The Imagination Machine II: Systems

Mark Tracy
Boston University
mrktracy@bu.edu

Introduction

This paper is the second part of a four-paper series titled *The Imagination Machine*. The first paper, *The Imagination Machine I: A View from Somewhere*, develops a formal epistemic framework for embedded observers and introduces the inference–implication loop that defines self-consistent world models. Within that framework, observations, classifiers, and world models all appear as structures internal to the same universe, and epistemic coherence arises as the closure of a representational loop rather than correspondence with an external vantage point.

The present paper develops a complementary layer of the project by introducing a general formalism for systems. Whereas the first paper describes the structure of representational closure for embedded epistemic systems, the present work describes the dynamical coupling between components of such systems, particularly in the case of agent–environment interaction. The goal is to define systems in an extremely general way so that the formalism has maximal expressiveness while making minimal assumptions.

In the first section, we develop a general definition of a system in terms of measurable variables, stochastic processes, and functional relations between inputs and outputs. In the following section we introduce optimization models and relate them to systems through the problem of system identification. The agent–environment framework developed later in the paper provides a general structure for modeling adaptive systems whose outputs influence the environment from which future inputs arise.

A minimal computational setting in which such agent–environment dynamics give rise to the recovery of relational invariants is developed in *The Imagination Machine III: Toy Model of Predictive Classification*. The final paper in the series, *The Imagination Machine IV: Institutional Intelligence*, extends the analysis further by examining how epistemic processes are stabilized and transmitted across communities and institutions.

1 General System Definition[1]

Define a set of variables that can be measured in practice. By necessity, this will be a countable set of variables, even if the underlying real system of interest has uncountable degrees of freedom. By measuring these variables over a set of time points I with minimal value t_0, we collect *data*. We distinguish between two different proper subsets of variables: *input variables* and *output variables*.

1.1 Input Variables

Without loss of generality, we will assume going forward that there is only one input variable. This is without loss of generality because any countable set of input variables may be represented by a tuple whose components are the simpler variables.

We represent the input variables with a random process. In particular, let $(\Omega_u, \mathcal{F}_u, \mathbb{P}_u)$ denote a probability space. Let U_{in} be the set of possible values the input variable may take.

Then the input process

$$u : \Omega_u \times I \to U_{\text{in}}$$

is measurable as a function from $(\Omega_u \times I, \mathcal{F}_u \otimes \mathcal{F}_I)$ to $(U_{\text{in}}, \mathcal{F}_{\text{in}})$, where $\mathcal{F}_I$ denotes a σ-algebra on the time set I, $\mathcal{F}_{\text{in}}$ denotes a σ-algebra on the input space U_{in}, and where the symbol $\otimes$ denotes the product σ-algebra. For each fixed time $t \in I$, the mapping

$$\omega \mapsto u(\omega, t)$$

is a random variable, and for every measurable set $A \subseteq U_{\text{in}}$ (i.e. $A \in \mathcal{F}_{\text{in}}$), the distribution of $u(t)$ is given by:

$$\mathbb{P}_u(\{\omega \in \Omega_u \mid u(\omega, t) \in A\})$$

We note, crucially, that although we are representing our input variable as a random process, input variables are often chosen to be those that one can deliberately vary over time. In such a case, the input variable may not be stochastic. In general, the input variable $u(t)$ at any time $t \in I$ may be a tuple in a product space of simpler, potentially degenerate random variables.

1.2 Output Variables

Complementary to the input variables are the output variables. Again, we will treat the case of a single output variable, since countably many output variables may be treated as a tuple.

Similarly to the input variable, we can represent the output variable as a random process. As before, let $(\Omega_y, \mathcal{F}_y, \mathbb{P}_y)$ denote a probability space. Let U_{out} be the set of possible values the output variable may take.

[1]Some language and structure adapted from *Introduction to Discrete Event Systems: Third Edition* by Christos G. Cassandras and Stéphane Lafortune. https://doi.org/10.1007/978-3-030-72274-6

Then the output process

$$y : \Omega_y \times I \to U_{\text{out}}$$

is measurable as a function from $(\Omega_y \times I, \mathcal{F}_y \otimes \mathcal{F}_I)$ to $(U_{\text{out}}, \mathcal{F}_{\text{out}})$, where $\mathcal{F}_{\text{out}}$ denotes a σ-algebra on the output space U_{out}, and where the symbol $\otimes$ denotes the product σ-algebra. For each fixed time $t \in I$, the mapping

$$\omega \mapsto y(\omega, t)$$

is a random variable, and for every measurable set $A \subseteq U_{\text{out}}$ (i.e. $A \in \mathcal{F}_{\text{out}}$), the distribution of $y(t)$ is given by:

$$\mathbb{P}_y(\{\omega \in \Omega_y \mid y(\omega, t) \in A\})$$

1.3 Relating Inputs to Outputs

The relation between the input variable and time, and the resulting output variable, is given by a functional g. A functional is a function whose domain is a Cartesian product of one or more sets of functions and zero or more other sets. In this case, the domain of the functional g is the Cartesian product of the set $\mathcal{U}$ of all measurable input processes $u : \Omega_u \times I \to U_{\text{in}}$ and the time set I. The codomain of g is $\mathcal{P}(U_{\text{out}}, \mathcal{F}_{\text{out}})$, the space of probability measures over the measurable space $(U_{\text{out}}, \mathcal{F}_{\text{out}})$. Explicitly:

$$g : \mathcal{U} \times I \to \mathcal{P}(U_{\text{out}}, \mathcal{F}_{\text{out}}),$$

The functional g satisfies:

$$y(t) \sim g[u, t]$$

Or, if modeling time as discrete, where t_{i+1} is a successor of t_i in a countable and strictly ordered time set I whose minimal element is t_0:

$$y(t_{i+1}) \sim g[u, t_i]$$

The symbol $\sim$ denotes random sampling or should be read as "is distributed according to." If the distribution is degenerate (i.e. there is no stochasticity), then the symbol may be treated as deterministic assignment, identically to an "equals" sign. In other words, determinism is represented as stochasticity with a degenerate distribution—assigning probability 1 to a single outcome.

Note that each component of the output variable (when considering a tuple of simpler variables) at time t can depend in general on the value of any component of the input variable (again, when considering a tuple of simpler variables) at any subset of time points, potentially including future points. While many physical systems are assumed to be "causal" (outputs depend only on present and past inputs), the mathematical formulation permits non-causal dependencies, allowing flexibility in modeling retroactive influence.

The functional g relating input and time to output may be an evaluation functional, which directly evaluates an input variable at a given time point, e.g.:

$$y(t) = g[u, t] = u(t)$$

It may also be a function of such evaluation functionals, e.g.,

$$y(t) = g[u, t] = u(t) + 3u(t - 1.3) - 76.8u(t + 4)^2$$

1.4 State

While the above is a general description of any system, in many cases, especially where history and memory matter, we find it useful to model the system's internal condition explicitly. This internal condition is what we call the system's *state*, which we can represent as a random process defined over some probability space $(\Omega_s, \mathcal{F}_s, \mathbb{P}_s)$. Letting U_{state} be the set of values that the state may take, we can write:

$$s : \Omega_s \times I \to U_{\text{state}}$$

Again, we consider the state $s(t)$ at time t to be a single random variable without loss of generality, since the state variable may be a tuple in a product space of simpler, potentially degenerate (i.e. determinate) random variables.

The evolution of the state may be represented in continuous time by a stochastic differential equation:

$$\dot{s}(t) \sim f[u, s, t], \quad s(t_0) \sim s_0 \quad \text{for some } s_0 \in \mathcal{P}(U_{\text{state}}, \mathcal{F}_{\text{state}})$$

where $\mathcal{F}_{\text{state}}$ is a σ-algebra on U_{state} and where

$$f : \mathcal{U} \times \mathcal{S} \times I \to \mathcal{P}(U_{\text{change}}, \mathcal{F}_{\text{change}}),$$

for some set U_{change} whose elements represent rates of change of the state and for $\mathcal{F}_{\text{change}}$ a σ-algebra on U_{change}; and where we denote by $\mathcal{S}$ the space of all measurable state processes $s : \Omega_s \times I \to U_{\text{state}}$.

If modeling time as discrete rather than continuous, then we may represent state dynamics as an update rule:

$$s(t_{i+1}) - s(t_i) \sim f[u, s, t_{i+1}], \quad s(t_0) \sim s_0 \quad \text{for some } s_0 \in \mathcal{P}(U_{\text{state}}, \mathcal{F}_{\text{state}})$$

where, similarly to before,

$$f : \mathcal{U} \times \mathcal{S} \times I \to \mathcal{P}(U_{\text{change}}, \mathcal{F}_{\text{change}}),$$

for some set U_{change} whose elements represent changes in the state, and where t_{i+1} is a successor of t_i in a countable and strictly ordered time set I whose minimal element is t_0.

The relation between the input variable, the system state, and time, and the resulting output variables may then be expressed as a functional:

$$y(t) \sim g[u, s, t]$$

or, in discrete time, where t_{i+1} is a successor of t_i in a countable and strictly ordered time set I whose minimal element is t_0:

$$y(t_{i+1}) \sim g[u, s, t_i]$$

where

$$g : \mathcal{U} \times \mathcal{S} \times I \to \mathcal{P}(U_{\text{out}}, \mathcal{F}_{\text{out}}).$$

2 Optimization Models

A functional, like those discussed above, is a special kind of function. An optimization model is a function approximator. An optimization model consists of a triplet $(\mathcal{H}, O, A)$ of:

1. A hypothesis space $\mathcal{H}$ (a set of functions);
2. An objective $O : \mathcal{H} \to \mathbb{R}$ (a functional whose domain is the hypothesis space and whose range is real numbers) which gives some signal as to the quality of the approximation; and
3. An optimization algorithm $A : \mathcal{H} \to \mathcal{H}$ (a rule for moving through the hypothesis space), in general utilizing the objective.

When learning inductively from data (that is, when attempting to move from particular examples to general principles), a few additional objects may be appended to the aforementioned triplet; in particular:

4. A dataset $\mathcal{D}$.
5. A (possibly unknown) random process P from which data points are sampled. In other words, data is collected empirically from the world during a time interval I_D with $d_i \sim P(t_i) \quad \forall d_i \in D$, where data point d_i is collected at time t_i. Note that in cases where data points may be assumed to be identically distributed and drawn independently, this amounts to a single distribution. In *active learning*, the algorithm A interacts with the random process P, influencing the empirical dataset $\mathcal{D}$ used during optimization. In other words, data points are not sampled according to P before the commencement of the algorithm A, but rather, the process of data collection is itself influenced by the optimization algorithm.
6. A dataset $\mathcal{D}_{\text{aug}}$, where for all $d_{\text{aug}} \in \mathcal{D}_{\text{aug}}$, there exists a function f, an integer N, and a tuple of elements $t \in \mathcal{D}^N$ such that $d_{\text{aug}} \sim f(t)$, where $\sim$ denotes, as before, stochastic sampling of the (possibly degenerate) random function f. In other words, every element of $\mathcal{D}_{\text{aug}}$ is a (potentially stochastic) function of elements of $\mathcal{D}$.
7. A random process P_{train} by which elements of $\mathcal{D}_{\text{aug}}$ are drawn by the optimization algorithm. In particular, the algorithm A draws at time t_i an element $d_i \sim P_{\text{train}}(t_i)$, where $P_{\text{train}}(t_i)$ is a distribution over $\mathcal{D}_{\text{aug}}$.

An inductive bias is a constraint on the hypothesis space. By traversing the hypothesis space algorithmically, an optimization model is intended to minimize the objective function and thus obtain a good approximation to the function that truly represents the system of interest.

3 System Identification

System identification is the process of utilizing an optimization model to find an approximation to the true dynamics of a system using measurements of its input and output variables. In particular, it is useful when the internal state of a system is not known or its internal dynamics–the stochastic differential (or difference) equation(s) and initial conditions governing the state's trajectory–are not known.

4 Agents

The agent–environment coupling introduced here provides the dynamical structure within which the representational closures described in *The Imagination Machine I: A View from Somewhere* may arise for embedded epistemic systems. In that framework, stable world models emerge as fixed points of an inference–implication loop defined over observations internal to the same universe. The systems formalism developed here provides a concrete representation of the interacting processes through which such observations and models may be generated.

An agent necessarily exists within and is co-constituted with an environment. An agent-environment pair comprises two systems, an agent A and environment E, which are in interchange (feeding back to one another); as well as an initial input to either the agent or the environment. In particular, A takes as input the output of E, and E takes as input the output of A, with the recursion beginning from some set of initial inputs to either system.

Formally, we may represent the recursive dependency between an agent A and an environment E as follows:

$$\begin{aligned} u^A(t) &= y^E(t) \quad \text{(agent receives environment's output as input)} \\ u^E(t) &= y^A(t) \quad \text{(environment receives agent's output as input)} \end{aligned}$$

where:

- $u^A(t)$ is the input to the agent at time t
- $y^A(t)$ is the agent's output at time t
- $u^E(t)$ is the input to the environment at time t
- $y^E(t)$ is the environment's output at time t

The recursion begins from a set of initial inputs:

$$\begin{aligned} u^E(t_0) &\sim u_0^E \quad \text{for some } u_0^E \in \mathcal{P}(U_{\text{in}}^E, \mathcal{F}_{\text{in}}^E) \quad \text{or} \\ u^A(t_0) &\sim u_0^A \quad \text{for some } u_0^A \in \mathcal{P}(U_{\text{in}}^A, \mathcal{F}_{\text{in}}^A) \end{aligned}$$

and the pair evolves together over time, potentially governed by their own internal state dynamics:

$$\dot{s}^A(t) \sim f^A[u^A, s^A, t], \qquad y^A(t) \sim g^A[u^A, s^A, t]$$
$$\dot{s}^E(t) \sim f^E[u^E, s^E, t], \qquad y^E(t) \sim g^E[u^E, s^E, t]$$

for some functionals defined analogously as before:

$$f^A : \mathcal{U}^A \times \mathcal{S}^A \times I \to \mathcal{P}(U^A_{\text{change}}, \mathcal{F}^A_{\text{change}})$$
$$g^A : \mathcal{U}^A \times \mathcal{S}^A \times I \to \mathcal{P}(U^A_{\text{out}}, \mathcal{F}^A_{\text{out}})$$
$$f^E : \mathcal{U}^E \times \mathcal{S}^E \times I \to \mathcal{P}(U^E_{\text{change}}, \mathcal{F}^E_{\text{change}})$$
$$g^E : \mathcal{U}^E \times \mathcal{S}^E \times I \to \mathcal{P}(U^E_{\text{out}}, \mathcal{F}^E_{\text{out}})$$

That is, each functional takes as input:

- a random input process over I,
- a random state process over I,
- and the current time $t \in I$,

and produces either a rate of change of the state (for f) or an output (for g), potentially by sampling randomly from a distribution of outputs.

4.1 Agents in Discrete Time

In many practical applications, especially in reinforcement learning, the agent-environment interaction is modeled in discrete time. This leads to the following slight change in representation:

$$u^A(t_{i+1}) = y^E(t_i) \quad \text{(agent receives environment's most recent output as input)}$$
$$u^E(t_{i+1}) = y^A(t_i) \quad \text{(environment receives agent's most recent output as input)}$$

where t_{i+1} is the successor of t_i in some ordered set of time points I (in particular, the time points are indexed by $i \in \mathbb{N}_0$), and where

$$s^A(t_{i+1}) - s^A(t_i) \sim f^A[u^A, s^A, t_{i+1}], \qquad y^A(t_{i+1}) \sim g^A[u^A, s^A, t_{i+1}]$$
$$s^E(t_{i+1}) - s^E(t_i) \sim f^E[u^E, s^E, t_{i+1}], \qquad y^E(t_{i+1}) \sim g^E[u^E, s^E, t_{i+1}]$$

for some functionals defined analogously as before:

$$f^A : \mathcal{U}^A \times \mathcal{S}^A \times I \to \mathcal{P}(U^A_{\text{change}}, \mathcal{F}^A_{\text{change}})$$
$$g^A : \mathcal{U}^A \times \mathcal{S}^A \times I \to \mathcal{P}(U^A_{\text{out}}, \mathcal{F}^A_{\text{out}})$$
$$f^E : \mathcal{U}^E \times \mathcal{S}^E \times I \to \mathcal{P}(U^E_{\text{change}}, \mathcal{F}^E_{\text{change}})$$
$$g^E : \mathcal{U}^E \times \mathcal{S}^E \times I \to \mathcal{P}(U^E_{\text{out}}, \mathcal{F}^E_{\text{out}})$$

That is, each functional takes as input:

- a random input process over I,
- a random state process over I,
- and the current time $t \in I$,

and produces either a state update (for f) or an output (for g), potentially by sampling randomly from a distribution over possible outputs.

4.2 Reinforcement Learning

Reinforcement learning is a special case of an optimization model, whereby the objective O depends on the history of interactions between an agent and its environment and where the algorithm A seeks to maximize the expected cumulative reward obtained through the agent and environment's dynamic coupling.

In the context of the present series, such agent–environment optimization dynamics provide a concrete setting in which representational models may be iteratively refined through interaction with structured environments. A minimal predictive example of such refinement is developed in *The Imagination Machine III: Toy Model of Predictive Classification*.

5 Becoming-Held-As-By: Subjects as Systems in Self-Representation

In the language developed in *The Imagination Machine I: A View from Somewhere*, a self-representing subject is an embedded epistemic system whose classifiers appear within its own observation space. The condition that classifiers are themselves observations allows a system to encounter and revise its own acts of classification. The present section approaches the same idea from the perspective of systems modeling: if the formalism developed above can represent any system, then it must also apply to the system performing the representation.

If the above framework above provides insight into how to represent any real system in mathematical terms, then a natural next step is to turn the inquiry on the modeler. In other words, in writing the above formalism I am confronted with the question, "Am I not a real system myself? Can I, then, be understood in these terms?"

I imagine a bubble around my body, and then I imagine it shrinks inward all around and approaches infinitely closely to the edge of my skin. Any measurable passing between this membrane is either input or output–and thus I conceive of agent and environment.

Pursuant to these aims, we now shift from formal system representation to a philosophical and phenomenological inquiry into how a system may represent itself as an agent, co-constituted and co-evolving with an environment. In this way, we move from a formalism for modeling system behavior from an external perspective to a vocabulary by which a self-modeling agentic system may represent its own reality from the internal perspective.

5.1 A Self-Referential Thesis

All may be called the Becoming-Held-As-By[2] (including its becoming held as this by me).

5.2 Potentiality and Representation

Suppose we take "existence" to mean "the quality, state, or event of becoming-held-as-by." We use the word "potentiality" to mean that from which existence emerges through representation. Potentiality is metaphysical substance itself—what we might call the pre-conceptual whatever-I-represent. Representation is the process or result of becoming-held-as-by. A subject is becoming-held-as-by-itself.

For example, to say that a particular cup "exists" is to say that some potentiality (which I could, for example, point to) is becoming held in mind by me as a unified and distinct "thing" which I represent as a cup. If the potentiality does not become held as anything by any subject, then it cannot be said that anything in particular exists there, though there may persist some potentiality for becoming-held-as-by (held as a cup, perhaps, or as something else, like a weapon or a hat, by any particular subject). To hold potentiality as something is not to define it once and for all, but to engage in a relationship that may change. The same potentiality may be held as many different things across time, across subjects, or even within the same subject in different moments.

One cannot properly imagine potentiality because all one *can* do is imagine potentiality, in the sense of bringing potentiality into representation. That is to say that to imagine potentiality is already to bring it into representation. Potentiality may have internal structure (e.g. change relative to some internal reference frame according to laws). Regardless, here is the big picture: potentiality (metaphysical substance) is translated into existence (the ontological) through its representation by the subject (the semiological and epistemological: perception, language, systems of meaning, knowledge claims). By this notion of existence, if every conscious being were to disappear suddenly, there would not be a universe at all—only potential for a universe to arise.

Reality, in this account, is enacted through the interplay of potentiality and representation: a process in which potential becomes held through representation, and representation constrains potentiality.

5.3 The Subject Becoming-Held-As-Agent-By-Itself

The most stable world-model I have yet realized is this: world as constituted of agent (self) co-evolving with environment, where the agent's state includes its representation of self, environment, and world; including, recursively, a representation of world as constituted of agent (self) co-evolving with environment, where the agent's state includes its representation of self, environment, and world.

This is a world that I hold as constituted of the agent-environment coupling, wherein a subject may coherently and productively become-held-as-agent-by-itself. The agent is not

[2]The hyphenation of "Becoming-Held-As-By" is deliberate: it reflects the interdependence and co-constitution of the becoming, the holding, the *as*-ness (representation), and the *by*-ness (the subject).

separable from the environment, though it may be ontologically separate in its own representation. The state of the agent includes its representation of potentiality: It is influenced by its environment's output and its own history, and, in turn, it influences the input to the environment through the output of the agent. Because of the inherent coupling of agent to environment, the subject becoming-held-as-agent-by-itself is to the Becoming-Held-As-By as a *holon* to its greater whole[3]: the subject may become-held-as a distinct object of analysis by itself, and yet it can simultaneously become-held by itself as a part in a larger system.

To use a human-centric analogy, the subject becoming-held-as-agent-by-itself is to the Becoming-Held-As-By as the mouth is to the body: the mouth is not the body, yet it is interconnected with the body; and the declaration that "I am the body" is made by means of the mouth. Similarly, the subject becoming-held-as-agent-by-itself is not the entirety of the Becoming-Held-As-By and yet is embedded (and participating) within it; and the writing and reading of statements like, "All may be called the Becoming-Held-As-By (including its becoming held as this by me)" is enacted by the subject becoming-held-as-agent-by itself.

5.4 The Limits of the Systems Formalism

A system is defined by its distinctions: inputs vs. outputs, internal vs. external state. The undivided Whole—that which contains all systems, distinctions, and environments—cannot itself be represented as a system. Since it has no external relation and no boundary, it admits no input/output mapping. Likewise, the complement of the undivided Whole—that is, nothingness, or void—admits no input/output mapping and as such may not be represented as a system.

5.5 Mathematics as Meta-Representation

Mathematics may derive from the structure of representation itself. That is to say, mathematics is a type of meta-representation: a representation of common structure across instances of representation. Accordingly, the representation of mathematical objects could potentially be invariant under change in subject if each subject can in principle abstract from their own instances of representation to arrive at the same mathematical meta-representations. For example, I can map a notion of two-ness to the same symbol that another subject can map theirs, and we can be reasonably sure that we agree on its meaning, because we both experience unity and difference in our representations of self and world. Likewise, I can map a notion of a function to the same symbol that another can map theirs, and we can be reasonably sure that we agree on its meaning, because we both represent and abstract from instances of change. Unity, difference, and change may be necessary structures of subjective representation, such that any subject with sufficient abstract reasoning capability can attribute the same meaning to the same meta-representations.

[3]The term "holon" was coined by Arthur Koestler in his 1967 book, *The Ghost in the Machine*. A holon is both a self-contained entity (hence it is a whole on its own) and at the same time is embedded within a larger containing system or systems (so it is part of a larger whole).

5.6 Haecceity and Qualia

Complementary to the notion of meta-representation in this account is the notion of haecceity, or the irreducible *this*-ness of an entity. Haecceity is what remains in representation modulo meta-representation—the particularity that is not captured by abstraction from representation to meta-representation. For a human, haecceity may correspond to the irreducible qualia of the experience of being *this particular self* in *this particular moment*.

5.7 Truth and Coherence

A proposition is a linguistic claim that may be judged true or false by a subject. Truth is a judgment of coherence among a collection of propositions. Formally, a proposition is judged false if it is shown that the proposition, potentially together with a collection of propositions judged to be true, implies contradiction of a proposition judged true. Therefore, a particular proposition is judged true only by virtue of its ongoing coherence with a collection of mutually non-contradictory propositions. It must be emphasized that propositions involving instantiation (of abstract classes) are among the propositions that must be coherent in a collection of truths. For example, a proposition like, "An electron evolves according to the Schrodinger equation," must cohere with such propositions of instantiation as, "This reading (referring to a particular representation in experience) is due to an electron," and, "This reading (at another time, perhaps) is not due to an electron," as well as propositions that are not instantiations like, "An electron has negative charge." This understanding of truth allows for pluralism while requiring that a worldview be consistently tethered to moments of becoming-held-as-by.

5.8 Conclusion

The central claim is that we are always describing the world from the inside: embedded within the Becoming-Held-As-By and co-evolving with our environment, seeing patterns in our seeing-patterns. We conscious beings are individually and collectively a self-representing network of interacting holonic subsystems. And yet, on the whole and within each part, haecceity remains.