

PCT Controller solution: MountainCarContinuous-v0: Description and Analysis

Author: Rupert Young
email: rupert@perceptualrobots.com
web: www.perceptualrobots.com
20 October 2026

Abstract

This document examines one verified hierarchical Perceptual Control Theory controller for Gymnasium's MountainCarContinuous-v0. The controller was generated through an evolution process. Across 100 unique reset seeds, it succeeded in 100 of 100 episodes (100%), required 76-93 episode steps, and achieved total reward ranging from 93.520 to 95.756. Its architecture and histories show a compact two-level feedback system that first drives left to accumulate momentum, then reverses to sustained rightward force and reaches the goal with positive velocity.

1. PCT Introduction

Perceptual Control Theory (PCT) explains behavior as the control of perception through negative feedback [1] [2]. A control unit receives a **reference signal** representing a desired perceptual state and a **perceptual signal** derived from the environment. A comparator subtracts perception from reference to produce an **error signal**. An output function converts error into an influence on the environment; environmental feedback then changes the next perceptual signal. The closed loop acts to limit the discrepancy between reference and perception despite disturbances [1] [2].

In hierarchical PCT, higher-level output signals set references for lower-level control units. Lower units regulate faster or more concrete perceptions, while higher units organize them toward broader goals [2]. This organization has been developed as a general architecture for artificial-life and robotics systems [3] and applied to control in real-world robotic settings [4]. In the controller examined here, a higher position-control unit supplies references to two lower velocity-control units, whose outputs jointly determine the force applied to the car.

2. Environment and Control Objective

The environment observes car position in $[-1.2, 0.6]$ and velocity in $[-0.07, 0.07]$. A scalar continuous action in $[-1, 1]$ applies force. The episode terminates when position reaches at least 0.45 with non-negative velocity; otherwise it is truncated after 999 steps. The car cannot climb directly from its randomized start. Successful behavior requires motion toward the left slope, accumulation of kinetic energy, force reversal, and a rightward climb.

The controller therefore has two coupled responsibilities: regulate perception relative to the goal while producing the temporarily goal-opposing action needed to build momentum. The histories below show how its feedback structure resolves that apparent conflict.

3. Controller Overview

It was selected from experiments based upon episode length for successful controllers. The verified [Comet experiment](#) [5] contains the controller configuration. A [YouTube replay video](#) [6] shows the controller completing the task as well as the progressive development of the controller through the generations of the evolutionary process.

Property	Verified value
Hierarchy	[2, 1]: two lower control units and one top-level unit
Top-level perception	Position
Top-level reference	0.45, fixed
Lower-level perceptions	Velocity only (both lower units)
Perception/reference functions	Smooth at level 0; linear at level 1

Property	Verified value
Output/action functions	Smooth outputs; action clipped to [-1, 1]
Verified replay	Seed 3986596453; success in 83 steps; total reward 94.6808

Table 1. Verified configuration of the selected controller.

The top unit perceives position with unit weight and compares it with the fixed reference 0.45. Its output gain of 0.7278 supplies two descending lower references through weights 2.0945 and -3.4060. Both lower units perceive velocity only. Their output gains, -1.4884 and 1.6989, combine into the action through weights -1.5757 and 0.6212. Different smoothing factors give the lower paths distinct temporal responses, allowing the weighted action to reverse without a hard-coded clock.

4. Controller Architecture

Figure 1 shows the PCT hierarchical architecture of the controller. The environment provides position and velocity observations, but they are routed to different control levels: position supplies the single top-level perceptual signal, while velocity supplies each of the two lower-level perceptual signals. At the top level, the position perception is compared with the fixed 0.45 reference. The resulting comparator error passes through the top output function, and the labelled descending arrows apply signed weights to create distinct references for the two lower velocity-control units.

Each lower comparator subtracts its velocity-only perception from its descending reference. Its error signal passes through a smoothed output function; the diagram's labels record the gains, connection weights, signs, and smoothing or time-response parameters that make the two pathways respond differently. The two lower outputs then follow weighted arrows to the action-combination node. Their sum is clipped to the environment's [-1, 1] action range and applied as continuous force. The resulting movement changes position and velocity, closing the feedback paths shown by the arrows from the environment back to the perceptual functions.

4.1 Diagram Reading Guide

Diagram element	Meaning
Vertical arrangement	Higher-level position control above two lower velocity-control pathways
P, R, C, and O node labels	Perception, reference, comparator, and output functions within each control unit
Observation routing	Position is routed to the top perception; velocity is routed to both lower perceptions
Descending arrows	Top output supplies separately weighted references to the lower comparators
Numeric labels and signs	Connection weights and gains determine magnitude and direction of influence
Smooth labels	Time-response functions shape how errors become outputs and how action changes
Action and feedback arrows	Weighted lower outputs form force; movement changes observations and closes the loop

Table 2. Reading guide for the architecture diagram.

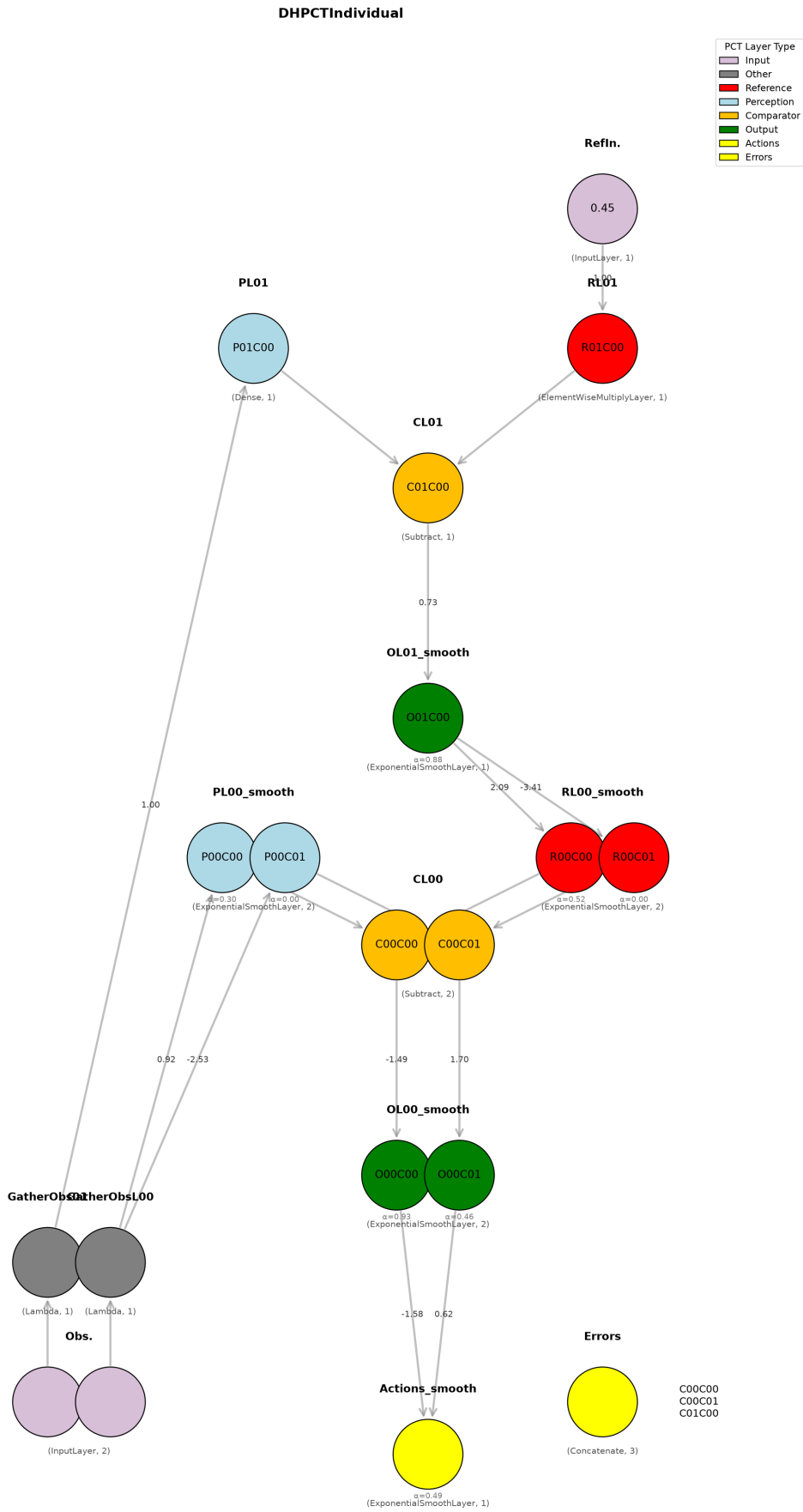


Figure 1. PCT hierarchical architecture of the selected controller.

5. Controller Behavior

The following histories come from the verified successful rollout at seed 3986596453. Each plot describes a different part of the same closed-loop event.

5.1 Observations

Figure 2 plots both environment observations over the episode. These traces describe environment state, not the perceptual input of every control unit: position is perceived at the top level, whereas velocity alone is perceived by both lower units. Position first moves left, reverses after the car has accumulated momentum, and then rises past 0.45 at step 83. Velocity is initially negative, crosses zero at the leftmost position, and becomes positive during the rightward climb. Moving away from the goal is therefore a functional momentum-building phase, not loss of control.

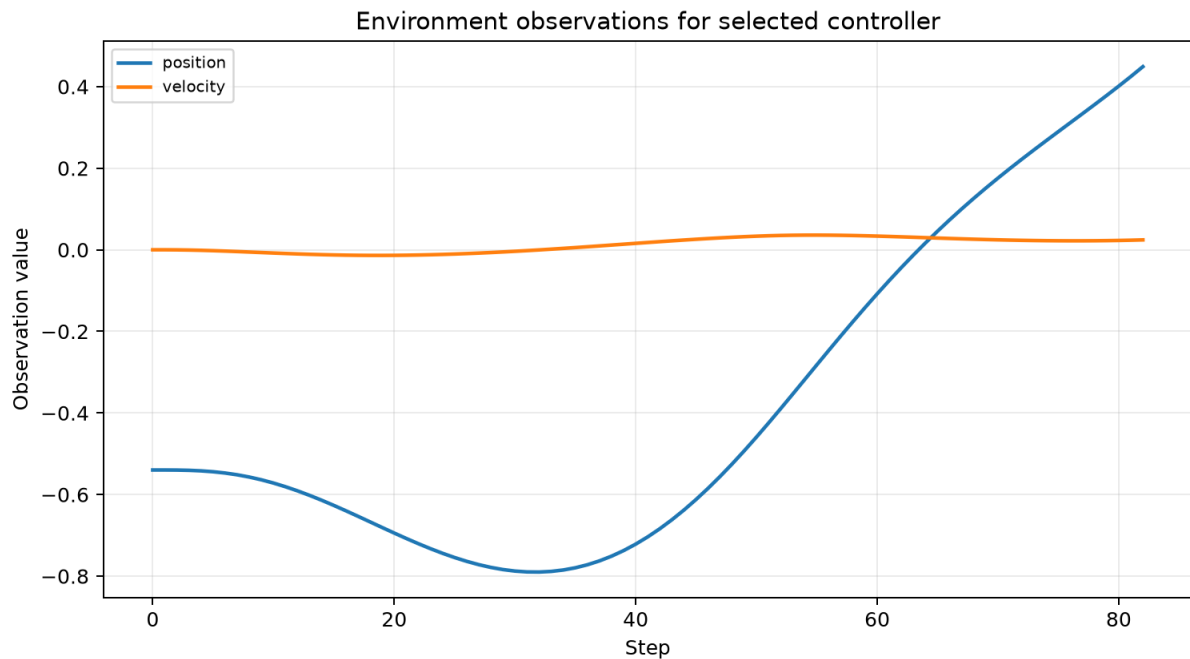


Figure 2. Environment position and velocity during the verified successful rollout.

5.2 Reference and Perception

Figure 3 shows the top reference fixed at 0.45 and the top perception tracking position. Perception initially moves away from the reference, so their difference grows while kinetic energy accumulates. After the turning point, perception rises rapidly toward the reference. This demonstrates that successful regulation can include a temporary increase in discrepancy when the environment requires momentum.

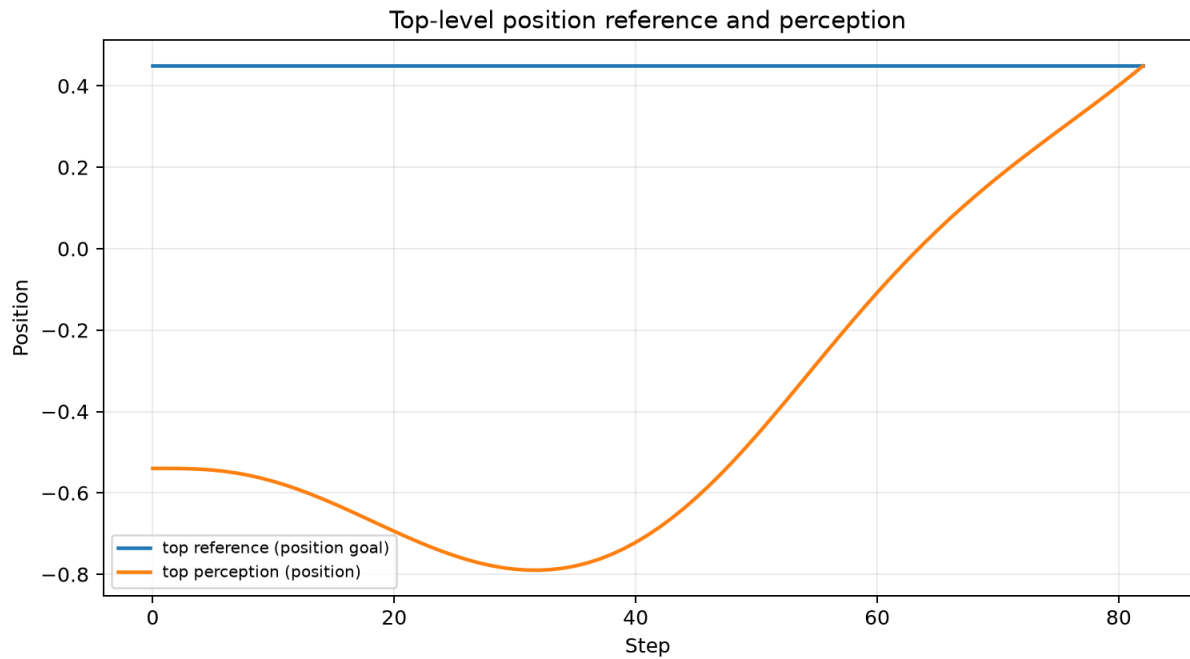


Figure 3. Fixed top-level position reference and controlled position perception.

5.3 Comparator Errors

Figure 4 contains the two lower comparator errors and the top-level error. The top error rises from about 0.87 to about 1.34 near the leftmost position and then falls toward zero. Each lower comparator error is formed from its descending reference and a velocity-only perception; position does not contribute directly to either lower perceptual signal. The lower errors differ in scale and timing because their descending-reference weights, gains, signs, and smoothing parameters differ. Those unequal error signals create the differential drive needed for the later action reversal.

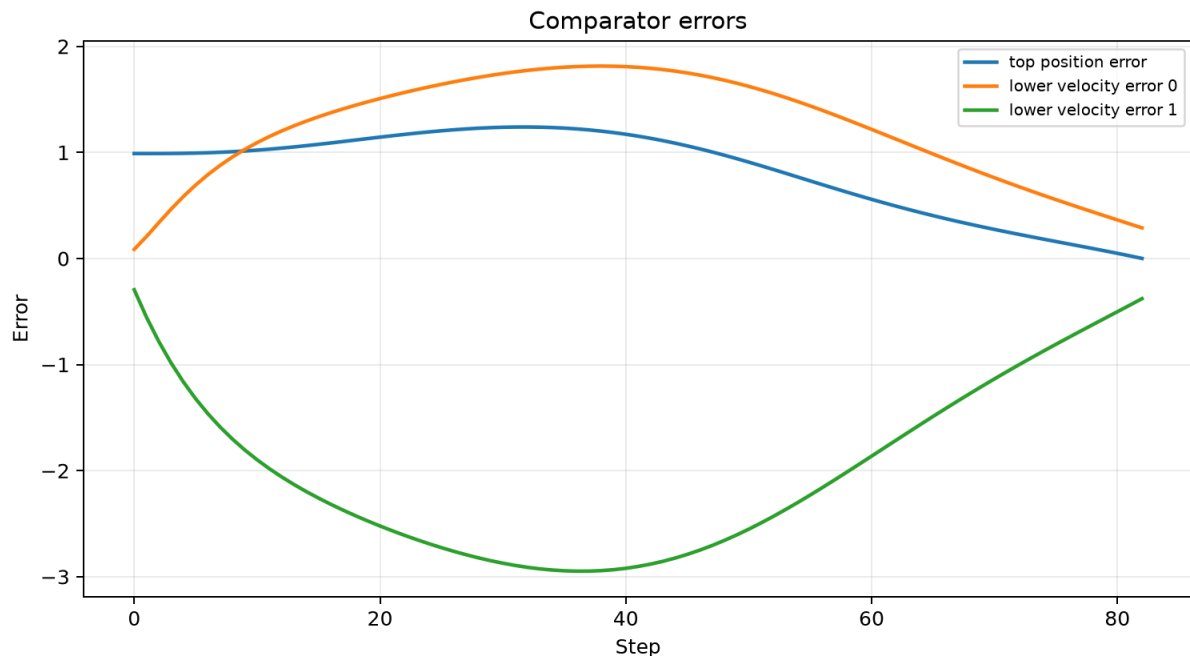


Figure 4. Comparator errors: two lower velocity-control errors and the top position-control error.

5.4 Lower-Level Outputs

Figure 5 plots the two smoothed lower outputs. Their different gains, signs, and time constants shift the weighted balance during the early momentum-building phase. The force reversal emerges from the interaction of these feedback outputs rather than from a predetermined switching time.

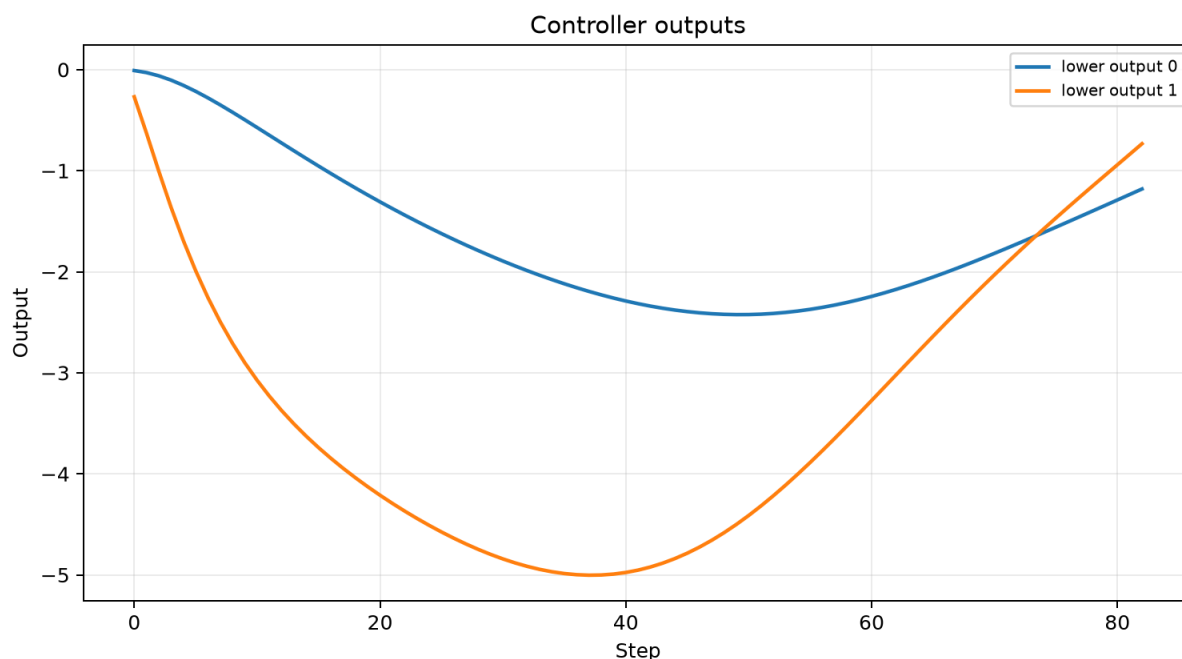


Figure 5. Smoothed outputs from the two lower velocity-control units.

5.5 Action

Figure 6 shows action initially directed left, then crossing into sustained rightward force for the climb. Negative force first drives the car up the left slope; positive force then converts stored momentum into the climb. Action clipping preserves the environment bounds even though internal outputs may have larger magnitudes.

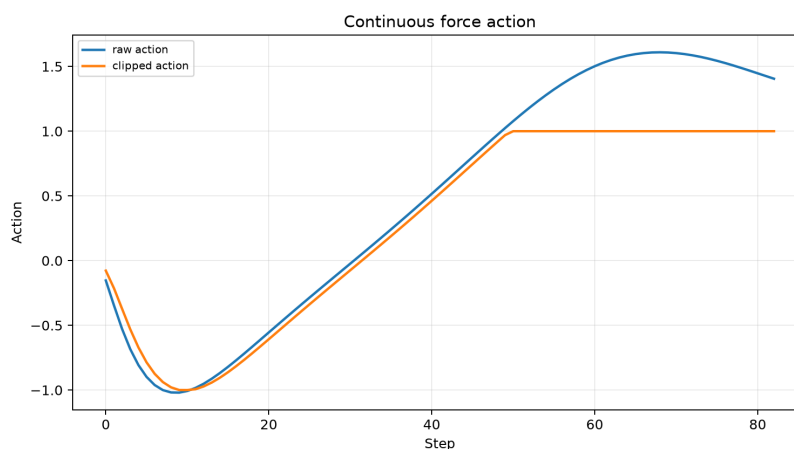


Figure 6. Continuous force action and its momentum-building sign reversal.

5.6 Reward

Figure 7 shows the per-step action cost and terminal goal bonus. Saturated force costs about -0.1 per step, which is visually small beside the terminal bonus near 100. The controller expends control effort to obtain the goal bonus, producing total reward 94.6808 in this rollout. The trace confirms that the force sequence ends in task success rather than merely reducing internal error.

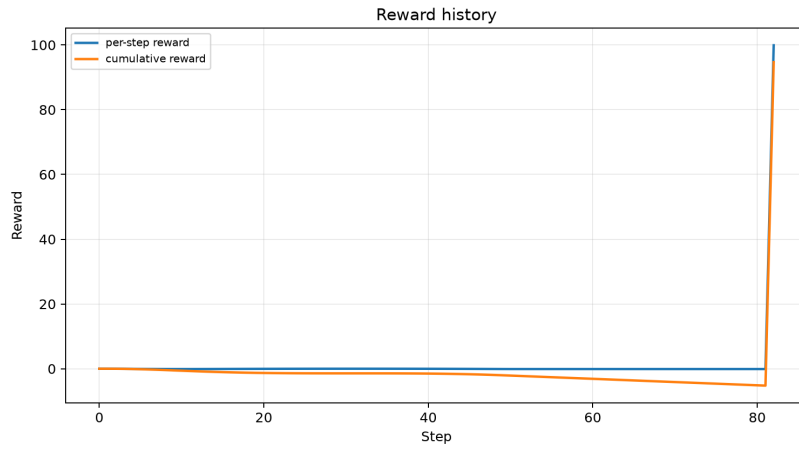


Figure 7. Per-step action cost and terminal goal bonus.

6. Replay Performance and Robustness

The replay audit explicitly assigned 100 unique seeds from 3986596453 through 3986596552. Every episode terminated successfully. Episode length ranged from 76 to 93 steps, and every total reward exceeded 93.520, showing that reset variation changes timing without altering the behavioral strategy.

Metric	Verified value
Unique seeds	3986596453-3986596552 / 100
Successful episodes	100/100 (100%)
Episode length	80.65 +/- 4.58 steps
Episode length median / range	79 / 76-93 steps
Total reward	95.051 +/- 0.615
Total reward median / range	95.228 / 93.520-95.756

Table 3. Verified 100-seed replay performance.

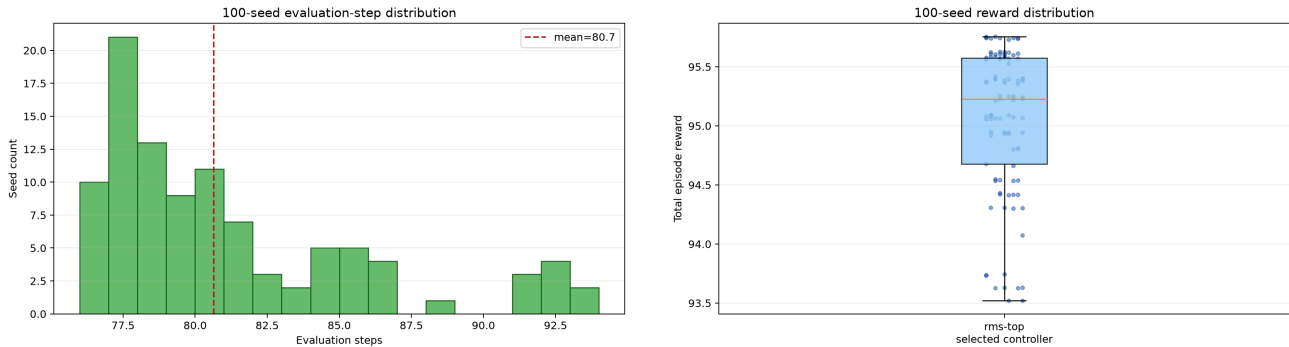


Figure 8. Episode length (left) and total reward (right) across 100 unique reset seeds.

Episode length is concentrated between 76 and 93 steps, and total reward spans 93.520 to 95.756. The distributions support a single robust control mechanism rather than seed-specific trajectories.

7. Behavioral Interpretation

The controller succeeds through four linked stages. First, the fixed top reference represents the target position while actual position supplies the top perception. Second, the top output establishes unequal lower references for two velocity-control units. Third, their velocity-only comparator errors drive weighted, smoothed outputs that apply strong left force and then reverse to strong right force. Fourth, position approaches 0.45, top error declines, and the car crosses the goal with positive velocity.

The central behavioral feature is tolerance of temporary top-level error growth. A purely immediate response would push toward the goal and fail to build enough energy. Here, hierarchical feedback and smoothing transform the growing discrepancy during the leftward descent into the delayed action reversal required by the

mountain-car dynamics. Complete success across the tested resets shows that this mechanism is stable across starting positions.

8. Conclusions

The selected controller is a compact, interpretable hierarchy of three PCT units: one top-level position-control unit and two lower velocity-control units. Its architecture exposes the complete path from references and perceptions through comparator errors and smoothed outputs to action. The histories explain a momentum-building leftward phase, a feedback-driven force reversal, and a sustained rightward climb.

Across 100 unique seeds, the controller achieved 100% success in 76-93 steps with mean total reward 95.051. The architecture, histories, and replay results therefore provide mutually consistent evidence that the same feedback mechanism controls the task robustly.

Data and Reproducibility

The controller configuration is available from the verified [Comet experiment](#). The following command replays the controller with the installed `dpct - run` package:

```
dpct-run "https://www.comet.com/mountaincarcontinuous-v0/rms-top/cef6e0ad66a344b5930f42758077a39d"
--run --steps 999 --render --early-termination
--history-dir /tmp/dpct-pipeline/reports/MountainCarContinuous-v0/replay --history-graphs all
```

The command downloads or reuses the experiment's `best_individual.json`, renders the controller in its configured environment with the stated step limit, and exports the controller architecture and rollout histories to the specified output directory.

References

- [1] Powers, W. T., Clark, R., & McFarland, R. (1960). *A general feedback theory of human behavior: Part I. Perceptual and Motor Skills*, 11(1), 71-88.
- [2] Powers, W. T. (1973). *Behavior: The control of perception*. Aldine de Gruyter.
- [3] Young, R. (2017). *A general architecture for robotics systems: A perception-based approach to artificial life*. *Artificial Life*, 23(2), 236-286.
- [4] Young, R. (2020). *Robotics in the real world: The perceptual control theory approach*. In W. Mansell (Ed.), *The Interdisciplinary Handbook of Perceptual Control Theory* (Chapter 14, pp. 517-556). Academic Press.
- [5] Comet. [Selected controller experiment and configuration](#).
- [6] Young, R. [MountainCarContinuous-v0: DPCT Evolution and Best Controller](#). YouTube.

Note

This report was generated by AI and supervised by a human.