The Universal Property of Possibilistic Belief Updating

Nathaniel Virgo
University of Hertfordshire, UK
n.virgo@herts.ac.uk

Matteo Capucci
University of Strathclyde, UK
Independent Researcher, IT
matteo.capucci@gmail.com

Manuel Baltieri
Araya Inc., Japan
manuel_baltieri@araya.org

Martin Biehl
Cross Labs,
Cross Compass, Japan
martin.biehl@cross-compass.com

We are interested in control problems with hidden state, where one system must interact with another in order to keep their combined state within some predefined set of "good" states. In such a setting it is natural to ask questions like "what does the controller know about the state of its environment?". In previous work [VBBC25] we showed how to make such questions formal, modelling controller X and environment Y as Moore and Mealy machines respectively. We have shown that under minimal hypotheses any successful controller admits a "semantic" map to the set $\mathcal{P}Y$ of subsets of the state space of the environment, thought of as "beliefs", and that this map commutes with a dynamics given by "possibilistic Bayesian conditioning".

Here we show that $\mathcal{P}Y$, together with the possibilistic updating procedure, generalises to a system (in the sense of categorical systems theory) with the universal property of a power object with respect to a suitable fibration of predicates.

Categorical systems theory [JM21, Mye22, LM25] is a framework for reasoning about open systems and their algebra of couplings. In its basic form, it concerns 'systems with interfaces' and ways to define coupling of systems at their interfaces, in such a way that both systems and interfaces form structured categories known as *theories of systems*. This allows compositional reasoning to be combined with category-theoretic reasoning.

In particular, the dynamics of systems can be described functorially, by providing a map of theories of systems targeting a so-called *behavioural theory*. A typical behavioural theory is constructed starting from a category $\mathbf{C}$ whose objects are seen as spaces of behaviours. The term 'behaviour' is very general, since behaviours can be used for many purposes, but for us it will roughly mean 'trajectory', as in a possible trajectory of a dynamical system. Often $\mathbf{C}$ will be a presheaf topos or a topos of sheaves on some space, allowing us to talk about the passage of time. We will also consider the case where $\mathbf{C} = \mathbf{Cat}$, the 1-category of small strict categories, with the interpretation that objects are states and morphisms are possible trajectories between states. This allows us to consider composition of trajectories and demonstrates that $\mathbf{C}$ need not be a topos.

For this extended abstract we don't need the full definition of a systems theory or a behavioural theory and can make do with this abridged fragment:

Definition 1. Given a category $\mathbf{C}$ with pullbacks and an object I of $\mathbf{C}$, the category of $\mathbf{C}$-*behavioural systems with interface* I is $\mathbf{C}_{/I}$.

Thus, a behavioural system with interface I is an object X of $\mathbf{C}$ equipped with a map $X \to I$ projecting out the behaviour of its interface. If we have systems X and Y with the same interface, we can couple them by forming the pullback of the cospan $X \to I \leftarrow Y$. The idea is that coupling the two systems constrains their behaviour since the two systems together can only undergo trajectories that agree on the interface. This idea has its origin in [Wil05], and appears in a categorical systems theory context in [SS19, SSV20, Lib20, Mye22].

Submitted to:
Applied Category Theory 2026 (extended abstract)

We want to consider control theory in this context. The relevant concepts from control theory can be expressed in terms of *forward-closed subsets* of the state space of some kind of dynamical system (e.g. Moore or Mealy machines). This is a subsets of states that trajectories cannot leave once entered (but trajectories can enter the set from outside, in general). These can be modelled behaviourally by a suitable choice of fibration on the category $\mathbf{C}$.

With this in hand we can formulate the following notion of *control problem*: consider a (plant) system Y, equipped with a subobject $G \hookrightarrow Y$, which is to be seen as its set of "good" states. A candidate controller for such a system might be a system X with the same interface, so that it can be coupled to Y. Given such a system we can pull back G along the pullback projection $X \times_I Y \to Y$, and then look for the largest forward-closed subobject contained within the resulting subobject. If this is nonempty we have a witness that some states of X can regulate Y in the form of a forward-closed 'relation' $U \subseteq X \times_I Y$: this means there are states of the coupled system such that if the system is initialised in such a state, the state of Y will remain inside G indefinitely.

In [VBBC25] we showed, in a much more concrete setting, that if a system X is able to control a system Y in this sense, then X can be seen as having 'beliefs' about Y, in the form of a map $\psi : X \to PY$, where PY is an object of subsets of Y with Bayesian-like deterministic dynamics which we defined ad-hoc. We show that, in this more abstract setting, PY enjoys a very natural universal property which seals its role as 'system of beliefs about Y'. This property is that of being a *power object* for the fibration of forward-closed predicates:

Definition 2. Let $F : \mathbf{E} \to \mathbf{B}$ be a fibration, such that the base category $\mathbf{B}$ has products. Then, given an object Y in $\mathbf{B}$, an *F-power object for Y* is an object $\mathcal{P}Y$ in $\mathbf{B}$ together with an object $\in_Y$ in E over $\mathcal{P}Y$, with the following universal property:

$$\begin{array}{c} \forall X : \mathbf{Set},\ U \subseteq X \times Y, \quad \exists!\, \psi : X \to \mathcal{P}Y \\ \text{such that there is a cartesian morphism} \end{array} \qquad \begin{array}{ccc} U & \xrightarrow{\text{cartesian}} & \in_Y \\ X \times Y & \xrightarrow{\psi \times \mathrm{id}_Y} & \mathcal{P}Y \times Y\,. \end{array} \tag{1}$$

In each of our examples, we can take $\mathbf{B}$ to be the category $\mathbf{C}_{/I}$ of systems with a given interface, and F the fibration of forward-closed subsets of their state spaces. Thus, this formal correspondence captures that between 'beliefs' $X \to PY$ and witnesses U that X solves a control problem with respect to environment Y. In each of our settings, all F-power objects exist.

This observation becomes interesting when we move down the ladder of abstraction and see what the power objects are in the example theories. In the talk we will show that they have the flavour of a map from states of X to subsets of the state space of Y, which *update over time* in a way that makes sense for belief states: informally, an updated belief state contains all the states of Y that are possible given the information received by X. Indeed, the 'possibilistic Bayesian updating with forgetting' of [VBBC25] arises as a special case. Thus we recover and generalise the result that for a successful controller belief semantics exist automatically. We believe this is an important step towards a general theory of agency, which may find applications in neuroscience, biology and machine learning as well as computer science and control theory.

Acknowledgments

The second author is funded by Advanced Research + Invention Agency grant MSAI-PRO-01. The other three authors are funded by the Advanced Research + Invention Agency (ARIA) through project code MSAI-SE01-P011

References

[JM21] David Jaz Myers (2021): *Double Categories of Open Dynamical Systems (Extended Abstract)*. *Electronic Proceedings in Theoretical Computer Science* 333, p. 154–167, doi:10.4204/eptcs.333.11. Available at http://dx.doi.org/10.4204/EPTCS.333.11.

[Lib20] Sophie Libkind (2020): *An Algebra of Resource Sharing Machines*. arXiv:2007.14442.

[LM25] Sophie Libkind & David Jaz Myers (2025): *Towards a double operadic theory of systems*. arXiv:2505.18329.

[Mye22] David Jaz Myers (2022): *Categorical Systems Theory*. Unpublished book draft. Available at http://davidjaz.com/Papers/DynamicalBook.pdf.

[SS19] Patrick Schultz & David I. Spivak (2019): *Temporal Type Theory: A Topos-Theoretic Approach to Systems and Behavior*. Springer International Publishing, doi:10.1007/978-3-030-00704-1. Available at http://dx.doi.org/10.1007/978-3-030-00704-1.

[SSV20] Patrick Schultz, David I. Spivak & Christina Vasilakopoulou (2020): *Dynamical Systems and Sheaves*. *Applied Categorical Structures* 28(1), pp. 1–57, doi:10.1007/s10485-019-09565-x.

[VBBC25] Nathaniel Virgo, Martin Biehl, Manuel Baltieri & Matteo Capucci (2025): *A "Good Regulator Theorem" for Embodied Agents*. In: *ALIFE 2025: Ciphers of Life: Proceedings of the Artificial Life Conference 2025*, *isal2025* 37, MIT Press, pp. 419–428, doi:10.1162/isal.a.874. Available at http://dx.doi.org/10.1162/ISAL.a.874.

[Wil05] Jan C Willems (2005): *The behavioral approach to systems theory*. In: *Proceedings of the Conference on New Approaches in System and Control Theory, Istanbul, Turkey*, pp. 59–71.