting sensations with respect to an initial categorization induced by h_0 (Figure 9C).

6 This can either be in a real environment or in a realistic simulation.

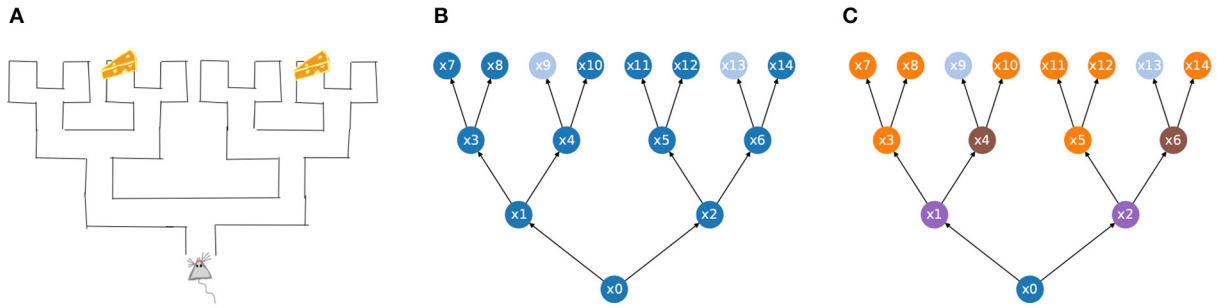

FIGURE 9
(A) Cheese maze defined in Example 4.27 **(B)** Labeled automaton with initial labeling h_0 corresponding to the cheese-maze example. **(C)** Minimal sufficient refinement of h_0. Self-loops at the leaf nodes are not shown in the figure.

Example 4.27. Consider an agent (a mouse) that is placed in a maze where certain paths lead to cheese and others do not (see Figure 9A). At each intersection the agent can go either left or right and it can not go back. Hence, at each step the agent takes one of the two actions; go right or go left. Figure 9B shows the corresponding automaton with 15 states describing the agent-environment system together with the initial labeling h_0 that partitions the state space into states in which the agent has reached a cheese (light blue) and others (dark blue). The initial state x_0 is when the agent is at the entrance of the maze. Once the end of the maze is reached (a leaf node) the state does not change regardless of which action is taken. After a predetermined number of steps the system reverts back to the initial state, similar to an episode in the reinforcement learning terminology (see, for example, Sutton and Barto, 2018). However, despite the system going back to the initial state the history information state still includes the prior actions and sensations. Figure 10 reports the updates of h, initialized at h_0, by Algorithm 1 being run for 1,000 steps. It converged to a final labeling h (Figure 10R), that is the minimal sufficient refinement of h_0, in 435 steps. For 20 initializations of Algorithm 1 for the same problem, on average, it took 364 steps to converge to a minimal sufficient refinement of h_0 (Figure 9C).

We have also applied the same algorithm to variations of this example with different depths of maze and different number of cheese and cheese placements (varying h_0). Empirical evidence shows that the same algorithm was capable of consistently finding the minimal sufficient refinement of the initial labeling. However, it is likely that it might fail for more complicated problems, for example, when the number of actions are significantly larger. It remains an open problem finding a provably correct algorithm for computing the minimal sufficient refinement of h_0 from the agent's perspective.

4.5. Sufficiency for coupled SM-systems

Section 2 introduced SM-systems, including the special class of quasifilters. We showed that quasifilters can be thought of as labeled transition systems, and we worked with such systems in Sections 4, 4.4. Let us see how do the concepts introduced in those sections work for SM-systems. We also defined *coupling* of SM-systems (Definition 2.22), but we have not defined what it means for a coupling to be "good." We will use sufficiency to approach this subject.

Let $\mathcal{E} = (E, (S \times M), T)$ and $\mathcal{I} = (I, S \times M, B)$ be SM-systems. We think intuitively of $\mathcal{E}$ as "the environment" and $\mathcal{I}$ as the "agent," even though they share the set of sensorimotor parameters $S \times M$. When is the coupling $\mathcal{E} * \mathcal{I}$ "successful"? Given another $\mathcal{I}' = (I', S \times M, B')$, how can we compare $\mathcal{I}$ and $\mathcal{I}'$ in the context of $\mathcal{E}$? The coupled system $\mathcal{E} * \mathcal{I}$ is not labeled; therefore, we cannot apply the definition of sufficiency. However, as soon as we apply some labeling to it, we can. There are many different ways to do it, intuitively corresponding to the "agent's perspective," the "environment's perspective" and a "god's perspective" (or "global perpsective").

The first one is the labeling $h\colon E \times I \to I$, which is the projection to the right coordinate, $h_I(e, i) = i$. The second one is the projection to the left coordinate $h_E(e, i) = i$, and the third one is the labeling of states by themselves, $h_G(e, i) = (e, i)$. Clearly, h_G is a refinement of both h_E and h_I. Yet another option is to use the sensory data as labelings, which is a coarser labeling than h_I. Or perhaps there was already a labeling $h\colon E \to S$ to begin with, so then we can ask about the property of $\hat{h}\colon E \times I \to S$ defined by $\hat{h}(e, i) = h(e)$. We focus on what we called the agent's perspective, h_I, for the rest of this section.

Recall Definition 4.11 of the degree of insufficiency. Given SM-systems $\mathcal{E}$ (environment) and $\mathcal{I}$ (agent), we can ask what is the degree of insufficiency of h_I in $\mathcal{E} * \mathcal{I}$? The smaller the

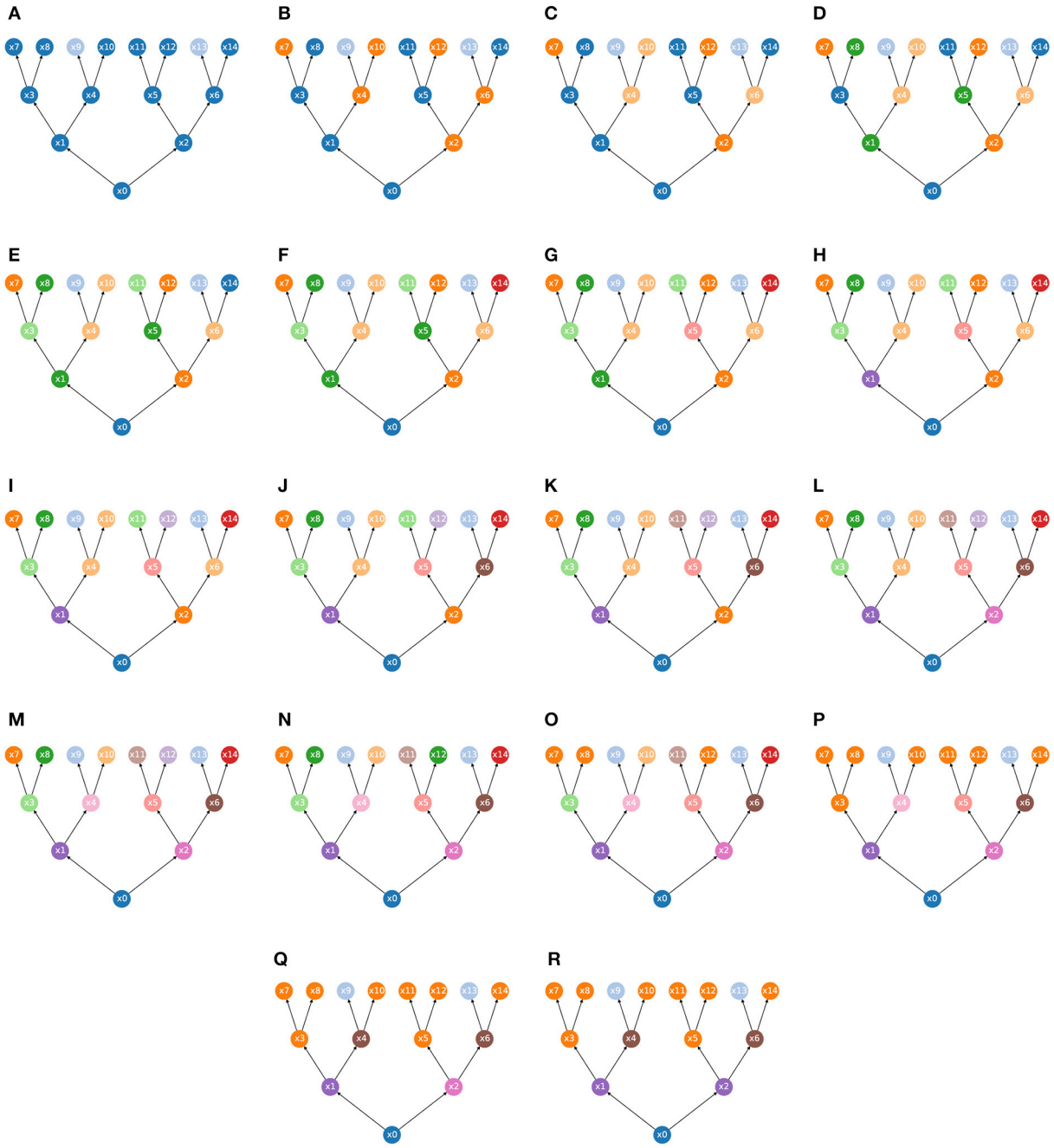

FIGURE 10
(A) Labeled automaton with labefigure/ling function $h = h_0$; same colored states belong to the figure/same equivalence class. **(B–Q)** Updating h by Algorithm 1 through splitting and merging of the labels. **(R)** Labeled automaton with the labeling function h that is the minimal sufficient refinement of h_0.

degree, the better the agent is attuned to the environment. This says something about the way in which the agent is adapted or attuned to the environment without attributing contentful states or representations to the agent in alignment with (EA2) and (EA4).

Let $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{I}'$ be SM-systems. When is degins($\mathcal{E} * \mathcal{I}, h_I$) < degins($\mathcal{E} * \mathcal{I}', h_{I'}$)? Of course, if $\mathcal{I}$ is fully constrained (Definition 2.31), then degins($\mathcal{E} * \mathcal{I}$) = ∞. This corresponds to the agent never engaging in any sensorimotor interaction with the environment. No wonder that it can always "predict" the

result of such passive existence. Assume, however, that there some constraints on the coupling. For example, we may demand that the agent must regularly visit states of some particular type to survive. Subject to such constrains, what can we say about degins($\mathcal{E} * \mathcal{I}$)? This seems to be a good preliminary notion[7] of attunement.

5. Discussion

In the introduction we defined our basic enactivist tenets:

(EA1) Embodiment and the inseparability of the brain-body-environment system,
(EA2) Grounding in sensorimotor interaction patterns, not in contentful representations.
(EA3) Emergence from embodiment, enactment of the world,
(EA4) Attunement, adaptation, and skill as possibilities to reliably engage in complicated patterns of activity with the environment.
(EA5) Perception as sensorimotor skills.

We developed a model of sensorimotor systems and coupling for which the purpose is to account for cognition mathematically, but in congruence with the principles (EA1)–(EA5). The principle (EA1) is intrinsic in the ways SM-systems are supposed to model brain-body and body-environment dynamics. The central ingredient is the control set $S \times M$ in all of those systems which include "motor" and "sensory" part; it is *impossible* in our framework to model say the environment without acknowledging the way in which the body is *part of* it. The approach that the actions of an agent depend solely on the history of its sensorimotor interactions with the environment, our approach is well in the scope of (EA2). We do not assume any representational or symbolic content possessed by the SM-systems. We do not evaluate them normatively by the "correctness" of their internal states, but rather by the ways in which they are, or can be, coupled to the environment and whether their sensory apparatus generates a sufficient sensor mapping or not. Coupling of SM-systems is defined so that two systems constrain each other. Thus, when an agent is coupled to the environment, they constrain each other, thereby creating new global properties of the body-environment system.

The principle (EA4) is mostly discussed in connection with minimal sufficient refinements. Given a labeling, or a categorization, or an equivalence relation on the state space, one can ask how well does this labeling "predict itself." The interpretation of this labeling can be anything from a sensor mapping to the labeling of environmental states by the internal states of the agent which coincide with them (this is not representation, this is mere co-occurence; see enactivist interpretation of the place cells in Hutto and Myin (2017) for comparison). A sufficient sensor mapping can be achieved in many different ways. In Section 4.4 we present a way in which the agent "develops" new sensors to be better attuned to the environment and in that way finds a sufficient sensor mapping. Another way for the agent would be to learn to act in a way that excludes "unpredictability." Both are examples of situations where the agent "structures" its own body-environment reality and gains skill. Finally, perception (EA5) can be understood as sensorimotor patterns on a microlevel. On the other hand, the agent engage in a sensorimotor activity locally without making big moves, such as moving the eyes without moving the body. The result of such sensorimotor interaction is another labeling function on a macro level.

In this paper, we not only presented mathematical definitions, but proved a number of propositions and theorems about them. There would be (and we hope there will be!) much more of them, but they did not fit in this expository work for which the main purpose was to demonstrate the connection of the mathematics in question with the enactive philosophy of mind.

We have already developed more concepts and theorems on top of this framework, including notions of *degree of insufficiency*, *universal covers*, *hierarchies*, and *strategic sufficiency*, but these are omitted here due to space limitations.

In other, more mathematical work, we plan to concentrate on working out mathematical and logical details of the proposed theory as well as applying the ideas to fundamental questions in robotics and autonomous systems, control theory, machine learning, and artificial intelligence.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

VW, BS, and SL developed the mathematical theory together over the past 2 years after extensive collaborative sessions. The primary author is VW, who wrote the most among the authors. VW contributed more to mathematical proofs. BS contributed more to computation. In addition to individual contributions, SL also played a supervisory role. All authors contributed to writing.

Funding

This work was supported by a European Research Council Advanced Grant (ERC AdG, ILLUSIVE: Foundations of Perception Engineering, 101020977), Academy of Finland (projects PERCEPT 322637, CHiMP 342556), and Business

7 Further research will indicate how much of this will be accepted by the most radical enactivists.

Finland (project HUMOR 3656/31/2019). All authors are with the Center for Ubiquitous Computing, Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland.

Acknowledgments

The author VW wishes to thank Dr. Otto Lappi for numerous discussion concerning enactivist and other agency which helped in shaping many of the ideas of this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2022.846982/full#supplementary-material

References

Başar, T., and Olsder, G. J. (1995). *Dynamic Noncooperative Game Theory, 2nd Edn*. London: Academic.

Bertsekas, D. P. (2001). *Dynamic Programming and Optimal Control, Vol. I, 2nd Edn*. Belmont, MA: Athena Scientific.

Blum, M., and Kozen, D. (1978). "On the power of the compass (or, why mazes are easier to search than graphs)," in *Proceedings Annual Symposium on Foundations of Computer Science* (Ann Arbor, MI), 132–142.

Choset, H., Lynch, K. M., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L. E., et al. (2005). *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge, MA: MIT Press.

Donald, B. R. (1995). On information invariants in robotics. *Artif. Intell. J.* 72, 217–304. doi: 10.1016/0004-3702(94)00024-U

Donald, B. R., and Jennings, J. (1991). "Sensor interpretation and task-directed planning using perceptual equivalence classes," in *Proceedings 1991 IEEE International Conference on Robotics and Automation* (Sacramento, CA: IEEE), 190–197.

Erdmann, M. A. (1993). Randomization for robot tasks: using dynamic programming in the space of knowledge states. *Algorithmica* 10, 248–291. doi: 10.1007/BF01891842

Fodor, J. (2008). *LOT 2: The Language of Thought Revisited*. Oxford: OUP Oxford.

Fraigniaud, P., Ilcinkas, D., Peer, G., Pelc, A., and Peleg, D. (2005). Graph exploration by a finite automaton. *Theor. Comput. Sci.* 345, 331–344. doi: 10.1016/j.tcs.2005.07.014

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Fuchs, T. (2020). The circularity of the embodied mind. *Front. Psychol.* 11, 1707. doi: 10.3389/fpsyg.2020.01707

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.

Gallagher, S. (2018). *Decentering the Brain: Embodied Cognition and the Critique of Neurocentrism and Narrow-Minded Philosophy of Mind*. Wollongong, NSW: Constructivist Foundations.

Gallistel, C. R., and King, A. (2009). *Memory and the Computational Brain*. Malden, MA: Wiley-Blackwell.

Ghallab, M., Nau, D., and Traverso, P. (2004). *Automated Planning: Theory and Practice*. San Francisco, CA: Morgan Kaufman.

Goranko, V., and Otto, M. (2007). "5 model theory of modal logic," in *Handbook of Modal Logic, volume 3 of Studies in Logic and Practical Reasoning*, eds P. Blackburn, J. Van Benthem, and F. Wolter (Cambridge: Elsevier), 249–329.

Hager, G. D. (1990). *Task-Directed Sensor Fusion and Planning*. Boston, MA: Kluwer.

Hopcroft, J. (1971). "An n log n algorithm for minimizing states in a finite automaton," in *Theory of Machines and Computations* (Haifa: Elsevier), 189–196.

Hutto, D. D., and Myin, E. (2012). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press.

Hutto, D. D., and Myin, E. (2017). *Evolving Enactivism*. Cambridge, MA: MIT Press.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artif. Intell. J.* 101, 99–134. doi: 10.1016/S0004-3702(98)00023-X

Kechris, A. S. (1994). *Classical Descriptive Set Theory, Vol. 156*. Verlag: Springer-Verlag; Graduate Texts in Mathematics.

Kumar, P. R., and Varaiya, P. (1986). *Stochastic Systems*. Englewood Cliffs, NJ: Prentice-Hall.

LaValle, S. M. (2006). *Planning Algorithms*. Cambridge, U.K: Cambridge University Press.

LaValle, S. M. (2012). *Sensing and Filtering: A Fresh Perspective Based on Preimages and Information Spaces, volume 1, 4 of Foundations and Trends in Robotics Series*. Delft: Now Publishers.

LaValle, S. M. (2019). "Sensor lattices: structures for comparing information feedback," in *2019 12th International Workshop on Robot Motion and Control (RoMoCo)* (Poznan: IEEE), 239–246.

Lozano-Pérez, T., Mason, M. T., and Taylor, R. H. (1984). Automatic synthesis of fine-motion strategies for robots. *Int. J. Rob. Res.* 3, 3–24. doi: 10.1177/027836498400300101

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

O'Kane, J. M., and LaValle, S. M. (2008). Comparing the power of robots. *Int. J. Rob. Res.* 27, 5–23. doi: 10.1177/0278364907082096

O'Kane, J. M., and Shell, D. A. (2017). Concise planning and filtering: hardness and algorithms. *IEEE Trans. Autom. Sci. Eng.* 14, 1666–1681. doi: 10.1109/TASE.2017.2701648

O'Regan, J. K., and Block, N. (2012). Discussion of j. kevin o'regan's "why red doesn't sound like a bell: understanding the feel of consciousness". *Rev. Philos. Psychol.* 3, 89–108. doi: 10.1007/s13164-012-0090-7

O'Regan, J. K., and Noëe, A. (2004, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/S0140525X01000115

Paolo, E. A. D. (2018). "The enactive conception of life," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. de Bruin, and S. Gallagher (Oxford: Oxford University Press), 71–95.

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Front. Psychol.* 2, 5. doi: 10.3389/fpsyg.2011.00005

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rescorla, M. (2016). Bayesian sensorimotor psychology. *Mind Lang.* 31, 3–36. doi: 10.1111/mila.12093

Roy, N., and Gordon, G. (2003). "Exponential family PCA for belief compression in POMDPs," in *Proceedings Neural Information Processing Systems* Vancouver, BC.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge, U.K: Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Techn. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1952). "Presentation of a maze-solving machine," in *Transaction of the 8th Cybernetics Conference* (Josiah Macy, Jr., Foundation), 173–180.

Suomalainen, M., Nilles, A. Q., and LaValle, S. M. (2020). "Virtual reality for robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Las Vegas, NV: IEEE), 11458–11465.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*. Cambridge, MA: MIT Press.

Tovar, B., Cohen, F., Bobadilla, L., Czarnowski, J., and LaValle, S. M. (2014). Combinatorial filters: sensor beams, obstacles, and possible paths. *ACM Trans. Sens. Networks* 10, 2594767. doi: 10.1145/2594767

Tschacher, W., and Dauwalder, J. P. (2003). *The Dynamical Systems Approach to Cognition: Concepts and Empirical Paradigms Based on Self-organization, Embodiment, and Coordination Dynamics*. Singapore: World Scientific.

Varela, F., Rosch, E., and Thompson, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.