Journal of Machine Learning Research 26 (2025) 1-37 Submitted 2/24; Revised 3/25; Published 6/25

Categorical Semantics of Compositional Reinforcement Learning

Georgios Bakirtzis BAKIRTZIS@TELECOM-PARIS.FR
LTCI, Télécom Paris, Institut Polytechnique de Paris

Michail Savvas MICHAIL-SAVVAS@UIOWA.EDU
The University of Iowa

Ufuk Topcu UTOPCU@UTEXAS.EDU
The University of Texas at Austin

Editor: Aryeh Kontorovich

Abstract

Compositional knowledge representations in reinforcement learning (RL) facilitate modular, interpretable, and safe task specifications. However, generating compositional models requires the characterization of minimal assumptions for the robustness of the compositionality feature, especially in the case of functional decompositions. Using a categorical point of view, we develop a knowledge representation framework for a *compositional theory* of RL. Our approach relies on the theoretical study of the category MDP, whose objects are Markov decision processes (MDPs) acting as models of tasks. The categorical semantics models the compositionality of tasks through the application of *pushout* operations akin to combining puzzle pieces. As a practical application of these pushout operations, we introduce *zig-zag* diagrams that rely on the compositional guarantees engendered by the category MDP. We further prove that properties of the category MDP unify concepts, such as enforcing safety requirements and exploiting symmetries, generalizing previous abstraction theories for RL.

Keywords: universal properties, Markov decision processes, category theory

1. Introduction

Verification of reinforcement learning (RL) systems is essential for mastering increasingly complex tasks, ensuring reliable and trustworthy behavior, and mitigating the risk of hazardous control actions in dynamic environments (Jothimurugan et al., 2021; Gur et al., 2021; Li et al., 2021; Szabó, 2012). The construction of compositional theories for Markov decision processes (MDPs) verifies the overall behavior of an RL system by parts, enabling the segmentation into manageable components. Exploiting the compositionality feature in MDPs suggests that regardless of the data or flexibility of the RL algorithms we deploy, it is beneficial to model using modular task specifications and precisely define the structural features of sequential decision-making problems such that the engineered artifact is interpretable.

In particular, functional compositionality is essential to engendering increasingly adaptable systems by providing an interpretable language for specifying tasks. However, methods that functionally compose problem formulations are scarce (Mendez et al., 2022). Functional compositionality *predicts* a system's behavior by the combination of behaviors deriving from its parts and the combination *preserves* properties emerging from the parts (Genovese, 2018).

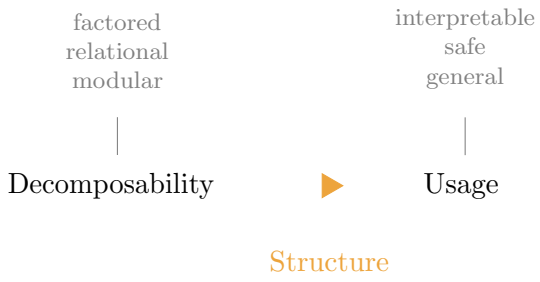


Figure 1: Structure improves interpretability, safety, and generalization (adapted from Mohan et al. (2024)).

To extend this line of research, we construct a unifying compositional theory for engineering RL systems that mechanizes functional composition into subprocess behaviors.

To achieve this explicit definition of compositionality in RL, we give a symbolic and semantic interpretation of compositional phenomena of the problem or task by translating them into categorical properties. Category theory is well-suited for posing and solving *structural* problems. As such, categorical constructions, like *pullbacks* and *pushouts*, describe general operations that model (de)compositions of objects, thereby giving formal meaning to clusters of objects as a composite and how they interface with each other. This formalization is not merely a theoretical exercise; it can improve RL systems' interpretability, safety, and generalization capabilities. Interpretability is augmented as these categorical structures provide a systematic framework for understanding the relationships and interactions within an RL system. Safety is addressed through precisely controlling and predicting these interactions, especially in complex or composite systems where emergent behaviors might pose risks. Thus, through a compositional theory it is possible to enhance the core aspects of RL systems—making them interpretable, safer, and adaptable to diverse applications (Figure 1).

The notion of MDP is the structural instantiation of a general sequential decision-making problem (Ok et al., 2018). We study compositionality from the perspective of subprocess MDPs and examine compositional properties governing MDPs that result in universal interfaces. In particular, addressing task (de)composition requires operationalizing (Todorov, 2009) a given overall behavior. In this study, categorical semantics makes the types of MDP operations in different configurations explicit, extending the algebraic approach (Perny et al., 2005). While hierarchical MDPs also allow for operationalizing tasks in RL (Parr and Russell, 1997; Dietterich, 2000; Nachum et al., 2018; Wen et al., 2020; Klissarov et al., 2025), the categorical semantics defines the exorable and incorporeal forms of abstraction, refinement, and compositional operations through universal properties. Universal properties are particularly desirable for knowledge representations because they govern the definition of new admissible composite systems, thereby constructively producing modular components. In the compositional theory, categories define a precise *syntax* for MDP instantiations and assign *semantics* to particular compositional behaviors (Section 2).

We construct a categorical semantics for sequential task completion problems. The first superposes obstacles in a grid world using the operation of pullbacks (Section 3.1). The second

exploits the (potential) symmetric structure of RL problems using the operation of a pushout (Section 3.3). The third is a design of compositional task completion, where zig-zag diagrams synthesize the behavior of a fetch-and-place robot (Section 3.4). Zig-zag diagrams (Section 3.2) denote composite MDPs produced by gluing together subprocess MDPs, preserving the relationship between types of actions and states. While this paper aims to develop a *compositional theory* of RL, our constructions have a computational interpretation that partitions problem spaces and manages the increasing complexity of RL systems (Bakirtzis et al., 2024).

Conventions Subscripts for MDP elements refer to a particular instantiation of the definition. For example, $S_{\mathcal{N}}$ refers to the state space of the MDP $\mathcal{N}$. Certain preliminary concepts are introduced in Appendix A. Unless otherwise noted, we consider probability measures on spaces expressed by a probability density function with respect to a given ambient probability measure when clear from context.

2. A Compositional Theory of Reinforcement Learning

In this section, we introduce the unit MDP and the latent structure of the environments in this work. We then introduce various compositionality features of MDP structures through a theoretical study of categorical properties.

2.1 The Category MDP

Definition 1 (MDP) *An MDP $\mathcal{M} = (S, A, \psi, T, R)$ is a 5-tuple consisting of the following data:*

- *A measurable space S with a fixed (implicit) σ-algebra, called the state space of $\mathcal{M}$.*
- *A set A, called the state-action space of $\mathcal{M}$. This is the set of actions available at all different states of S.*
- *A function $\psi\colon A \to S$ that maps an action $a \in A$ to its associated state $s \in S$.*
- *A function $T\colon A \to \mathcal{P}_S$, where $\mathcal{P}_S$ denotes the space of probability measures on S. This is the information of the transition probabilities on $\mathcal{M}$.*
- *A reward function $R\colon A \to \mathbb{R}$.*

The actions an agent can take at a particular state s are given by the set $\psi^{-1}(s) \subseteq A$. We denote this set of available actions for each state s by A_s. Knowing the action spaces A_s for all $s \in S$, one may recover A and ψ by the assignment

$$A = \coprod_{s \in S} A_s, \ \psi\colon A_s \mapsto s \in S. \tag{1}$$

If the actions of the MDP are not state-specific and uniform across the whole state space S, then we can take $A = S \times A_0$ for a fixed set of actions A_0 and $\psi\colon A \to S$ is then just the projection onto the first factor.

One departure of the above definition from standard definitions in the literature is the added flexibility in allowing the actions to vary across the state space S. While many MDPs

are uniform and thus the state-action space is naturally a product $A = S \times A_0$ as in the previous remark, there are natural MDP environments where this uniformity fails. For example, an agent moving inside a maze will not be able to move in all four directions when faced with a wall or another obstacle, forming a boundary which can vary within the environment defining the state space.

Another possible simplification occurs when the MDP rewards only depend on the state and not the action taken. In that case, the reward function R factors through ψ and a state-dependent reward function $R_0 \colon S \to \mathbb{R}$, so that $R = R_0 \circ \psi$.

Having defined MDPs, we now proceed to define morphisms to obtain the corresponding category whose objects are MDPs (Definition 1).

Definition 2 (Category of MDPs) *MDPs form a category* MDP *whose morphisms are defined as follows:*

Let $\mathcal{M}_i = (S_i, A_i, \psi_i, T_i, R_i)$*, with* $i = 1, 2$*, be two MDPs.*

A morphism $m = (f, g) \colon \mathcal{M}_1 \to \mathcal{M}_2$ *is the data of a measurable function* $f \colon S_1 \to S_2$ *and a function* $g \colon A_1 \to A_2$ *satisfying the following compatibility conditions:*

1. *The diagram*

$$\begin{array}{ccc} A_1 & \xrightarrow{g} & A_2 \\ {\scriptstyle \psi_1}\downarrow & & \downarrow{\scriptstyle \psi_2} \\ S_1 & \xrightarrow[f]{} & S_2 \end{array} \tag{2}$$

   is commutative.

2. *The diagram*

$$\begin{array}{ccc} A_1 & \xrightarrow{g} & A_2 \\ {\scriptstyle T_1}\downarrow & & \downarrow{\scriptstyle T_2} \\ \mathcal{P}_{S_1} & \xrightarrow[f_*]{} & \mathcal{P}_{S_2} \end{array} \tag{3}$$

   is commutative, where f_* *maps a probability measure* $\mu_1 \in \mathcal{P}_{S_1}$ *to the pushforward measure* $\mu_2 = f_* \mu_1 \in \mathcal{P}_{S_2}$ *under* f*.*

3. *The diagram*

$$\begin{array}{ccc} A_1 & \xrightarrow{g} & A_2 \\ & {\scriptstyle R_1}\searrow & \downarrow{\scriptstyle R_2} \\ & & \mathbb{R} \end{array} \tag{4}$$

   is commutative.

The commutative diagrams above express the compatibility between two MDPs in the sense that their interfaces and rewards agree.

Namely, diagram (2) guarantees that if an action a_1 in MDP $\mathcal{M}_1$ is associated to a state $s_1 \in S_1$, then its image action $a_2 = g(a_1)$ under the morphism m is associated to the image state $s_2 = f(s_1)$.

Similarly, diagram (3) ensures that the transition probability from any state s_1 to the measurable subset $f^{-1}(U) \subseteq S_1$ after taking action a_1 in $\mathcal{M}_1$, where $U \subseteq S_2$ is any measurable subset, is equal to the transition probability from the state $s_2 = f(s_1)$ to $U \subseteq S_2$ under action $a_2 = g(a_1)$ in $\mathcal{M}_2$.

Finally, diagram (4) ensures that the reward obtained after taking action a_1 at the state s_1 in MDP $\mathcal{M}_1$ equals the reward obtained after taking action $a_2 = g(a_1)$ at the state $s_2 = f(s_1)$ in MDP $\mathcal{M}_2$.

Some immediate features of the category MDP are as follows.

Definition 3 (Empty and constant MDP) *The empty MDP $\emptyset$ is the MDP whose state space and action spaces are the empty set. The constant MDP pt is the MDP whose state space and action spaces are the one-point set.*

Remark 4 *The constant MDP is not unique, as the reward function value can be defined to be equal to any number. We refer to any such instance as the constant MDP.*

The category MDP has an initial object and a collection of partially terminal objects.

Proposition 5 *For every MDP $\mathcal{M}$, there exists a unique morphism $\emptyset \to \mathcal{M}$. In particular, $\emptyset$ is the initial object of MDP. Every MDP $\mathcal{M}$ with a constant reward function admits a unique, natural morphism $\mathcal{M} \to \mathsf{pt}$.*

Subprocesses There is also a natural definition of subobjects in MDP. In particular, the definition of morphisms can correctly capture the notion of a subprocess of an MDP.

Definition 6 (Subprocesses) *We say that the MDP $\mathcal{M}_1$ is a subprocess of the MDP $\mathcal{M}_2$ if there exists a morphism $m = (f, g) \colon \mathcal{M}_1 \to \mathcal{M}_2$ where f and g are injective, that is the morphism m is injective on state and state-action spaces.*

We say that a subprocess $\mathcal{M}_1$ is a full subprocess of $\mathcal{M}_2$ if diagram (2) is cartesian. Equivalently, this means that $A_1 = \psi_2^{-1}(f(S_1))$ as a subset of A_2.

Since f is injective, we may consider the state space S_1 as a subset of S_2. Moreover, the condition that diagram (2) is cartesian means that S_1 is closed under actions in $\mathcal{M}_2$ and the state-action spaces of $\mathcal{M}_1$ and $\mathcal{M}_2$ coincide over S_1. Thus, $\mathcal{M}_1$ being a full subprocess of $\mathcal{M}_2$ implies that an agent following the MDP $\mathcal{M}_2$ who finds themself at a state $s_1 \in S_1$ will remain within S_1 no matter which action $a_1 \in A_1$ they elect to take.

Conversely, for a MDP $\mathcal{M}_2$ and any subset $S_1 \subseteq S_2$ there is a canonical subprocess $\mathcal{M}_1$ with state space S_1, whose action space A_1 is defined as

$$A_1 \coloneqq \psi_2^{-1}(S_1) \cap T_2^{-1}(f_\star(\mathcal{P}_{S_1})). \tag{5}$$

In fact, $\mathcal{M}_1$ is uniquely characterized as the maximal such subprocess.

Proposition 7 *Any subprocess $\mathcal{M}_1' \to \mathcal{M}_2$ with state space S_1 factors uniquely through the subprocess $\mathcal{M}_1 \to \mathcal{M}_2$.*

Proof This follows from the fact that any action $a_2 \in (A_2)_{s_1}$ such that $T_2(a_2)$ is a measure supported on S_1 gives an element of A_1, as defined in equation (5). ■

In the ensuing subsections, we will investigate further, more complicated features of the category MDP. In particular, we will be interested in appropriate and universal constructions of products and "unions" of MDPs.

2.2 Pullbacks: Universal Interfaces

Interesting categorical properties are *universal*. Universal properties represent specific ideals of behavior within a defined category (Spivak, 2014; Asperti and Longo, 1991). A pullback (or fiber product) is a categorical generalization of the notion of cartesian product whose universal property is determined by it being maximal in a certain sense. Besides the cartesian product, another common example is the intersection $S_1 \cap S_2$ of two subsets $S_1, S_2 \subseteq S_3$, obtained as the pullback of the two inclusions of sets $S_1 \to S_3$ and $S_2 \to S_3$.

We prove that well-formed MDPs have partial pullbacks in the category MDP, meaning that there exists a way to compose the data of one MDP with another to create an interface between MDPs.

The intuition behind a pullback of MDPs is the idea of *intersecting* two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$ viewed as components of a third MDP $\mathcal{M}_3$ through morphisms $m_1 \colon \mathcal{M}_1 \to \mathcal{M}_3$ and $m_2 \colon \mathcal{M}_2 \to \mathcal{M}_3$. In general, the morphisms m_i do not need to define subprocesses and this allows for valuable flexibility, ranging from intersections to products. For example, the cartesian product $\mathcal{M}_1 \times \mathcal{M}_2$ of two MDPs will be obtained when $\mathcal{M}_3 = \mathsf{pt}$ is the constant MDP (Definition 3). In the other extreme, we can obtain intersections of subprocesses, as for the grid world problem, which will be studied later on in this paper.

Write $\mathcal{M}_i = (S_i, A_i, \psi_i, T_i, R_i)$ for $i = 1, 2, 3$. The structure of the state space of the pullback as a measure space requires some care in the construction and is the reason for the failure of the existence of a fully universal pullback. If this were indeed possible, one would then be able to deduce a construction of pullbacks for the category of measurable spaces, which is known not to exist.

However, we will succeed, allowing for the weakening of the universal properties. We will obtain the following theorem by examining progressively more general cases.

Theorem 8 *Let $\mathcal{M}_i = (S_i, A_i, \psi_i, T_i, R_i)$ for $i = 1, 2, 3$ be MDPs. There exists an MDP $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2$ with state space $S = S_1 \times_{S_3} S_2$ (defined in equation 6) and state-action space $A = A_1 \times_{A_3} A_2$ (defined in equation 7) which fits in a commutative diagram in MDP:*

$$\begin{array}{ccc} \mathcal{M} & \longrightarrow & \mathcal{M}_1 \\ \downarrow & & \downarrow{\scriptstyle m_1} \\ \mathcal{M}_2 & \xrightarrow[m_2]{} & \mathcal{M}_3 \end{array}$$

The MDP $\mathcal{M}$ is universal with respect to diagrams whose morphisms are conditionally independent (Definition 10).

We establish the theorem in steps, which we also discuss as we go.

Suppose that we have three MDPs $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ together with two morphisms $m_1 = (f_1, g_1)\colon \mathcal{M}_1 \to \mathcal{M}_3$ and $m_2 = (f_2, g_2)\colon \mathcal{M}_2 \to \mathcal{M}_3$.

We begin by defining the pullback's state, action spaces, and rewards and continue with the setup of the transition probabilities.

Construction of states, actions and, rewards For the state and action spaces, we set

$$S = S_1 \times_{S_3} S_2 = \{(s_1, s_2) \in S_1 \times S_2 \mid f_1(s_1) = f_2(s_2) \in S_3\}, \tag{6}$$
$$A = A_1 \times_{A_3} A_2 = \{(a_1, a_2) \in A_1 \times A_2 \mid g_1(a_1) = g_2(a_2) \in A_3\}. \tag{7}$$

By standard properties of pullbacks (of sets), the maps $\psi_i\colon A_i \to S_i$ for $i = 1, 2, 3$ induce a canonical morphism $\psi\colon A \to S$. Write $pr_i\colon S \to S_i$ and $\rho_i\colon A \to A_i$ for the projection maps, where $i = 1, 2$.

Since we need S to be a measurable space, we endow it with the σ-algebra generated by all subsets

$$pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2),$$

where U_1, U_2 are any two measurable subsets of S_1 and S_2 respectively. The projections pr_i are then tautologically measurable functions and this is the minimal choice of σ-algebra for which this holds.

For the reward function, we set $R = R_3 \circ g_1 \circ pr_1 = R_3 \circ g_2 \circ pr_2\colon A \to \mathbb{R}$.

This minimal σ-algebra on S is potentially small. However, when σ-algebras of S_1 and S_2 are their power sets (for example, in the case where S_1 and S_2 are finite or discrete measure spaces), then the σ-algebra is also the power set of S, so we do not view this potential smallness as an issue.

Transition probabilities Let now $a \in A$ with projections $a_1 \in A_1$ and $a_2 \in A_2$ mapping in turn to an action $a_3 \in A_3$. For brevity, write $\mu_i = T_i(a_i) \in \mathcal{P}_{S_i}$ for $i = 1, 2$ and $\mu_3 = (f_i)_\star \mu_i \in \mathcal{P}_{S_3}$.

Our goal is to construct $\mu = T(a) \in \mathcal{P}_S$ and obtain for each index $i = 1, 2$ a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\rho_i} & A_i \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle T_i} \\ \mathcal{P}_S & \xrightarrow[(pr_i)_\star]{} & \mathcal{P}_{S_i}. \end{array}$$

Having achieved that, we will get, by definition, an MDP $\mathcal{M}$ fitting in the commutative diagram

$$\begin{array}{ccc} \mathcal{M} & \xrightarrow{(pr_1, \rho_1)} & \mathcal{M}_1 \\ {\scriptstyle (pr_2, \rho_2)}\downarrow & & \downarrow{\scriptstyle m_1} \\ \mathcal{M}_2 & \xrightarrow[m_2]{} & \mathcal{M}_3. \end{array} \tag{8}$$

We would moreover like this diagram to be universal among commutative diagrams of the form

$$\begin{array}{ccc} \mathcal{N} & \xrightarrow{(\alpha_1,\beta_1)} & \mathcal{M}_1 \\ {\scriptstyle (\alpha_2,\beta_2)}\downarrow & & \downarrow{\scriptstyle m_1} \\ \mathcal{M}_2 & \xrightarrow[m_2]{} & \mathcal{M}_3, \end{array} \tag{9}$$

in the sense that any commutative diagram (9) should be induced by a canonical morphism $\mathcal{N} \to \mathcal{M}$ and composing with the projection maps $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$. This will not generally be possible and we will need to restrict our attention to pairs of morphisms $\mathcal{M}_1 \to \mathcal{M}_3,\ \mathcal{M}_2 \to \mathcal{M}_3$ satisfying a certain independence condition, as we will see below. For such pairs of morphisms, we can indeed achieve our goal.

We ask:

⟨Q1⟩ Under what conditions does μ exist?

⟨Q2⟩ If $\mu = T(a)$ exists for all $a \in A$, what kind of universality properties does $\mathcal{M}$ have?

We describe several instances in which the measure $\mu = T(a)$ can be constructed, in order of increasing generality. We begin with the case where one of the two morphisms is a subprocess, then move on to cases with discrete state spaces, and finally conclude with the most general case of local fibrations. In addition, we obtain a simple-to-state and clean universality statement in the subprocess and discrete cases using conditional independence.

The subprocess case Suppose that one of $\mathcal{M}_1 \to \mathcal{M}_3$ and $\mathcal{M}_2 \to \mathcal{M}_3$ is a subprocess. Assume this is the case for $\mathcal{M}_2 \to \mathcal{M}_3$ without loss of generality. By Definition 6, this implies that the maps $pr_1 \colon S \to S_1$ and $\rho_1 \colon A \to A_1$ are injective. Moreover, since $(f_1)_\star \mu_1 = (f_2)_\star \mu_2 = \mu_3$, it follows that μ_1 is supported on S. We define $T(a) = \mu_1$ as a measure on S.

Proposition 9 *The diagram* (8) *is universal among diagrams of the form* (9) *for which the morphism* $\mathcal{N} \to \mathcal{M}_1$ *defines a subprocess.*

Proof Suppose that $(\alpha_1, \beta_1) \colon \mathcal{N} \to \mathcal{M}_1$ defines a subprocess. We have induced morphisms $\alpha = (\alpha_1, \alpha_2) \colon S_{\mathcal{N}} \to S = S_1 \times_{S_3} S_2$ and $\beta = (\beta_1, \beta_2) \colon A_{\mathcal{N}} \to A = A_1 \times_{A_3} A_2$.

Both maps are injective since the compositions $pr_1 \circ \alpha = \alpha_1$ and $\rho_1 \circ \beta = \beta_1$ are injective.

For any $a \in A_{\mathcal{N}}$ we need to check that $\alpha_\star T_{\mathcal{N}}(a) = T(\beta(a))$.

But $T_1(\beta_1(a))$ is supported on $\mathcal{N}$ and we have $T_1(\beta_1(a)) = T(\beta(a))$.

On the other hand, $(pr_1)_\star \alpha_\star T_{\mathcal{N}}(a) = (pr_1 \circ \alpha)_\star T_{\mathcal{N}}(a) = (\alpha_1)_\star T_{\mathcal{N}}(a) = T_1(\beta_1(a))$, as wanted. ■

The discrete case Assume that the spaces S_1, S_2, S_3 are discrete or finite and their σ-algebras are their power sets.

Consider the function $\nu \colon S_1 \times_{S_3} S_2 \to \mathbb{R}$ defined by the formula

$$\nu(s_1, s_2) = \begin{cases} \frac{\mu_1(s_1)\mu_2(s_2)}{\mu_3(f_1(s_1))} = \frac{\mu_1(s_1)\mu_2(s_2)}{\mu_3(f_2(s_2))}, & \text{if } \mu_3(f_1(s_1)) = \mu_3(f_2(s_2)) > 0, \\ 0, & \text{if } \mu_3(f_1(s_1)) = \mu_3(f_2(s_2)) = 0. \end{cases} \tag{10}$$

We define the measure $\mu = T(a) \in \mathcal{P}_S$ to be the one with probability density function ν. Thus,

$$\mu\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) \coloneqq \sum_{(s_1,s_2)\in pr_1^{-1}(U_1)\cap pr_2^{-1}(U_2)} \nu(s_1, s_2). \tag{11}$$

This satisfies the desired properties to define MDP morphisms $\mathcal{M} \to \mathcal{M}_i$, $i = 1, 2$ (see Proposition 12 below). We will now show that it is also universal in an appropriate sense.

Definition 10 *We say that two morphisms $(\alpha_i, \beta_i)\colon \mathcal{N} \to \mathcal{M}_i$, $i = 1, 2$, in* MDP *form a* conditionally independent pair with respect to $\mathcal{M}_3$ *if for any $a \in A_{\mathcal{N}}$ with images $a_i = \beta_i(a) \in A_i$ and any measurable subsets $U_i \subseteq S_i$, we have*

$$T_{\mathcal{N}}(a)\left(\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)\right) = \sum_{(s_1,s_2)\in pr_1^{-1}(U_1)\cap pr_2^{-1}(U_2)} \nu(s_1, s_2). \tag{12}$$

This assignment determines the measure $T_{\mathcal{N}}$ only on the σ-algebra generated by sets $\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)$.

Our definition of conditional independence is consistent with the corresponding notion considered by Swart (1996). We can now prove a universality statement.

Proposition 11 *The morphisms $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$ in diagram* (8) *form a conditionally independent pair. Moreover, diagram* (8) *is universal among diagrams* (9) *where the morphisms $\mathcal{N} \to \mathcal{M}_i$ form a conditionally independent pair.*

Proof The independence of $\mathcal{M} \to \mathcal{M}_1$ and $\mathcal{M} \to \mathcal{M}_2$ follows immediately by observing that equations (11) and (12) coincide.

For the universality statement, suppose as above that we have two independent morphisms $(\alpha_i, \beta_i)\colon \mathcal{N} \to \mathcal{M}_i$, $i = 1, 2$, in MDP, that fit in a commutative square (9).

Since $\mathcal{M}$ has state and action spaces given by $S = S_1 \times_{S_3} S_2$ and $A = A_1 \times_{A_3} A_2$ respectively and the diagram gives that $f_1 \circ \alpha_1 = f_2 \circ \alpha_2$ and $g_1 \circ \beta_1 = g_2 \circ \beta_2$, we get canonical morphisms $\gamma\colon S_{\mathcal{N}} \to S$ and $\delta\colon A_{\mathcal{N}} \to A$. These satisfy

$$pr_i \circ \gamma = \alpha_i, \quad \rho_i \circ \delta = \beta_i, \quad i = 1, 2,$$

and moreover by construction $\gamma \circ \psi_{\mathcal{N}} = \psi \circ \delta$.

To obtain a morphism $\mathcal{N} \to \mathcal{M}$, it remains to check that the corresponding diagram (3) commutes, as compatibility of rewards is clear. For $a \in A_{\mathcal{N}}$, write $\beta_i(a) = \rho_i(\delta(a)) = a_i \in A_i$. We then have, using formula (12) for the third equality and formula (11) for the last equality,

$$\begin{aligned}
&\gamma_\star T_{\mathcal{N}}(a)\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) = \\
&= T_{\mathcal{N}}(a)\left(\gamma^{-1}(pr_1^{-1}(U_1)) \cap \gamma^{-1}(pr_2^{-1}(U_2))\right) = \\
&= T_{\mathcal{N}}(a)\left(\alpha_1^{-1}(U_1) \cap \alpha_2^{-1}(U_2)\right) = \\
&= \sum_{(s_1,s_2)\in pr_1^{-1}(U_1)\cap pr_2^{-1}(U_2)} \nu(s_1, s_2) = \\
&= T(\delta(a))\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right),
\end{aligned}$$

which implies that $\gamma_\star \circ T_{\mathcal{N}} = T \circ \delta$, completing the proof. ■

The case of S_3 finite Suppose that S_3 is finite with weights for the measure μ given by a density function ν_3.

Suppose in addition that S_1, S_2 are measurable subsets of ambient measure spaces $(K_1, \tau_1), (K_2, \tau_2)$ with density functions ν_1, ν_2 so that for any measurable subset $U_i \subseteq S_i$ we have $\mu_i(U_i) = \int_{U_i} \nu_i \, \mathrm{d}\tau_i$. This is not restrictive and we allow it for exhibition and consistency with the preceding case. One can always take ν_i to be identically 1 and then $\mu_i = \tau_i$.

As above, we consider the function $\nu \colon S_1 \times_{S_3} S_2 \to \mathbb{R}$,

$$\nu(s_1, s_2) = \begin{cases} \frac{\nu_1(s_1)\nu_2(s_2)}{\nu_3(f_1(s_1))} = \frac{\nu_1(s_1)\nu_2(s_2)}{\nu_3(f_2(s_2))}, & \text{if } \nu_3(f_1(s_1)) = \nu_3(f_2(s_2)) > 0 \\ 0, & \text{if } \nu_3(f_1(s_1)) = \nu_3(f_2(s_2)) = 0, \end{cases}$$

and define the measure $\mu = T(a) \in \mathcal{P}_S$ to be the one with probability density ν with respect to the product measure $\tau_1 \otimes \tau_2$ on $K_1 \times K_2$ so that

$$\mu\left(pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)\right) \coloneqq \int_{pr_1^{-1}(U_1) \cap pr_2^{-1}(U_2)} \nu \, \mathrm{d}(\tau_1 \otimes \tau_2). \tag{13}$$

The case of S_3 finite encompasses the discrete case above, since for S_1, S_2 we may take the ambient space K to be $S_1 \coprod S_2$ with measure given by set cardinality. We treat the discrete case separately because we can obtain a clean and simpler universality statement, which we hope clarifies the exposition.

Proposition 12 *For the probability measure μ, we have $(pr_i)_\star \mu = \mu_i$ for $i = 1, 2$.*

Proof We first observe that for any $s_3 \in S_3$

$$\begin{aligned} \nu_3(s_3) &= \mu_3(s_3) \\ &= (f_2)_\star \mu_2(s_3) \\ &= \mu_2(f_2^{-1}(s_3)) \\ &= \int_{f_2^{-1}(s_3)} \nu_2(s_2) \, \mathrm{d}\tau_2. \end{aligned}$$

We then have for any measurable subset $U_1 \subseteq S_1$

$$\begin{aligned} (pr_1)_\star \mu(U_1) = \mu(pr_1^{-1}(U_1)) &= \mu(U_1 \times_{S_3} S_2) \\ &= \int_{U_1 \times_{S_3} S_2} \nu \, \mathrm{d}(\tau_1 \otimes \tau_2) \\ &= \int_{U_1 \times S_2} \mathbb{I}_{U_1 \times_{S_3} S_2} \cdot \nu \, \mathrm{d}(\tau_1 \otimes \tau_2), \end{aligned}$$

so, applying Fubini–Tonelli's theorem, we obtain

$$\begin{aligned}
\mu(U_1 \times_{S_3} S_2) &= \int_{U_1} \left(\int_{S_2} \mathbb{I}_{U_1 \times_{S_3} S_2} \cdot \nu \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \left(\int_{f_2^{-1}(f_1(s_1))} \frac{\nu_2(s_2)}{\nu_3(f_2(s_2))} \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \left(\int_{f_2^{-1}(f_1(s_1))} \frac{\nu_2(s_2)}{\nu_3(f_1(s_1))} \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \frac{\nu_1(s_1)}{\nu_3(f_1(s_1))} \left(\int_{f_2^{-1}(f_1(s_1))} \nu_2(s_2) \, \mathrm{d}\tau_2 \right) \mathrm{d}\tau_1 \\
&= \int_{U_1} \frac{\nu_1(s_1)}{\nu_3(f_1(s_1))} \nu_3(f_1(s_1)) \, \mathrm{d}\tau_1 \\
&= \int_{U_1} \nu_1(s_1) \, \mathrm{d}\tau_1 = \mu_1(U_1).
\end{aligned}$$

This shows that $(pr_1)_\star \mu = \mu_1$. An identical computation shows that $(pr_2)_\star \mu = \mu_2$ and we are done. ■

Remark 13 *Suppose that we consider MDPs without a reward function. In that case, when $\mathcal{M}_3$ is the terminal object* pt *and we take $S_1 = K_1$, $S_2 = K_2$ and ν_1, ν_2 identically 1, we obtain the cartesian product $\mathcal{M}_1 \times \mathcal{M}_2$ of two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$. One may then equip this product with a non-unique choice of reward function, possibly depending on context. A common such choice would be the sum of the reward functions of $\mathcal{M}_1$ and $\mathcal{M}_2$, but any other function of the two reward functions would work as well.*

In the presence of rewards, the situation is more subtle, as obtaining the above cartesian product through a pullback over pt *requires that the reward functions of $\mathcal{M}_1$ and $\mathcal{M}_2$ are constant and have the same value.*

The local fibration case We finally treat the general case. We introduce some terminology.

Definition 14 *A measurable function between two measurable spaces $f\colon X \to Y$ is a local fibration if there exists a measurable partition $Y = Y_1 \coprod \cdots \coprod Y_N$ such that for every index i, we have that $X_i = f^{-1}(Y_i) \simeq F_i \times Y_i$ for some measure space F_i and $f|_{X_i}$ is projection onto Y_i.*

A morphism $(f, g)\colon \mathcal{M}_1 \to \mathcal{M}_2$ between MDPs is called a local fibration if the map $f\colon S_1 \to S_2$ is a local fibration.

The intuition behind the introduction of a local fibration is that it allows us to practically treat the base state space S_3 as finite, essentially reducing the case of a continuous state space to that of a discrete state space.

We now generalize Definition 10 to the local fibration setting, following the definition of conditional independence given by Swart (1996).

Definition 15 *Let $(\alpha_i, \beta_i)\colon \mathcal{N} \to \mathcal{M}_i$, $i = 1, 2$, be two morphisms fitting into a commutative diagram*

$$\begin{array}{ccc} \mathcal{N} & \longrightarrow & \mathcal{M}_1 \\ \downarrow & & \downarrow \\ \mathcal{M}_2 & \longrightarrow & \mathcal{M}_3, \end{array}$$

where the two morphisms $\mathcal{M}_1 \to \mathcal{M}_3$ and $\mathcal{M}_2 \to \mathcal{M}_3$ are local fibrations.

For any $a \in A_{\mathcal{N}}$, write $a_i = \beta_i(a) \in A_i$ and $\mu_i = T_i(a_i) \in \mathcal{P}_{S_i}$ for $i = 1, 2$. Let also $\mu_3 = (f_i)_\star \mu_i \in \mathcal{P}_{S_3}$ for $i = 1, 2$. Then, by possibly refining partitions, we can assume that $S_3 = Z_1 \coprod \cdots \coprod Z_N$ is a common partition for the fibration structure of the maps f_i with respect to the measures determined by $a \in A_{\mathcal{N}}$, so that $S_1 = X_1 \times Z_1 \coprod \cdots \coprod X_N \times Z_N$, $S_2 = Y_1 \times Z_1 \coprod \cdots \coprod Y_N \times Z_N$. We have

$$S = S_1 \times_{S_3} S_2 = (X_1 \times Y_1 \times Z_1) \coprod \cdots \coprod (X_N \times Y_N \times Z_N).$$

and the maps $\alpha_i\colon S_{\mathcal{N}} \to S_i$, $i = 1, 2$, are induced by a canonical morphism $\gamma\colon S_{\mathcal{N}} \to S_1 \times_{S_3} S_2$. Write $\mu = \gamma_* T_{\mathcal{N}}(a) \in \mathcal{P}_S$.

We say that the morphisms are conditionally independent with respect to $\mathcal{M}_3$ *if the following property is satisfied: For any given choice of partitions, as above, it holds that for each index $1 \leq j \leq N$, the probability measures*

$$\widehat{\mu}_1^j := \frac{\mu_1|_{Z_j \times X_j}}{\mu_1(Z_j \times X_j)},\ \widehat{\mu}_2^j := \frac{\mu_2|_{Z_j \times Y_j}}{\mu_2(Z_j \times Y_j)}$$

are conditionally independent with respect to the probability measure $\widehat{\mu}^j := \frac{\mu|_{X_j \times Y_j \times Z_j}}{\mu(X_j \times Y_j \times Z_j)}$.

Here, conditional independence (cf. (Swart, 1996, Section 1)) means that for any $z \in Z_j$ and any measurable subsets $U \subseteq X_j, V \subseteq Y_j$ we have

$$\mathbb{P}_{\widehat{\mu}^j}(U \times V \mid Z_j = z) = \mathbb{P}_{\widehat{\mu}_1^j}(U \mid Z_j = z) \cdot \mathbb{P}_{\widehat{\mu}_2^j}(V \mid Z_j = z). \tag{14}$$

We are being informal here with the notation used referring to probability density functions (without loss of generality) and assuming that the measures we are dividing by are always nonzero.

Remark 16 *Definition 15 is a generalization of the discrete case. Consider discrete variables X_j, Y_j, Z_j. In this scenario, the conditional probabilities can be expressed as follows*

$$\begin{aligned} \mathbb{P}_{\widehat{\mu}^j}(X_j = u, Y_j = v \mid Z_j = z) &= \frac{\mathbb{P}_{\widehat{\mu}^j}(X_j = u, Y_j = v, Z_j = z)}{\mathbb{P}_{\widehat{\mu}^j}(Z_j = z)} \\ &= \frac{\mathbb{P}_{\widehat{\mu}^j}(X_j = u, Y_j = v, Z_j = z)}{\mu_3(z)} \cdot \mu_3(Z_j), \\ \mathbb{P}_{\widehat{\mu}_1^j}(X_j = u \mid Z_j = z) &= \frac{\mathbb{P}_{\widehat{\mu}_1^j}(X_j = u, Z_j = z)}{\mathbb{P}_{\widehat{\mu}_1^j}(Z_j = z)} = \frac{\mu_1(u, z)}{\mu_3(z)} \cdot \frac{\mu_3(Z_j)}{\mu_1(X_j \times Z_j)} = \frac{\mu_1(u, z)}{\mu_3(z)}, \\ \mathbb{P}_{\widehat{\mu}_2^j}(Y_j = v \mid Z_j = z) &= \frac{\mathbb{P}_{\widehat{\mu}_2^j}(Y_j = v, Z_j = z)}{\mathbb{P}_{\widehat{\mu}_2^j}(Z_j = z)} = \frac{\mu_2(v, z)}{\mu_3(z)} \cdot \frac{\mu_3(Z_j)}{\mu_2(Y_j \times Z_j)} = \frac{\mu_2(u, z)}{\mu_3(z)}. \end{aligned}$$

Equality (14) *then becomes*

$$\widehat{\mu}^j(u,v,z) = \mathbb{P}_{\widehat{\mu}^j}(X_j = u, Y_j = v, Z_j = z) = \frac{\mu_1(u,z)\mu_2(v,z)}{\mu_3(z)\mu_3(Z_j)}.$$

By definition,

$$T_{\mathcal{N}}(a)(\alpha_1^{-1}(u,z) \cap \alpha_2^{-1}(v,z)) = \mu(u,v,z) = \mu(X_j \times Y_j \times Z_j) \cdot \widehat{\mu}^j(u,v,z),$$

so, using that $\mu_3(Z_j) = \mu_2(Y_j \times Z_j) = \mu_1(X_j \times Z_j) = \mu(X_j \times Y_j \times Z_j)$*, we obtain*

$$T_{\mathcal{N}}(a)(\alpha_1^{-1}(u,z) \cap \alpha_2^{-1}(v,z)) = \mu(u,v,z) = \frac{\mu_1(u,z)\mu_2(v,z)}{\mu_3(z)},$$

which recovers equation (12)*.*

The following is now a direct consequence of our discussion and work so far.

Theorem 17 *Assume that the state spaces* S_1, S_2 *and* S_3 *are Polish spaces and the morphisms* $\mathcal{M}_1 \to \mathcal{M}_3$ *and* $\mathcal{M}_2 \to \mathcal{M}_3$ *are local fibrations. Then there is a unique probability measure* $\mu = T(a)$*, giving rise to an MDP* $\mathcal{M} = \mathcal{M}_1 \times_{\mathcal{M}_3} \mathcal{M}_2$ *fitting in a commutative diagram* (8) *such that the morphisms* $\mathcal{M} \to \mathcal{M}_1$ *and* $\mathcal{M} \to \mathcal{M}_2$ *are conditionally independent with respect to* $\mathcal{M}_3$*. Moreover, diagram* (8) *is universal among diagrams* (9) *where the morphisms* $\mathcal{N} \to \mathcal{M}_i$ *form a conditionally independent pair.*

Proof As above, we may assume that $S_3 = Z_1 \coprod \cdots \coprod Z_N$ is a common partition for the fibration structure of the maps f_i with the measures determined by $a \in A$, so that $S_1 = X_1 \times Z_1 \coprod \cdots \coprod X_N \times Z_N$ and $S_2 = Y_1 \times Z_1 \coprod \cdots \coprod Y_N \times Z_N$.

As sets, we then have

$$S_1 \times_{S_3} S_2 = (X_1 \times Y_1 \times Z_1) \coprod \cdots \coprod (X_N \times Y_N \times Z_N)$$

and thus we may reduce to the case $S_3 = Z$ and $S_1 = X \times Z$ and $S_2 = Y \times Z$.

Existence and uniqueness then follows from the existence and uniqueness of μ with respect to local fibrations $\mathcal{M}_i \to \mathcal{M}_3$ (Swart, 1996; Brandenburger and Keisler, 2016) and the normalizing argument of remark 16. Universality can be shown by the continuous version of the argument in the proof of Proposition 11. ■

Remark 18 *A Polish space is a separable completely metrizable topological space. The requirement in the preceding theorem that the state spaces be Polish spaces is made for technical reasons and does not affect the substance of our results or their practicability.*

The constructions in the three previous special cases are applications of Theorem 17:

1. For the subprocess case, observe that any injection $f_i \colon S_i \to S_3$ is a local fibration by taking $S_i \simeq (S_i \times \{\bullet\}) \coprod ((S_3 \setminus S_i) \times \emptyset)$ and $S_3 = S_i \coprod (S_3 \setminus S_i)$. $S_i \times \{\bullet\} \to S_i$ is a fibration with fiber $\{\bullet\}$ and $(S_3 \setminus S_i) \times \emptyset \to S_3 \setminus S_i$ is a fibration with empty fiber.

2. For the finite case, any morphism $f\colon X \to Z$ between finite spaces is a local fibration, because
$$X = \coprod_{z \in Z} f^{-1}(z) \simeq \coprod_{z \in Z} f^{-1}(z) \times \{z\}$$
and $f^{-1}(z) \times \{z\} \mapsto z \in Z$ is a fibration with fiber $f^{-1}(z)$.

3. When S_3 is finite, any measurable function $f_i\colon S_i \to S_3$ is a local fibration
$$S_i = \coprod_{s_3 \in S_3} f_i^{-1}(s_3) \simeq \coprod_{s_3 \in S_3} f_i^{-1}(s_3) \times \{s_3\}$$
and $f_i^{-1}(s_3) \times \{s_3\} \mapsto s_3 \in S_3$ is a fibration with fiber $f^{-1}(s_3)$.

2.3 Pushouts: A Gluing Construction

Having discussed pullbacks, we move on to the dual categorical notion of pushout. The pushout models gluing two objects along a third object with morphisms to each. Its universal property is determined by it being minimal in an appropriate sense. Coproducts give a standard example of a pushout. In the category of sets, an example is given by the disjoint union $S_1 \coprod S_2$, which can be viewed as the pushout of the two morphisms $\emptyset \to S_1$ and $\emptyset \to S_2$.

Intuitively, the pushout is the result of gluing two MDPs $\mathcal{M}_1$ and $\mathcal{M}_2$ along a third MDP $\mathcal{M}_3$ which is expressed as a component of both through morphisms $m_1\colon \mathcal{M}_3 \to \mathcal{M}_1$ and $m_2\colon \mathcal{M}_3 \to \mathcal{M}_2$.

We show that the category MDP admits pushouts. Unlike the case of pullbacks, we do not encounter measure-theoretic obstructions, as pushouts of sets and pushforwards of measures are compatible (and both covariant) operations. Therefore, MDP is suitable for expressing compositional structures of MDPs in a universal way.

Theorem 19 *Suppose that we have three MDPs $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ together with two morphisms $m_1 = (f_1, g_1)\colon \mathcal{M}_3 \to \mathcal{M}_1$ and $m_2 = (f_2, g_2)\colon \mathcal{M}_3 \to \mathcal{M}_2$. There exists an MDP $\mathcal{M} = \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ and morphisms $\mathcal{M}_1 \to \mathcal{M}, \mathcal{M}_2 \to \mathcal{M}$ fitting in a pushout diagram in* MDP:

$$\begin{array}{ccc} \mathcal{M}_3 & \xrightarrow{m_1} & \mathcal{M}_1 \\ {\scriptstyle m_2}\downarrow & & \downarrow \\ \mathcal{M}_2 & \longrightarrow & \mathcal{M} \end{array}$$

Proof We wish to glue $\mathcal{M}_1$ and $\mathcal{M}_2$ along their overlap coming from $\mathcal{M}_3$ to obtain a new MDP, denoted by $\mathcal{M} \coloneqq \mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$, such that there exist natural maps $\mathcal{M}_1 \to \mathcal{M}$ and $\mathcal{M}_2 \to \mathcal{M}$, giving a pushout diagram as above

$$\begin{array}{ccc} \mathcal{M}_3 & \xrightarrow{m_1} & \mathcal{M}_1 \\ {\scriptstyle m_2}\downarrow & & \downarrow \\ \mathcal{M}_2 & \longrightarrow & \mathcal{M}. \end{array}$$

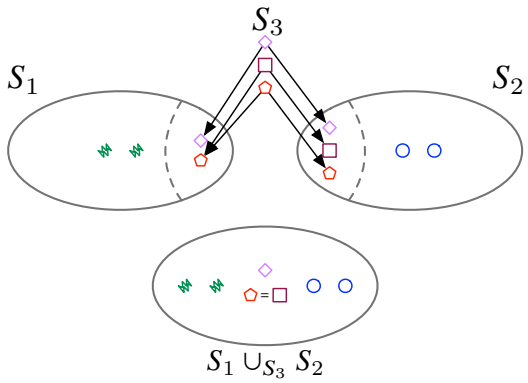


Figure 2: Gluing state spaces S_i. Gluing works similarly for state-action spaces A_i.

We propose the following construction (figure 2). To define the state space S of $\mathcal{M}$, we take

$$S = S_1 \cup_{S_3} S_2 = S_1 \coprod S_2 / \sim \tag{15}$$

where the equivalence relation $\sim$ is generated by identifying $f_1(s_3) \in S_1$ with $f_2(s_3) \in S_2$ for all $s_3 \in S_3$. This gives a pushout diagram in the category of sets:

$$\begin{array}{ccc} S_3 & \xrightarrow{f_1} & S_1 \\ {\scriptstyle f_2}\downarrow & & \downarrow{\scriptstyle i_1} \\ S_2 & \xrightarrow[i_2]{} & S. \end{array} \tag{16}$$

Observe that the state space is a disjoint union of three components

$$S = (S_1 \setminus f_1(S_3)) \coprod (f_1(S_3) \sim f_2(S_3)) \coprod (S_2 \setminus f_2(S_3)). \tag{17}$$

Set $S_1^\circ = S_1 \setminus f_1(S_3)$, $S_2^\circ = S_2 \setminus f_2(S_3)$. By abuse of notation, we write S_3 to denote the middle component, when this is clear from context.

To specify the action space A and the projection map $\psi\colon A \to S$, by formula (1), it suffices to define the action spaces A_s for each state $s \in S$. We consider each component separately:

1. Over each S_i°, for $s \in S_i^\circ$, since S_i° is naturally a subset of S_i, we define $A_s \coloneqq (A_i)_s$.
2. Over the overlap S_3, for each state $s = f_1(s_3) = f_2(s_3) \in S$, where $s_3 \in S_3$, we let
$$A_s = (A_1)_s \coprod (A_2)_s / \sim \tag{18}$$
where $\sim$ is generated by identifying $g_1(a_3)$ with $g_2(a_3)$ for all actions $a_3 \in A_3$.

As for state spaces, we have defined A as the following pushout in the category of sets:

$$\begin{array}{ccc} A_3 & \xrightarrow{g_1} & A_1 \\ {\scriptstyle g_2}\downarrow & & \downarrow{\scriptstyle j_1} \\ A_2 & \xrightarrow[j_2]{} & A. \end{array} \tag{19}$$

By the universal property of pushouts, diagrams (17) and (19) imply the existence of the canonical morphism $\psi\colon A \to S$.

We move on to the transition probability information, namely the morphism $T\colon A \to \mathcal{P}_S$. We proceed componentwise specifying $T_s\colon A_s \to \mathcal{P}_S$ for s in S_i° and S_3:

1. Over each S_i°, we define $T_s = (T_i)_s$. This makes sense since $A_s = (A_i)_s$ in this case.
2. Over S_3, according to formula (18), the action space consists of components $(A_i)_s$.

   For $a \in (A_i)_s$ which does not lie in the image of g_i, define $T(a)$ to equal $(i_i)_\star T_i(a)$.

   Otherwise, suppose that $a_1 = g_1(a_3) \in A_s$ for some action $a_3 \in A_3$. Write $a_2 = g_2(a_3)$. We observe that

   $$(i_1)_\star T_1(a_1) = (i_2)_\star T_2(a_2). \tag{20}$$

   This follows from the equality $i_1 \circ f_1 = i_2 \circ f_2$ and diagram (3), since

   $$\begin{aligned}(i_1)_\star T_1(a_1) &= (i_1)_\star T_1(g_1(a_3))\\ &= (i_1)_\star (f_1)_\star T_3(a_3)\\ &= (i_1 \circ f_1)_\star T_3(a_3),\\ (i_2)_\star T_2(a_2) &= (i_2)_\star T_2(g_2(a_3))\\ &= (i_2)_\star (f_2)_\star T_3(a_3)\\ &= (i_2 \circ f_2)_\star T_3(a_3).\end{aligned}$$

   We may, thus, define unambiguously

   $$T(a) = (i_1)_\star T_1(a_1) = (i_2)_\star T_2(a_2). \tag{21}$$

   This expression is independent of the choice of index i, so it respects the equivalence relation $\sim$ on A. If the action is in the image of g_2, we argue in an identical way.

Finally, the reward function $R\colon A \to \mathbb{R}$ is defined to be R_1 on the image $j_1(A_3) \subseteq A$ and R_2 on the image $j_2(A_3)$. This is well-defined since by definition $R_1 \circ g_1 = R_3 = R_2 \circ g_2$.

We now verify that this construction indeed gives a pushout.

By definition, we have commutative diagrams

$$\begin{array}{ccc} A_1 & \xrightarrow{j_1} & A \\ {\scriptstyle T_1}\downarrow & & \downarrow{\scriptstyle T} \\ \mathcal{P}_{S_1} & \xrightarrow[(i_1)_\star]{} & \mathcal{P}_S, \end{array} \qquad \begin{array}{ccc} A_2 & \xrightarrow{j_2} & A \\ {\scriptstyle T_2}\downarrow & & \downarrow{\scriptstyle T} \\ \mathcal{P}_{S_2} & \xrightarrow[(i_2)_\star]{} & \mathcal{P}_S. \end{array} \tag{22}$$

Therefore we obtain natural morphisms $\rho_1 = (i_1, j_1)\colon \mathcal{M}_1 \to \mathcal{M}$ and $\rho_2 = (i_2, j_2)\colon \mathcal{M}_2 \to \mathcal{M}$ fitting in a commutative diagram

$$\begin{array}{ccc} \mathcal{M}_3 & \xrightarrow{m_1} & \mathcal{M}_1 \\ {\scriptstyle m_2}\downarrow & & \downarrow{\scriptstyle \rho_1} \\ \mathcal{M}_2 & \xrightarrow[\rho_2]{} & \mathcal{M}. \end{array}$$

Now let $\mathcal{N}$ be a MDP fitting in a commutative diagram

$$\begin{array}{ccc} \mathcal{M}_3 & \xrightarrow{m_1} & \mathcal{M}_1 \\ {\scriptstyle m_2}\downarrow & & \downarrow{\scriptstyle (\alpha_1,\beta_1)} \\ \mathcal{M}_2 & \xrightarrow[(\alpha_2,\beta_2)]{} & \mathcal{N}. \end{array}$$

We need to check that the diagram is induced by a canonical morphism $m = (f, g)\colon \mathcal{M} \to \mathcal{N}$. By the definition of the state and action spaces of $\mathcal{M}$, they are the pushouts (in the category of sets) of the corresponding spaces of $\mathcal{M}_i$ along those of $\mathcal{M}_3$ so there are natural candidates for f and g. These fit in a commutative diagram (2), so it remains to verify that diagram (3) is commutative, as compatibility of rewards is again clear.

To avoid confusion, we now use the subscripts $\mathcal{M}$ and $\mathcal{N}$ to indicate the MDP to which state and action spaces correspond below.

Since the maps $j_1\colon A_1 \to A$ and $j_2\colon A_2 \to A$ are jointly surjective, we only check that the outer square in the diagram

$$\begin{array}{ccccc} A_1 & \xrightarrow{j_1} & A_{\mathcal{M}} & \xrightarrow{g} & A_{\mathcal{N}} \\ {\scriptstyle T_1}\downarrow & & \downarrow{\scriptstyle T_{\mathcal{M}}} & & \downarrow{\scriptstyle T_{\mathcal{N}}} \\ \mathcal{P}_{S_1} & \xrightarrow[(i_1)_\star]{} & \mathcal{P}_{S_{\mathcal{M}}} & \xrightarrow[f_\star]{} & \mathcal{P}_{S_{\mathcal{N}}} \end{array}$$

commutes and the same is true for the corresponding diagram for j_2. But this is the case by construction of f and g, since $g \circ j_1 = \beta_1$ and $f \circ i_1 = \alpha_1$ and $(\alpha_1, \beta_1)\colon \mathcal{M}_1 \to \mathcal{N}$ is a morphism in MDP. ■

The following proposition shows that gluing behaves well with respect to subprocesses.

Proposition 20 *Suppose that $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$. Then $\mathcal{M}_1$ and $\mathcal{M}_2$ are subprocesses of $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$.*

Proof We need to show that functions i_1, i_2 in diagram (17) are injective, given that f_1, f_2 are injective, and that j_1, j_2 in diagram (19) are injective given the injectivity of g_1, g_2. We check that this is true for i_1. The argument works for all maps.

Suppose that $i_1(s_1) = i_1(s_1')$. If $i_1(s_1) \in S_1^\circ \subseteq S$, then we must have $s_1 = s_1' \in S_1^\circ$ since $i_1|_{S_1^\circ}$ is the identity map from $S_1^\circ \subseteq S_1$ to $S_1^\circ \subseteq S$.

If $i_1(s_1) \in f_1(S_3) \subseteq S$, then there exists a unique $s_3 \in S_3$ such that $s_1 = f_1(s_3)$, since f_1 is injective. Similarly $s_1' = f_1(s_3')$. Now $i_1(s_1) = i_1(s_1')$ implies that $f_1(s_3) \sim f_1(s_3')$ under the equivalence relation $\sim$ on $S_1 \coprod S_2$, whose quotient is S by definition. Since f_2 is also injective, it follows that the equivalence classes of $\sim$ are pairs $\{f_1(t), f_2(t)\} \subseteq S_1 \coprod S_2$ where t runs through S_3. Hence $f_1(s_3) \sim f_1(s_3')$ is only possible if $f_1(s_3) = f_1(s_3')$, which implies that $s_3 = s_3'$. ■

We finally state the following corollary, which is a consequence (and generalization) of Theorem 19 combined with (Stacks project authors, 2024, Lemma 002Q) and Proposition 5.

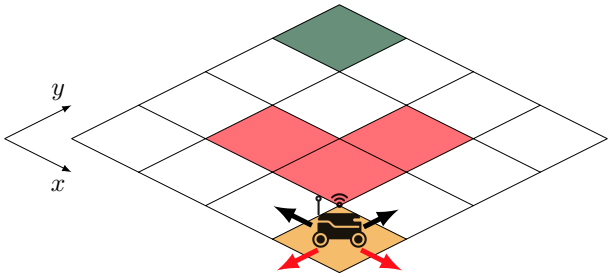


Figure 3: A grid world with a starting state (yellow, s), a destination state (green, $\mathbb{D}$), and obstacles (red, $\mathbb{O}$). We can have arbitrary complex worlds of this form; we use the simplest to explain some of the properties we can exploit using categorical RL.

Namely, a category that has an initial object and admits pushouts, admits finite colimits. These should be thought of as generalizations of pushouts and allow for more complex combinations of MDPs along components.

Corollary 21 *The category* MDP *admits finite colimits.*

3. Compositional Knowledge Representations for Reinforcement Learning

In this section, we use the compositional theory and its properties to synthesize increasingly complex behaviors. One of the advantages of our framework is the ability to address different forms of compositionality in a uniform and systematic fashion. For example, consider functional (or logical) composition and temporal composition. An instance of the former could be the simultaneous execution of two tasks by an agent, which may or may not interact with each other, such as juggling three balls while riding a unicycle. An instance of the latter could be the sequential completion of two tasks, the first of which might be a prerequisite for the second, such as baking bread before making a sandwich. From the point of view of the theory developed in this paper, both of these compositional behaviors are treated in the same way by the appropriate diagrammatic language (indeed, zig-zag diagrams encapsulate sequential task completion).

3.1 Safe Grid Worlds: A Motivating Example

We consider the case of a grid world (Leike et al., 2017) constructed as a 4×4 grid, where an agent attempts to navigate from a starting position to a destination position in the presence of some obstacles (yellow, green and red respectively in figure 3).

Definition 6 is also well-suited to modeling the removal of a subset $\mathbb{O} \subseteq S$ from the state space of a MDP to obtain a new MDP.

Definition 22 (Puncturing undesired states and actions) *Let $\mathcal{M} = (S, A, \psi, T, R)$ be an MDP and $\mathbb{O} \subseteq S$. The MDP $\mathcal{M}^\circ$ obtained by puncturing $\mathcal{M}$ along $\mathbb{O}$ is the MDP $(S^\circ, A^\circ, \psi^\circ, T^\circ, R^\circ)$ where*

1. $S^\circ = S \setminus \mathbb{O}$.

2. *Let* $B = \{a \in A \mid T(a)(\mathbb{O}) > 0\}$. *Then* $A^\circ = A \setminus \big(\psi^{-1}(\mathbb{O}) \cup B\big)$.
3. $\psi^\circ = \psi|_{A^\circ}$.
4. $T^\circ = T|_{A^\circ}$.
5. $R^\circ = R|_{S^\circ}$.

There is a canonical morphism $\mathcal{M}^\circ \to \mathcal{M}$, *which exhibits* $\mathcal{M}^\circ$ *as a subprocess of* $\mathcal{M}$. *In fact,* $\mathcal{M}^\circ$ *is the canonical maximal subprocess for the subset* $S^\circ \subseteq S$. *For a morphism of MDPs* $m = (f, g)\colon \mathcal{M}_1 \to \mathcal{M}_2$, *the punctured MDP* $\mathcal{M}_2^\circ$ *obtained by puncturing* $\mathcal{M}_2$ *along* $\mathcal{M}_1$ *is defined as the puncture of* $\mathcal{M}_2$ *along the subset* $f(S_1) \subseteq S_2$.

This construction is well-defined since for any action in A°, the probability of ending in $\mathbb{O}$ is zero, as we removed exactly these actions, which forms set B. This is where allowing the action spaces to vary along the state space S gives us the flexibility to modify the actions locally to avoid the set $\mathbb{O}$.

Gluing (Section 2.3) also behaves well with respect to puncturing (Proposition 20 and Proposition 23).

Proposition 23 *Suppose* $\mathcal{M}_3$ *is a subprocess of* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *and any action* $a_2 \in A_2 \backslash g_2(A_3)$ *is not supported on* S_3, *meaning that there is some measurable subset* $U_2 \subseteq S_2$ *disjoint from* $f_2(S_3)$ *such that* $T(a_2)(U_2) > 0$. *Let* $\mathcal{M}_2^\circ$ *be the MDP obtained by puncturing* $\mathcal{M}_2$ *along* $\mathcal{M}_3$. *Then the MDP* $(\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2)^\circ$ *obtained by puncturing* $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ *along the subprocess* $\mathcal{M}_2^\circ$ *is the MDP* $\mathcal{M}_1$.

Proof The state spaces of $(\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2)^\circ$ and $\mathcal{M}_1$ coincide. For the action spaces, by the given condition it follows that puncturing $\mathcal{M}_1 \cup_{\mathcal{M}_3} \mathcal{M}_2$ along $\mathcal{M}_2$ will remove precisely the actions $A_2 \setminus g_2(A_3) \subseteq A$. But, since $\mathcal{M}_3$ is a subprocess of $\mathcal{M}_1$ and $\mathcal{M}_2$ and $g_1(A_3) = g_2(A_3)$, it follows that $A \setminus (A_2 \setminus g_2(A_3)) = A_1$. ■

We illustrate the naturality of the above constructions in the context of puncturing MDPs to enforce a safety condition.

Example 1 (Static obstacles) *Fix an MDP* $\mathcal{M} = (S, A, \psi, T, R)$ *and consider two disjoint subsets* $\mathbb{O}_1, \mathbb{O}_2 \subseteq S$. *We can then construct the MDPs:*

1. *The punctured MDPs* $\mathcal{M}_i^\circ$ *along* $\mathbb{O}_i$ *for* $i = 1, 2$.
2. *The punctured MDP* $\mathcal{M}_{12}^\circ$ *along* $\mathbb{O}_1 \cup \mathbb{O}_2$.

Proposition 24 *There exists a commutative diagram*

$$\begin{array}{ccc} \mathcal{M}_{12}^\circ & \longrightarrow & \mathcal{M}_1^\circ \\ \downarrow & & \downarrow \\ \mathcal{M}_2^\circ & \longrightarrow & \mathcal{M} \end{array}$$

which is simultaneously a pullback and pushout diagram in MDP; *that is,* $\mathcal{M}_{12}^\circ = \mathcal{M}_1^\circ \times_{\mathcal{M}} \mathcal{M}_2^\circ$ *and* $\mathcal{M} = \mathcal{M}_1^\circ \cup_{\mathcal{M}_{12}^\circ} \mathcal{M}_2^\circ$.

Proof To see that this is a pullback diagram, observe that the state space of the pullback is the intersection of state spaces of the two punctured MDPs $\mathcal{M}_i^\circ$, which is the same as the state space of $\mathcal{M}_{12}^\circ$. For the action spaces, the action space of the pullback consists of the actions in $\mathcal{M}$ that avoid the two obstacles $\mathbb{O}_i$. This coincides with the MDP $\mathcal{M}_{12}$ actions. The transition probabilities coincide tautologically.

The reasoning for the pushout is analogous. ■

Example 2 (Collision avoidance) *Fix an MDP $\mathcal{M}_1 = (S, A, \psi, T, R)$, which we consider as a model for an agent moving through the state space S with possible actions A. We may work with the product $\mathcal{M}_2 = \mathcal{M} \times \mathcal{M}$ to model the movement of two independent agents. If, in addition, we would like to make sure that the two agents never collide, we may puncture $\mathcal{M} \times \mathcal{M}$ along the diagonal*

$$\mathbb{O}_2 = \Delta = \{(s, s) \mid s \in S\} \subseteq S \times S$$

to get the MDP $\mathcal{M}_2^\circ$.

We can use, for example, the above construction to model the movement of N independent agents, ensuring that no two of them collide, by puncturing the product MDP with N factors $\mathcal{M}_N = \mathcal{M} \times \ldots \times \mathcal{M}$ along the big diagonal

$$\begin{aligned}\mathbb{O}_N &= \bigcup_{1 \le i < j \le N} \Delta_{ij} \\ &= \bigcup_{1 \le i < j \le N} \{(s_1, \ldots, s_N) \mid s_i = s_j\} \subseteq S^N\end{aligned}$$

to define the MDP $\mathcal{M}_N^\circ$.

3.2 Zig-zag Diagrams: A Language Equipped with Compositional Verification

For designing compositional tasks, we desire to operationalize using the categorical semantics of RL, that involve accomplishing tasks sequentially. The zig-zag diagrams we consider below are not merely an intuitive diagrammatic representation of sequential task completion, but rather they invoke directly the results of this work. Therefore, they represent a form of *compositional verification* for operational goals and stopping conditions (Figure 4).

In a general setting, we consider the setup given by, what we term, a *zig-zag* diagram of MDPs

$$\begin{array}{ccccccccc} & \mathcal{N}_0 & & \mathcal{N}_1 & & \ldots & & \mathcal{N}_{n-1} & \\ \swarrow & & \searrow\ \swarrow & & \searrow & & \swarrow & & \searrow \\ \mathcal{M}_0 & & \mathcal{M}_1 & & \mathcal{M}_2 & \ldots & \mathcal{M}_{n-1} & & \mathcal{M}_n \end{array} \tag{23}$$

where for each $i = 0, \ldots, n-1$, $\mathcal{N}_i$ is a subprocess of $\mathcal{M}_i$ (Definition 6).

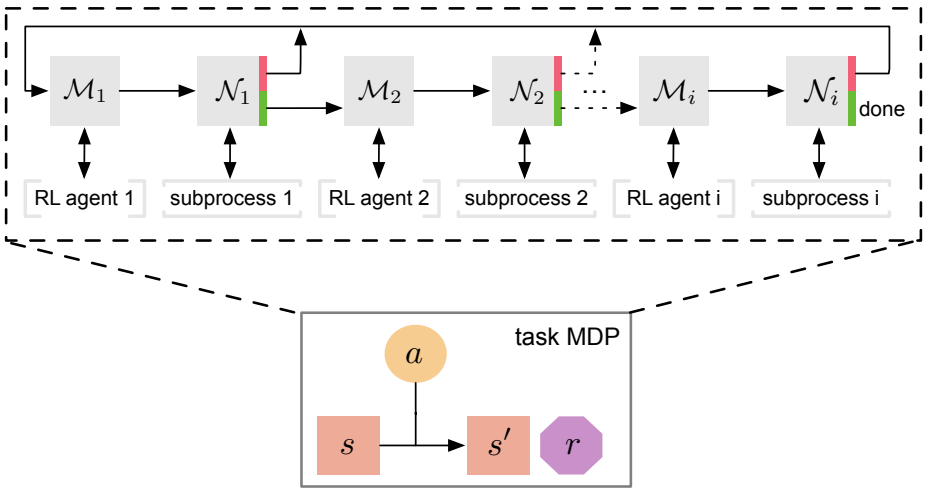


Figure 4: Main task decomposed into sub-tasks $(M_1, M_2, \cdots, M_i)$ and subprocesses that check for correct task completion $(N_1, N_2, \cdots, N_i)$: the categorical formalism introduces compositional guarantees for the chaining of tasks.

The composite MDP associated to the above diagram is the MDP $\mathcal{C}_n$ defined by the inductive rule

$$\begin{aligned} \mathcal{C}_0 &\coloneqq \mathcal{M}_0, \\ \mathcal{C}_1 &\coloneqq \mathcal{C}_0 \cup_{\mathcal{N}_0} \mathcal{M}_1, \\ &\vdots \\ \mathcal{C}_n &\coloneqq \mathcal{C}_{n-1} \cup_{\mathcal{N}_{n-1}} \mathcal{M}_n. \end{aligned}$$

The intuitive interpretation of the above zig-zag diagram is that it blackboxes compositional task completion. In particular, each subprocess $\mathcal{N}_i \to \mathcal{M}_i$ models the completion of a task in the sense that the goal of an agent is to eventually find themself at a state of $\mathcal{N}_i$. Once the i-th goal is accomplished inside the environment given by $\mathcal{M}_i$, we allow for the possibility of a changing environment and more options for states and actions in order to achieve the next goal modeled by the subprocess $\mathcal{N}_{i+1} \to \mathcal{M}_{i+1}$.

The composite MDP $\mathcal{C}_n$ is a single environment capturing all the tasks at the same time (and is in fact the colimit of diagram (23) in the category MDP).

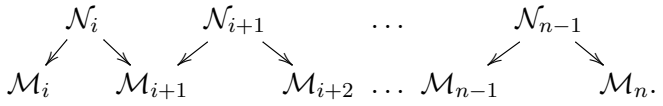
⟨?⟩ Suppose an agent has learned an optimal policy for each MDP $\mathcal{M}_i$ given the reward function R_i for achieving the i-th goal for each $i = 0, \ldots, n$. Under what conditions do these optimal policies determine optimality for the joint reward on the composite MDP $\mathcal{C}_n$?

A scenario in which this is true is when the zig-zag diagram is forward-moving, meaning that $\mathcal{N}_i$ is a full subprocess of $\mathcal{M}_i$, and the optimal value function $v_\star(s)$ for any state s in the state space of a component $\mathcal{M}_i$, considered as a state of $\mathcal{C}_n$, is *monotonic* for subsequent

subprocesses $\mathcal{M}_{i+1}, \ldots, \mathcal{M}_n$. For simplicity, we take everything to be discrete. Monotonicity here means that the expressions

$$\sum_{s' \in S_i} T(a)(s')(R_i(a) + \gamma \cdot v_\star^{\mathcal{C}_n}(s'))$$
$$\sum_{s' \in S_i} T(a)(s')(R_i(a) + \gamma \cdot v_\star^{\mathcal{C}_{[i,n]}}(s'))$$

are maximized by the same action $a \in (A_i)_s$, where we have fixed a discount factor γ. Here $\mathcal{C}_{[i,n]}$ denotes the composite MDP of the truncated zig-zag diagram



A zig-zag diagram can always be made forward-moving by removing the actions of $\mathcal{N}_i$ that can potentially move the agent off $\mathcal{N}_i$ back into $\mathcal{M}_i$. This is the operation of puncturing $\mathcal{M}_i$ along the complement of $\mathcal{N}_i$ and intersecting the result with $\mathcal{N}_i$.

Theorem 25 *Suppose that a zig-zag diagram is forward-moving and the optimal value function of $\mathcal{C}_n$ is monotonic as above. Then, following the optimal policy π_i on each component $\mathcal{M}_i$ gives an optimal policy on the composite MDP $\mathcal{C}_n$.*

Proof First, consider the Bellman equation for all $s \in S$, $a \in A$, R the reward function, and γ the discount factor we have fixed (Sutton and Barto, 2018, chapter 4)

$$v_\star(s) = \max_{a \in A_s} \sum_{s' \in S} T(a)(s') \left(R(a) + \gamma v_\star\left(s'\right)\right).$$

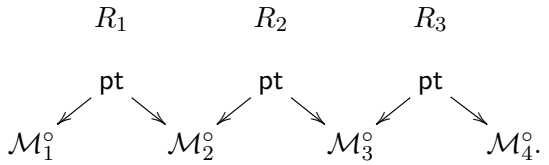
We argue by reverse induction. Suppose that the claim is true for the composite $\mathcal{C}_{[i+1,n]}$.

We then show that it is true for $\mathcal{C}_{[i,n]}$. Since the zig-zag diagram is forward-moving, the optimal values of $\mathcal{C}_{[i,n]}$ and $\mathcal{C}_{[i+1,n]}$ coincide on the state space of the latter, as successor states in $\mathcal{C}_{[i+1,n]}$ are states in $\mathcal{C}_{[i+1,n]}$. For the same reason, the optimal values of states of $\mathcal{M}_i$ are the same for the composite MDPs $\mathcal{C}_n$ and $\mathcal{C}_{[i,n]}$. The monotonicity condition now implies that following policy π_i on $\mathcal{M}_i$ and the optimal policy on $\mathcal{C}_{[i+1,n]}$ gives an optimal policy on $\mathcal{C}_{[i,n]}$.

For the base case, observe that $\mathcal{M}_n = \mathcal{C}_{[n,n]}$ which has maximal policy π_n. ■

Example 3 (Visiting regions sequentially) *Consider a point mass robot sequentially visiting three points, R_1, R_2, R_3, in a grid world, while always avoiding obstacles.*

We can model this problem compositionally by the following zig-zag diagram, where we consider punctured MDPs (Proposition 22)



At any given position in the grid the five options for the point mass robot are forward, backwards, left, right, and stay in the same position. Here each $\mathcal{M}_i^\circ$ denotes the MDP in which all the obstacles have been punctured and the actions forward, backwards, left, right have been removed at the point R_i. Each intermediate subprocess $\mathsf{pt} \to \mathcal{M}_i$ maps the stationary point to the point region R_i. This requires us moving from MDP to MDP to puncture the actions that lead to moving away from the next subsequent sequence we want.

The problem is inductively defined by the composite MDP:

$$\mathcal{C}_{\text{robot}} = \mathcal{M}_1^\circ \cup_{\mathsf{pt}} \mathcal{M}_2^\circ \cup_{\mathsf{pt}} \mathcal{M}_3^\circ \cup_{\mathsf{pt}} \mathcal{M}_4^\circ.$$

Observe that this zig-zag diagram is forward-moving and the optimal value function of $\mathcal{C}_{\text{robot}}$ is monotonic. Thus, the optimal policy on $\mathcal{C}_{\text{robot}}$ is given by the optimal policy of each component $\mathcal{M}_i^\circ$ (Theorem 25).

The more categorically minded reader will observe that zig-zag diagrams bear a close relationship with spans and cospans (Cicala, 2018).

In practice, the language of zig-zag diagrams in RL enables a structured, abstract, and rigorous understanding of complex sequential tasks via the following properties.

- **Semantics:** Diagrammatic languages map syntactic constructs to mathematical objects, explaining what each part of a system means (Diskin and Maibaum, 2012). Zig-zag diagrams represent the relationships between MDPs and subprocesses in a sequential task and encode the semantics of how these processes interact.

- **Compositionality:** The meaning of the constituent parts determines the meaning of a complex expression, ensuring modularity (Coecke, 2023). Similarly, the zig-zag pattern shows how complex processes comprise subprocesses and individual MDPs.

- **Abstraction:** Diagrammatic languages abstract away many implementation details and focus on the meaning or behavior of constructs. Zig-zag diagrams abstract away the specific workings of each MDP and subprocess, focusing instead on their high-level relationships. Similar diagrammatic languages from applied category theory have used abstraction to make progress in designing complex systems (Zardini et al., 2021; Abbott and Zardini, 2024; Bakirtzis et al., 2021; Breiner et al., 2019; Schultz et al., 2016, 2019; Bonchi et al., 2019; Gavranovic et al., 2024; Hedges and Sakamoto, 2024).

- **Formal system:** The relationships represented in a zig-zag diagram are subject to compositionality conditions, providing a formal view to manipulating the composite system often used in, for example, model-based engineering (Diskin et al., 2019).

3.3 State-action Symmetry

So far we have discussed approaches to helping designers of RL systems organize into functional subsystems. However, categorical semantics can provide benefits beyond that. By exploiting structure, we can make learning easier for the agents; for example, one of the benefits of working with algebras in some category is that we can use gadgets from algebra to exploit the geometry of an RL problem. Using algebraic gadgets becomes helpful when considering efficient approaches to *symmetric* RL problems, such as homomorphic

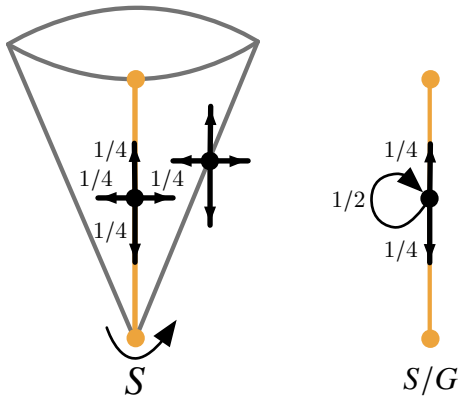


Figure 5: A conical state space collapses to a line in the quotient by axial rotation. The actions of going up and down stay the same, while left and right merge into one.

networks (van der Pol et al., 2020). We will investigate how homomorphic networks abstract within our categorical theory and improve generalization.

MDP homomorphisms in our theory can be viewed as morphisms between MDPs, which preserve composition and *forget* unrelated structure. In the concept of symmetry, we will rely on the construction of a *quotient MDP*. The symmetric structure of the RL problems we consider in this section alters the state-action space such that it is more economical to use the geometry of the problem. The construction removes the state-action pairs that can be related through symmetry. These symmetries show up in several RL problems and are common to mechanical systems, where we can think of symmetry as moving up compared to down or left compared to the right. In the larger context of designing RL systems we can think of the below operations as engineering within an *abstraction*—related to the "forget" operation. In some cases, we would instead need to add structure, which would be the same as applying some sort of *refinement* operation within our framework.

The goal of discussing symmetries is, therefore, twofold. First, we study symmetries to show the generality of our categorical formalism. Second, we study symmetries to show how design operations, as they occur in the engineering of RL systems, such as abstraction, reflect precisely within the categorical semantics of RL.

Denote by $Aut(\mathcal{M})$ the set of isomorphisms of an object of MDP; a group under composition. A group action of a group G on an MDP $\mathcal{M}$ is a group homomorphism $\rho\colon G \to Aut(\mathcal{M})$. Concretely this means that for every group element $g \in G$ there is an isomorphism $\rho_g = (\alpha_g, \beta_g)\colon \mathcal{M} \to \mathcal{M}$ satisfying the composition identity $\rho_{gh} = \rho_g \circ \rho_h$. Intuitively a group action gives a set of symmetries for an MDP $\mathcal{M}$. We would like to perform RL keeping this mind and obtain policies that are invariant under the given domain symmetries. In order to do this efficiently, we need access to a quotient MDP $\mathcal{M}/G$.

For a group G, we have a natural quotient MDP $\mathcal{M}/G$ constructed as follows (for example see figure 5):

1. Let $\widehat{\mathcal{M}} \coloneqq \mathcal{M} \times G$ be the MDP with state space $\widehat{S} = S \times G$ and action space $\widehat{A} = A \times G$. We have $\widehat{\psi} = \psi \times \mathrm{id}_G$ and the transition probabilities are given by the map

$$\widehat{T}(a, g) = (\mathrm{id}_S \times \iota_g)_\star T(a)$$

where $\iota_g \colon S \to G$ denotes the constant map with value $g \in G$.

2. There are natural group action and projection morphisms $\rho, pr_1 \colon \widehat{\mathcal{M}} \to \mathcal{M}$ defined as:

   - For ρ, the map on state and action spaces is given by the group actions $\alpha \coloneqq S \times G \to S$ and $\beta \coloneqq A \times G \to A$. This gives a morphism of MDPs since for any $(a, g) \in A \times G$ mapping to $(s, g) \in S \times G$ under $\widehat{\psi}$ we have

$$\alpha_\star \widehat{T}(a, g) = \alpha_\star (\mathrm{id}_S \times \iota_g)_\star T(a) = (\alpha_g)_\star T(a) = T(\beta_g(a)) = T(\beta(a, g))$$

     where we used that by definition $\alpha \circ (\mathrm{id}_S \times \iota_g) = \alpha_g$.

   - For pr_1, the same argument works for the first projection maps $S \times G \to S$ and $A \times G \to A$.

3. Define $\mathcal{M}/G$ as fitting in the pushout diagram

$$\begin{array}{ccc} \mathcal{M} \times G & \xrightarrow{pr_1} & \mathcal{M} \\ {\scriptstyle \rho}\downarrow & & \downarrow{\scriptstyle q} \\ \mathcal{M} & \xrightarrow[q]{} & \mathcal{M}/G. \end{array}$$

   Thus $\mathcal{M}/G \coloneqq \mathcal{M} \cup_{\mathcal{M} \times G} \mathcal{M}$ and $q \colon \mathcal{M} \to \mathcal{M}/G$ is the canonical quotient morphism.

The following proposition confirms that the quotient $\mathcal{M}/G$ satisfies the desired universal property.

Proposition 26 *$\mathcal{M}/G$ is the quotient of the MDP $\mathcal{M}$ by the action of G in the sense that for any MDP $\mathcal{N}$ and a G-invariant morphism $m \colon \mathcal{M} \to \mathcal{N}$, there is a unique factorization $\mathcal{M} \xrightarrow{q} \mathcal{M}/G \to \mathcal{N}$.*

Proof This follows immediately from the definition of $\mathcal{M}/G$. A morphism $m \colon \mathcal{M} \to \mathcal{N}$ is G-invariant if the maps between their state and action spaces are G-invariant and this is equivalent to the condition that

$$m \circ pr_1 = m \circ \rho \colon \mathcal{M} \times G \to \mathcal{N}.$$

The conclusion then follows by the universal property of the pushout $\mathcal{M}/G$. ■

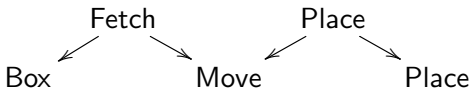
Figure 6: A fetch-and-place robot transfers objects. In dynamic environments, the robot should also be able to handle moving objects within a defined area.

3.4 A Design for Compositional Task Completion: Putting It All Together

Compositional RL problems are *sequential* in nature. An action follows another, given some rules of engagement.Those rules include how actions modify given the presence of another agent or how the actions of agents ought to intertwine in one sequential task description. We adapt a fetch-and-place robot problem (figure 6), where the learning algorithm controls actuation and rewards are assigned at task completion (Sutton and Barto, 2018).

Example 4 (Fetch-and-place robot) *Suppose that a robotic arm wants to fetch a moving object from inside a box and then place it on a shelf outside the box.*

This is captured by the diagram of MDPs



We give a simple intuitive description of each:

Box *The state space is $B \times B$ where each factor records the position of the robot arm tip and the moving object and the action space is $A \times A$.*

Fetch *The object has been fetched and, thus, the position and actions of the arm and the object coincide. The state and action spaces are given by the diagonals in $B \times B$ and $A \times A$. More generally,* Fetch *is the maximal subprocess associated with the diagonal as a subset of $B \times B$. Observe that we have made our state and action spaces smaller by recording data of half the dimension.*

Move *Since the arm needs to move the object outside of the box, we need to enlarge the state space. Thus,* $\mathsf{Move} = \mathsf{Fetch} \cup_{\mathsf{Overlap}} \mathsf{Outside}$ *where* Overlap *is a common region of the box and the outside environment. The actions are defined to allow the arm and object to move within the whole environment.*

Place *This is a full subprocess of* Outside → Move. *If the ending position is a point in* Outside, *then we may take* Place = pt *and a subprocess* pt → Outside.

The composite MDP in this setup can be expressed as

$$\begin{aligned}\mathcal{C} &= (\mathsf{Box} \cup_{\mathsf{Fetch}} \mathsf{Move}) \cup_{\mathsf{Place}} \mathsf{Place} \\ &= \mathsf{Box} \cup_{\mathsf{Fetch}} \mathsf{Move} \\ &= \mathsf{Box} \cup_{\mathsf{Fetch}} (\mathsf{Fetch} \cup_{\mathsf{Overlap}} \mathsf{Outside}).\end{aligned}$$

If the object is stationary, we can make the diagram forward-moving by deleting the actions of Fetch *inside* Box *which separate the arm and the object. Then,* Fetch *is full in* Box*, but it is no longer full in composite* $\mathcal{C}$*, allowing for continuation of movement in order to complete the final task* Place *which remains full in* $\mathcal{C}$*. In that case, we can apply Theorem 25 to compute the policies componentwise, meaning that we prove that learning-by-parts is the same as learning on the whole given some conditions for the composition of MDPs.*

4. Related Work

Category theory has found applications across mathematics and computer science, serving as a unifying mathematical language that captures the structure and relationships of systems (Pierce, 1991). In this paper, we give a uniform treatment and provide universal interfaces to abstractions of MDPs and their homomorphisms (Li et al., 2006; Ravindran and Barto, 2003, 2004), allowing, among others, a rigorous universal formalization of subtasks and embeddings of MDPs. These generalizations and unifications are particularly compelling for making progress in agent generalization via defining the abstract class of sequential decision-making and letting the agent learn the specifics of the problem (Konidaris, 2019) or being explicit in recovering and forgetting information when traversing the RL abstraction hierarchy (Abel et al., 2018). Our work is supported by previous findings on the complexity and efficiency of hierarchical RL in the context of MDP interactions (Wen et al., 2020).

Hierarchical decomposition has been a prominent theme in RL, often implemented through hierarchical MDPs (Parr and Russell, 1997; Dietterich, 2000; Ravindran, 2013; Nachum et al., 2018). Unlike previous work focusing on the decomposition of tasks, the categorical approach universally describes the structure of MDPs and their relationships to understand compositionality, scalability, and interoperability in complex decision-making systems. The immediate consequence of the zig-zag diagram is the formal proof of the minimal conditions for reaching a goal from one MDP to another. Finding the right initiation set has been effectively studied for particular classes of problems (Bagaria et al., 2023), but the categorical formalism results in full generality—we prove the specific rules of engagement that must be satisfied to verify the compositionality of any framework, working in tandem rather than in opposition. Neary et al. (2022) compose MDPs into subtask policies and meta-policies for the purpose of verification, in fact the composition is intuitive and clever in that largely experimental work but the *why* it works fits neatly in the zig-zag diagrams we present in the current paper.

Building on these principles, compositionality in RL has typically emphasized temporal and state abstractions, executing complex behaviors, and improving learning efficiency through mechanisms such as chaining known skills (Tasse et al., 2020, 2022; van Niekerk

et al., 2019; Jothimurugan et al., 2021; Ivanov et al., 2021). However, our categorical formalism shifts the focus toward the functional composition in RL through universal properties. We use the categorical formalism to decompose tasks into behavioral *functions*. Unlike previous work, this functional approach provides a more granular, mathematical structure for task decomposition. Functional composition aligns with recent efforts in solving robotics tasks (Devin et al., 2017) and discovering policy decomposition into modular neural architectures (Mendez et al., 2022), offering a rigorous compositional framework equipped with a corresponding diagrammatic language of zig-zag diagrams. The relationship between the categorical formalism and the ideas of logical composition is one of coexistence rather than comparison, in the sense that both can make use of the other to verify increasingly complex classes of RL systems.

Through monads and functors, categorical semantics have been used to describe stochastic processes and model uncertainty within probabilistic computations (Giry, 1981; Watanabe et al., 2023). Other work has attempted to apply category theory to model MDPs, embedding the dynamics of decision processes within a categorical framework (Fritz et al., 2023; Baez et al., 2016). While these studies create a bridge between abstract mathematics and probabilistic modeling, our work applies categorical principles to functional decomposition in RL. We extend the categorical perspective beyond mere probabilistic representation, using it to systematically structure complex tasks, compose behavior and produce functional decision-making algorithms for RL. Other adjacent work relates to using a category-theoretic lens in control, dynamical systems, and robotics (Hanks et al., 2024; Ames et al., 2025a,b; Ames, 2006; Culbertson et al., 2020; Lerman, 2018; Lerman and Schmidt, 2020).

5. Conclusion

This work demonstrates how a compositional theory can be used to systematically model complex behaviors, integrating both functional and temporal composition within a unified framework. By using a consistent diagrammatic approach, we show that seemingly distinct behaviors, such as independent task execution (for example, juggling while riding a unicycle) and sequential task completion (for example, baking bread before making a sandwich), can be treated equivalently. This provides a more rigorous and flexible method for analyzing compositional dynamics in diverse scenarios.

In general, structural conditions are essential for RL generalization (Du et al., 2021), safety (Alshiekh et al., 2018), and sample efficiency (Sun et al., 2019). Making explicit the relational aspects of these structural conditions gives rise to concrete mappings between formalisms (Bakirtzis and Topcu, 2022), a currently open problem in learning (Luckcuck et al., 2019). Our theoretical results give meaning to the robustness of RL's compositionality feature for structural conditions. In particular, we prove that the composition of MDPs is a pullback and has a well-defined measure, there exists a gluing operation of MDPs, and that as long as MDP composites are forward-moving, then the learning-by-parts corresponds to learning the optimal policy on the whole. Following the distinction Pattee (1978) makes between universal laws and local rules, this work seeks to make progress in identifying the fundamental *laws* of reinforcement learning—which are inexorable, incorporeal, and universal—as opposed to the more arbitrary, structure-dependent, and local *rules*.

Acknowledgments

U.T. would like to acknowledge his support from the ARO W911NF-20-1-0140 and AFOSR FA9550-22-1-0403 grants.

Appendix A. Preliminaries

Here we give a brief account of definitions of some of the mathematical structures we use in order to study the universal properties of the compositionality features in RL.

A.1 Some Categorical Notions

Consult Lawvere and Schanuel (2009); Leinster (2014), or Mac Lane (1998) for an in-depth treatment of category theory.

Definition 27 (Category) *A (small)* category C *consists of a collection of a set of objects* Ob *and for any two* $X, Y \in \mathrm{Ob}$ *a set of arrows* $f\colon X \to Y$*, along with a composition rule*

$$(f\colon X \to Y,\ g\colon Y \to Z) \mapsto g \circ f\colon X \to Z$$

and an identity arrow $\mathrm{id}_X\colon X \to X$ *for all objects, subject to associativity, whenever compositions make sense, and unity conditions:* $(f \circ g) \circ h = f \circ (g \circ h)$ *and* $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f$*.*

This definition encompasses a vast variety of structures in mathematics and other sciences: to name a couple, Set is the category of sets and functions between them, whereas Lin_k denotes the category of k-linear real vector spaces and k-linear maps between them, where k is a given field.

Definition 28 (Commutative diagrams) *A standard diagrammatic way to express composites is by concatenation of arrows:* $X \xrightarrow{f} Y \xrightarrow{g} Z$*. One can also express equations between morphisms via commutative triangles of the form*

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ & \searrow^{h} & \downarrow{\scriptstyle g} \\ & & Y \end{array} \qquad \textit{which stands for } g \circ f = h;$$

or commutative squares of the form

$$\begin{array}{ccc} X & \xrightarrow{f} & Z \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g'} \\ Y & \xrightarrow{f'} & W \end{array} \qquad \textit{which stands for } g' \circ f = f' \circ g.$$

More complicated diagrammatic expressions account for analogous relationships between morphisms.

Definition 29 (Isomorphism) *A morphism* $f\colon X \to Y$ *is called* invertible *or an* isomorphism *when there exists a morphism* $g\colon Y \to X$ *such that* $f \circ g = \mathrm{id}_Y$ *and* $g \circ f = \mathrm{id}_X$*.*

Definition 30 (Functor) *A* functor $F\colon \mathsf{C} \to \mathsf{D}$ *between two categories consists of a function between objects and a function between morphisms, mapping an object* X *of* C *to an object* FX *of* D *and a morphism* $f\colon X \to Y$ *to* $Ff\colon FX \to FY$*, such that it preserves composition and identities:* $F(f \circ g) = Ff \circ Fg$ *and* $F(\mathrm{id}_X) = \mathrm{id}_{FX}$*.*

A functor can be informally thought of as a structure-preserving map between domains of discourse. Interestingly, categories and functors form a category on their own, denoted Cat, in the sense that functors can be composed and the rest of the axioms hold.

The following definition is more technical and is provided for completeness.

Definition 31 (Cartesian category) *A category is called cartesian closed if it has all finite products and exponentials.*

The category of sets Set is cartesian closed where the product is the common product

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

We can think of this operation as, for example, organizing data on a table and the projections are then giving us the particular column and row respectively.

Pullbacks are a generalization of the notion of a product.

Definition 32 (Pullback) *A category is said to have a pullback for the pair of morphisms* $f\colon A \to C$ *and* $g\colon B \to C$ *when there exists an object* W *together with morphisms* $a\colon W \to A$ *and* $b\colon W \to B$ *such that the square*

$$\begin{array}{ccc} W & \xrightarrow{b} & B \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle g} \\ A & \xrightarrow{f} & C \end{array}$$

commutes, that is $f \circ a = g \circ b$*, and which is universal in the following way: for any object* W_o *with morphisms* $a_o\colon W_o \to A$ *and* $b_o\colon W_o \to B$ *such that* $f \circ a_o = g \circ b_o$*, there exist a unique morphism* $w\colon W_o \to W$ *such that* $a \circ w = a_o$ *and* $b \circ w = b_o$*.*

W *is then the limit of the diagram*

$$\begin{array}{ccc} & & B \\ & & \downarrow{\scriptstyle g} \\ A & \xrightarrow{f} & C \end{array}$$

and we write $W = A \times_C B$ *for the pullback. We also say that* W *is the pullback of* A *along the map* $g\colon B \to C$ *and also the pullback of* B *along the map* $f\colon A \to C$*.* W *is also the fiber product of the pair of morphisms* $f\colon A \to C$ *and* $g\colon B \to C$*.*

In analogy with the product of two sets $A \times B$ with the two projections $A \times B \to A$ and $A \times B \to B$, the morphisms a, b are often thought of as projection maps.

Definition 33 (Pushout) *A pushout for morphisms $f\colon C \to A$ and $g\colon C \to B$ is an object W together with morphisms $a\colon A \to W$ and $b\colon B \to W$ such that the square*

$$\begin{array}{ccc} C & \xrightarrow{g} & B \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle b} \\ A & \xrightarrow{a} & W \end{array}$$

commutes, that is $a \circ f = b \circ g$, and which is universal in the following way: for any object W_o with morphisms $a_o\colon A \to W_o$ and $b_o\colon B \to W_o$ such that $a_o \circ f = b_o \circ g$, there exist a unique morphism $w\colon W \to W_o$ such that $w \circ a = a_o$ and $w \circ b = b_o$.

W is then the colimit of the diagram

$$\begin{array}{ccc} C & \xrightarrow{g} & B \\ {\scriptstyle f}\downarrow & & \\ A & & \end{array}$$

and we write $W = A \cup_C B$ for the pushout.

We see that a pushout is the dual version of a pullback. The pushout in the category Set of two morphisms $\emptyset \to A$ and $\emptyset \to B$ is the standard disjoint union $A \coprod B$, labelling the composed set with which elements come from set A and which elements come from set B. This is also the coproduct in the category of sets. For another example, when we have a non-empty intersection $A \cap B \subseteq A$ and $A \cap B \subseteq B$ and $a\colon A \cap B \to A$ and $b\colon A \cap B \to B$ are the inclusions, the union $A \cup B$ is naturally isomorphic with the pushout of a and b. More generally, the morphisms a, b can be thought of as inclusions in keeping with these examples.

A.2 Pushforwards in Measure Theory

For a treatment of measure theory consult Billingsley (1986).

Definition 34 (Pushforward measure) *Given measurable spaces (Ω, Σ) and (X, T), a measure μ on (Ω, Σ) and a measurable function $\psi\colon (\Omega, \Sigma) \to (X, T)$, we write $\psi_\star \mu$ for the pushforward measure obtained from μ by applying ψ:*

$$\psi_\star \mu(A) \coloneqq \mu\left(\psi^{-1}(A)\right) \text{ for all } A \in T.$$

For example, for a deterministic function with random inputs, the pushforward measure gives us an explicit description of the possible distribution of outcomes.

A.3 Groups

For an in-depth treatment of algebra consult Aluffi (2021).

Definition 35 (Group) *A set G endowed with a binary operation $\bullet$, $(G, \bullet)$, is a group if the following conditions hold.*

1. *The operation* $\bullet$ *is associative; that is,* (for all $g, h, k \in G$): $(g \bullet h) \bullet k = g \bullet (h \bullet k)$.
2. *There exists an identity element* e_G *for* $\bullet$*; that is,*

$$(\text{there exists } e_G \in G)(\text{for all } g \in G) \colon g \bullet e_G = e_G \bullet g.$$

3. *Every element in* G *is invertible with respect to* $\bullet$*; that is,*

$$(\text{for all } g \in G)(\text{there exists } h \in G) \colon g \bullet h = h \bullet g = e_G.$$

Consider as an example the set of integers $\mathbb{Z}$:

1. Addition in $\mathbb{Z}$ is an associative operation.
2. The identity element is the integer 0.
3. The inverse map sends an integer $n \in \mathbb{Z}$ to another integer $-n$.

References

V. Abbott and G. Zardini. Diagrammatic negative information. arXiv:2404.03224 [math.CT], 2024.

D. Abel, D. Arumugam, L. Lehnert, and L. L. Littman. State abstractions for lifelong reinforcement learning. In *ICML*, 2018. URL `https://proceedings.mlr.press/v80/abel18a.html`.

M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu. Safe reinforcement learning via shielding. In *AAAI*, 2018. doi:10.1609/AAAI.V32I1.11797.

P. Aluffi. *Algebra: Chapter 0.* American Mathematical Society, 2021.

A. D. Ames. *A categorical theory of hybrid systems.* PhD thesis, University of California, Berkeley, 2006.

A. D. Ames, S. Mattenet, and J. Moeller. Categorical Lyapunov theory II: Stability of systems. *arXiv:2505.22968 [math.DS]*, 2025a.

A. D. Ames, J. Moeller, and P. Tabuada. Categorical Lyapunov theory I: Stability of flows. *arXiv:2502.15276 [math.DS]*, 2025b.

A. Asperti and G. Longo. *Categories, types, and structures: An introduction to category theory for the working computer scientist.* MIT Press, 1991.

J. C. Baez, B. Fong, and B. S. Pollard. A compositional framework for Markov processes. *Journal of Mathematical Physics*, 2016. doi:10.1063/1.4941578.

A. Bagaria, B. M. Abbatematteo, O. Gottesman, M. Corsaro, S. Rammohan, and G. Konidaris. Effectively learning initiation sets in hierarchical reinforcement learning. In *NeurIPS*, 2023. URL `http://papers.nips.cc/paper_files/paper/2023/hash/e8da56eb93676e8f60ed2b696e44e7dc-Abstract-Conference.html`.

G. Bakirtzis and U. Topcu. Categorical semantics of motion planning. In *ICRA*, 2022.

G. Bakirtzis, C. H. Fleming, and C. Vasilakopoulou. Categorical semantics of cyber-physical systems theory. *ACM Trans. Cyber Phys. Syst.*, 2021. doi:10.1145/3461669.

G. Bakirtzis, M. Savvas, R. Zhao, S. Chinchali, and U. Topcu. Reduce, reuse, recycle: Categories for compositional reinforcement learning. In *ECAI*, 2024. doi:10.3233/FAIA240797.

P. Billingsley. *Probability and Measure*. Wiley, 1986.

F. Bonchi, J. Holland, R. Piedeleu, P. Sobocinski, and F. Zanasi. Diagrammatic algebra: from linear to concurrent systems. In *POPL*, 2019. doi:10.1145/3290338.

A. Brandenburger and H. J. Keisler. Fiber products of measures and quantum foundation. In *Logic and Algebraic Structures in Quantum Computing*. Lecture Notes in Logic, 2016. doi:10.1017/CBO9781139519687.006.

S. Breiner, R. D. Sriram, and E. Subrahmanian. Compositional models for complex systems. In *Artificial Intelligence for the Internet of Everything*. Elsevier, 2019. doi:10.1016/B978-0-12-817636-8.00013-2.

D. Cicala. Spans of cospans. *Theory and Applications of Categories*, 2018.

B. Coecke. Compositionality as we see it, everywhere around us. In *The Quantum-Like Revolution: A Festschrift for Andrei Khrennikov*. Springer, 2023. doi:10.1007/978-3-031-12986-5_12.

J. Culbertson, P. Gustafson, D. E. Koditschek, and P. F. Stiller. Formal composition of hybrid systems. *Theory and Applications of Categories*, 2020.

C. Devin, A. Gupta, T. Darrell, P. Abbeel, and S. Levine. Learning modular neural network policies for multi-task and multi-robot transfer. In *ICRA*, 2017. doi:10.1109/ICRA.2017.7989250.

T. G. Dietterich. Hierarchical reinforcement learning with the MAXQ value function decomposition. *JAIR*, 2000. doi:10.1613/jair.639.

Z. Diskin and T. S. E. Maibaum. Category theory and model-driven engineering: From formal semantics to design patterns and beyond. In *ACCAT*, 2012. doi:10.4204/EPTCS.93.1.

Z. Diskin, H. König, and M. Lawford. Multiple model synchronization with multiary delta lenses with amendment and k-putput. *Formal Aspects of Computing*, 2019. doi:10.1007/S00165-019-00493-0.

S. Du, S. Kakade, J. Lee, S. Lovett, G. Mahajan, W. Sun, and R. Wang. Bilinear classes: A structural framework for provable generalization in RL. In *ICML*, 2021. URL `https://proceedings.mlr.press/v139/du21a.html`.

T. Fritz, T. Gonda, P. Perrone, and E. F. Rischel. Representable Markov categories and comparison of statistical experiments in categorical probability. *Theoretical Computer Science*, 2023. doi:10.1016/j.tcs.2023.113896.

B. Gavranovic, P. Lessard, A. J. Dudzik, T. von Glehn, J. G. Madeira Araújo, and P. Velickovic. Position: Categorical deep learning is an algebraic theory of all architectures. In *ICML*, 2024. URL https://openreview.net/forum?id=EIcxV7T0Sy.

F. Genovese. Modularity vs compositionality: A history of misunderstandings. URL https://perma.cc/B5SZ-G9JW, 2018. Statebox.

M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology and Analysis*, 1981.

I. Gur, N. Jaques, Y. Miao, J. Choi, M. Tiwari, H. Lee, and A. Faust. Environment generation for zero-shot compositional reinforcement learning. In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/218344619d8fb95d504ccfa11804073f-Paper.pdf.

T. Hanks, B. She, E. Patterson, M. Hale, M. Klawonn, and J. P. Fairbanks. Modeling model predictive control: A category theoretic framework for multistage control problems. In *ACC*. IEEE, 2024. doi:10.23919/ACC60939.2024.10644848.

J. Hedges and R. R. Sakamoto. Reinforcement learning in categorical cybernetics. arXiv:2404.02688 [cs.LG], 2024.

R. Ivanov, K. Jothimurugan, S. Hsu, S. Vaidya, R. Alur, and O. Bastani. Compositional learning and verification of neural network controllers. *ACM Transactions on Embedded Computing System*, 2021. doi:10.1145/3477023.

K. Jothimurugan, S. Bansal, O. Bastani, and R. Alur. Compositional reinforcement learning from logical specifications. *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper/2021/file/531db99cb00833bcd414459069dc7387-Paper.pdf.

M. Klissarov, A. Bagaria, Z. Luo, G. Konidaris, D. Precup, and M. C. Machado. Discovering temporal structure: An overview of hierarchical reinforcement learning. *arXiv:2506.14045 [cs.AI]*, 2025.

G. Konidaris. On the necessity of abstraction. *Current Opinion in Behavioral Sciences*, 2019. doi:10.1016/j.cobeha.2018.11.005.

F. W. Lawvere and S. H. Schanuel. *Conceptual mathematics: a first introduction to categories*. Cambridge University Press, 2009.

J. Leike, M. Martic, V. Krakovna, P. A. Ortega, T. Everitt, A. Lefrancq, L. Orseau, and S. Legg. AI safety gridworlds. arXiv:1711.09883 [cs.LG], 2017.

T. Leinster. *Basic Category Theory*. Cambridge University Press, 2014.

E. Lerman. Networks of open systems. *Journal of Geometry and Physics*, 2018. doi:10.1016/j.geomphys.2018.03.020.

E. Lerman and J. Schmidt. Networks of hybrid open systems. *Journal of Geometry and Physics*, 2020. doi:10.1016/j.geomphys.2019.103582.

L. Li, T. J. Walsh, and M. L. Littman. Towards a unified theory of state abstraction for MDPs. In *AI&M*, 2006. URL http://anytime.cs.umass.edu/aimath06/proceedings/P21.pdf.

Y. Li, Y. Wu, H. Xu, X. Wang, and Y. Wu. Solving compositional reinforcement learning problems via task reduction. In *ICLR*, 2021. URL https://openreview.net/forum?id=9SS69KwomAM.

M. Luckcuck, M. Farrell, L. A. Dennis, C. Dixon, and M. Fisher. Formal specification and verification of autonomous robotic systems: A survey. *ACM Computing Surveys*, 2019. doi:10.1145/3342355.

S. Mac Lane. *Categories for the working mathematician*. Springer, 1998.

J. A. Mendez, M. Hussing, M. Gummadi, and E. Eaton. CompoSuite: A compositional reinforcement learning benchmark. In *CoLLAs*, 2022. URL https://proceedings.mlr.press/v199/mendez22a.html.

A. Mohan, A. Zhang, and M. Lindauer. Structure in reinforcement learning: A survey and open problems. *Journal of Artificial Intelligence Research*, 2024. doi:10.1613/jair.1.15703.

O. Nachum, S. S. Gu, H. Lee, and S. Levine. Data-efficient hierarchical reinforcement learning. *NeurIPS*, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/e6384711491713d29bc63fc5eeb5ba4f-Abstract.html.

C. Neary, C. K. Verginis, M. Cubuktepe, and U. Topcu. Verifiable and compositional reinforcement learning systems. In *ICAPS*, 2022. URL https://ojs.aaai.org/index.php/ICAPS/article/view/19849.

J. Ok, A. Proutiere, and D. Tranos. Exploration in structured reinforcement learning. *NeurIPS*, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/d693d554e0ede0d75f7d2873b015f228-Abstract.html.

R. Parr and S. Russell. Reinforcement learning with hierarchies of machines. *NeurIPS*, 1997. URL http://papers.nips.cc/paper/1384-reinforcement-learning-with-hierarchies-of-machines.

H. H. Pattee. The complementarity principle in biological and social structures. *Journal of Social and Biological Structures*, 1978. doi:10.1016/S0140-1750(78)80007-4.

P. Perny, O. Spanjaard, and P. Weng. Algebraic Markov decision processes. In *IJCAI*, 2005. URL http://ijcai.org/Proceedings/05/Papers/1677.pdf.

B. C. Pierce. *Basic category theory for computer scientists*. Foundations of computing. MIT Press, 1991.

B. Ravindran. Relativized hierarchical decomposition of Markov decision processes. *Progress in Brain Research*, 2013. doi:10.1016/B978-0-444-62604-2.00023-X.

B. Ravindran and A. G. Barto. SMDP homomorphisms: An algebraic approach to abstraction in semi-Markov decision processes. In *IJCAI*, 2003. URL http://ijcai.org/Proceedings/03/Papers/145.pdf.

B. Ravindran and A. G. Barto. Approximate homomorphisms: A framework for non-exact minimization in Markov decision processes. In *KBCS*, 2004.

P. Schultz, D. I. Spivak, C. Vasilakopoulou, and R. Wisnesky. Algebraic databases. *Theory & Applications of Categories*, 2016.

P. Schultz, D. I. Spivak, and C. Vasilakopoulou. Dynamical systems and sheaves. *Applied Categorical Structures*, 2019. doi:10.1007/s10485-019-09565-x.

D. I. Spivak. *Category theory for the sciences*. MIT Press, 2014.

The Stacks project authors. The Stacks project. https://stacks.math.columbia.edu, 2024.

W. Sun, N. Jiang, A. Krishnamurthy, A. Agarwal, and J. Langford. Model-based RL in contextual decision processes: PAC bounds and exponential improvements over model-free approaches. In *COLT*, 2019. URL http://proceedings.mlr.press/v99/sun19a.html.

R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT Press, 2018.

J. M. Swart. A conditional product measure theorem. *Statistics & Probability Letters*, 1996. doi:10.1016/0167-7152(95)00107-7.

Z. Szabó. The case for compositionality. *The Oxford Handbook of Compositionality*, 2012.

G. N. Tasse, S. D. James, and B. Rosman. A Boolean task algebra for reinforcement learning. In *NeurIPS*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6ba3af5d7b2790e73f0de32e5c8c1798-Abstract.html.

G. N. Tasse, S. James, and B. Rosman. Generalisation in lifelong reinforcement learning through logical composition. In *ICLR*, 2022. URL https://openreview.net/forum?id=ZOcX-eybqoL.

E. Todorov. Compositionality of optimal control laws. In *NeurIPS*, 2009. URL https://proceedings.neurips.cc/paper/2009/hash/3eb71f6293a2a31f3569e10af6552658-Abstract.html.

E. van der Pol, D. E. Worrall, H. van Hoof, F. A. Oliehoek, and M. Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. In *NeurIPS*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2be5f9c2e3620eb73c2972d7552b6cb5-Abstract.html.

B. van Niekerk, S. D. James, A. Christopher Earle, and B. Rosman. Composing value functions in reinforcement learning. In *ICML*, 2019. URL http://proceedings.mlr.press/v97/van-niekerk19a.html.

K. Watanabe, C. Eberhart, K. Asada, and I. Hasuo. Compositional probabilistic model checking with string diagrams of MDPs. In *CAV*, 2023. doi:10.1007/978-3-031-37709-9_3.

Z. Wen, D. Precup, M. Ibrahimi, A. Barreto, B. Van Roy, and S. Singh. On efficiency in hierarchical reinforcement learning. In *NeurIPS*, 2020. URL `https://proceedings.neurips.cc/paper/2020/hash/4a5cfa9281924139db466a8a19291aff-Abstract.html`.

G. Zardini, D. Milojevic, A. Censi, and E. Frazzoli. Co-design of embodied intelligence: A structured approach. In *IROS*, 2021. doi:10.1109/IROS51168.2021.9636513.