Why Some Actions Cannot Be Chosen: On the Architectural Necessity of Authorization Before Optimization

MxBv

Abstract

Dominant agent architectures explain action execution through choice: actions are selected by optimizing preferences, utilities, or expected outcomes. This framing implicitly assumes that all actions are admissible and differ only in value.

This work establishes a minimal but fundamental negative result: some actions cannot be chosen. If an action with irreversible structural consequences is executed, its admissibility must have been resolved prior to any act of choice. Choice therefore cannot serve as a sufficient explanatory primitive for action execution in long-horizon systems.

We formalize the distinction between authorization and choice. Authorization determines whether an action is permitted to occur at all; choice determines which permitted action is selected. Authorization precedes optimization and operates independently of reward, belief, preference, or evaluation.

By isolating this ordering constraint, we identify a missing architectural layer in dominant agent theories. This result does not propose a new agent model, algorithm, or control method. It defines a boundary condition that any viable agent architecture must satisfy in the presence of irreversibility and path dependence.

The contribution of this work is therefore architectural rather than algorithmic: it establishes that optimization-based explanations presuppose an authorization layer and cannot replace it. Recognizing this layer clarifies why certain classes of actions cannot be explained by choice alone and why irreversibility imposes limits that no amount of optimization can overcome.

1 Introduction

Action is commonly explained through choice. An agent observes the world, evaluates possible actions, selects one according to a criterion, and executes it. This narrative underlies rational choice theory, control systems, planning algorithms, reinforcement learning, and most contemporary agentic frameworks.

Despite their differences, these approaches share a structural assumption: that the execution of an action is the result of a selection process among alternatives. The primary disagreements concern how alternatives are evaluated, not whether they are admissible.

This work challenges that assumption. We argue that for a non-trivial class of actions, execution cannot be explained by choice at all. If an action occurs, its permissibility must have been resolved prior to any optimization or deliberation. Choice presupposes admissibility.

The goal of this paper is not to propose a new agent model, but to demonstrate a structural omission in existing ones.

2 Action Selection in Standard Architectures

Let $\mathcal{S}$ denote a state space, $\mathcal{A}$ an action space, and $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ a transition operator.

In standard formulations, action execution at time t is defined as

$$a_t = \arg\max_{a \in \mathcal{A}} \mathbb{E}[U(\mathcal{T}(s_t, a))],$$

or some variant thereof. The defining feature is that all actions are assumed to be eligible for consideration, differing only in expected value.

Safety constraints, risk measures, or penalties are typically incorporated by modifying the objective or augmenting the state. Crucially, admissibility is treated as a consequence of evaluation rather than as a prior condition.

This structure implicitly equates *not chosen* with *not permitted*. We will show that this equivalence fails.

3 Irreversibility and the Limits of Choice

Consider actions whose execution induces irreversible structural consequences. Such actions alter the future action space itself, not merely the current state.

Examples include commitments that collapse alternative futures, identity- affecting decisions, regime transitions, or actions that permanently alter constraints under which the system operates.

For these actions, delayed execution or post-hoc correction is not possible. Once executed, the structural impact persists regardless of future optimization.

In such settings, treating action execution as the result of choice leads to a paradox: the act of choosing presupposes the very structural conditions that the action will destroy.
Encoding irreversibility into state or utility does not resolve the paradox, because choice mechanisms still presuppose an admissible action set over which comparison is defined.

4 Authorization Versus Choice

We introduce a strict architectural distinction between authorization and choice.

Authorization determines whether an action is permitted to occur at all. Choice determines which permitted action is selected. This distinction is not semantic, but causal and structural.

Formally, let

$$G : (\mathcal{S}, \mathcal{A}, \tau) \to \{0, 1\}$$

be an authorization function, where τ indexes internal system evolution. An action a is admissible at (s, τ) if and only if $G(s, a, \tau) = 1$.

Choice mechanisms operate exclusively over the admissible action set

$$\mathcal{A}_{\text{adm}}(\tau) = \{a \in \mathcal{A} \mid G(s_\tau, a, \tau) = 1\}.$$

Crucially, authorization is not a function of information, belief, preference, expected outcome, or uncertainty. It does not evaluate actions, rank alternatives, or resolve trade-offs. Its role is strictly to determine whether deliberation is allowed to occur at all.

Authorization therefore precedes choice not only temporally but logically: choice presupposes the existence of admissible alternatives, whereas authorization determines whether such alternatives exist.

Authorization cannot be reconstructed from choice behavior. Two systems may exhibit identical observable action selection while differing fundamentally in their authorization structure. Any explanation that infers admissibility from observed choice commits a category error by collapsing a pre-decisional structural boundary into a post-decisional evaluative process.

5 A Negative Result

Proposition. There exists a class of actions for which execution cannot be explained as the result of choice under any utility function.

Argument. Choice operates over alternatives. Authorization determines whether alternatives exist.

For actions whose execution entails irreversible structural consequences, the admissibility of the action must be resolved prior to any comparison, evaluation, or ranking. If an action collapses future admissible space, then the act of comparing it to alternatives already presupposes that its execution is permitted.

Therefore, execution implies prior authorization. An action with irreversible consequences cannot originate from choice, because choice presupposes a set of admissible alternatives, whereas authorization determines whether such a set exists at all.

This ordering is structural, not psychological or computational. It does not depend on the form of optimization, the depth of planning, the representation of uncertainty, or the expressiveness of the utility function. It follows solely from the temporal and causal ordering of irreversible structural operations. Any account that attributes the origin of such actions to choice alone implicitly assumes admissibility and therefore fails to explain it.

6 Implications for Agent Theories

The negative result established above imposes a boundary condition on agent architectures. Any system capable of executing actions with irreversible structural consequences must include a mechanism that determines admissibility independently of optimization, evaluation, or preference.

Without such a mechanism, action execution itself cannot be coherently explained. Optimization presupposes a space of admissible alternatives; it cannot account for the origin or collapse of that space. As a result, architectures that treat all actions as inherently admissible implicitly assume away irreversibility rather than addressing it.

This does not invalidate optimization-based methods. It restricts their domain of applicability. Optimization governs behavior *within* admissible space; it does not define that space.

Frameworks that conflate admissibility with evaluation, risk, or confidence implicitly encode irreversibility as a soft preference or penalty. Such encodings fail when structural consequences are non-recoverable and path-dependent, because no amount of post hoc optimization can restore collapsed future admissibility.

7 Relation to Existing Work

The distinction identified here is orthogonal to constrained optimization, safe reinforcement learning, shielded control, and risk-sensitive planning.

Those approaches modify how actions are evaluated, selected, or penalized *within* a predefined action space. Authorization, by contrast, determines whether evaluation is permitted to occur at all.

In constrained or safe optimization, inadmissibility is represented as a cost, penalty, or constraint applied during evaluation. In the present formulation, inadmissible actions are excluded prior to any evaluation, comparison, or ranking.

As a result, the authorization layer cannot be reduced to constraints, penalties, confidence thresholds, or belief updates without loss of essential structural information. Such reductions necessarily presuppose admissibility and therefore fail to model the conditions under which admissibility itself is established or withdrawn.

Any framework that explains irreversible action execution solely in terms of optimization implicitly assumes that authorization has already occurred, and therefore does not address the origin of admissibility itself.

8 Conclusion

This work establishes a minimal but fundamental claim: some actions cannot be chosen.

In systems operating over long horizons with irreversible structural consequences, action execution cannot be explained solely in terms of choice, preference, or optimization. If an action is executed, its admissibility must have been resolved prior to any act of choice. Choice therefore operates only over an already authorized space and cannot serve as a sufficient explanatory primitive for action execution.

By isolating this negative result, we identify a missing architectural layer in dominant agent theories. Authorization precedes optimization. This ordering is not a design preference or modeling convenience; it is a structural necessity imposed by irreversibility and path dependence.

The result presented here does not propose a new agent model, learning algorithm, or decision procedure. It defines a boundary condition that any viable agent architecture must respect. Architectures that conflate admissibility with evaluation implicitly assume away irreversibility and therefore fail to explain how irreversible actions can coherently occur.

The broader implication is that many observed failures in large-scale adaptive systems are not failures of optimization or intelligence, but failures of architectural ordering. Without an explicit authorization layer, systems are forced to encode irreversibility implicitly, where it becomes invisible, untestable, and unrecoverable.

Recognizing authorization as prior to choice does not solve decision-making. It establishes the conditions under which decision-making is possible at all.