

Beyond Dodge Rate: Testing Humanoid Dodgeball Under Camera Failures

Supplementary Material

Kingston Kuan Brandon Ong
Hebbian Robotics

<https://hebbian-robotics.github.io/beyond-dodge-rate/>

APPENDIX A ADDITIONAL EVALUATION DETAILS

This supplement accompanies *Beyond Dodge Rate: Testing Humanoid Dodgeball Under Camera Failures*. The main paper reports the primary recovery test, camera-input comparison, shared-throw effects, and their uncertainty. The details below describe the exploratory control and recorded-data checks.

A. Shared ball trajectories

For trajectory index $i = 0, \dots, 63$, donor seed $9000 + i$ supplies a healthy-condition throw from π_1 when i is even and π_2 when i is odd. This selects 32 trajectories per donor policy without filtering outcomes. Simulation seeds 10000–10063 replay the pool with both policies and all five camera conditions. These are 640 exploratory trials, not newly sampled physical trajectories or independent training runs.

At launch, the evaluator retains native collision settings, accounting, and random draws, then replaces only the ball’s initial position and velocity before integration. Exact throw parameters match across all ten policy-condition combinations per seed. Robot states can differ between policies. The target label describes the donor’s original aim, not the receiving robot’s body part. Twenty preflight trials use two separate simulation seeds. Two additional healthy-condition replays of the first control seed reproduce complete trial records and verify live ball position and velocity before integration.

B. Trajectory and uncertainty checks

Hits resolve at native contact; misses resolve when the simulator ends the throw. Resolution is not a closest-approach measurement. Area exits use 50 Hz samples without interpolation; a tie with resolution counts as a by-resolution exit. The 1 m score fails at the first sampled exit even if the robot later returns. For stale-input restoration pairs, launch states, throws, camera schedules, and all physical samples before restoration match exactly. Every stale-input throw resolves by two seconds within numerical tolerance.

Paired intervals resample complete evaluation seeds with all their conditions. The shared-throw input-effect difference retains both policies in each resampled seed. Bootstrap runs use 10,000 resamples and seed 20261014. Intervals are nominal per comparison, not multiplicity-adjusted. The trial archives

TABLE I
EXPLORATORY SHARED-THROW CONTROL: THE SAME 64 RECORDED TRAJECTORIES FOR BOTH POLICIES. ALL FIVE CONDITIONS ARE INCLUDED; FALLS AND ABORTED THROWS ARE ZERO. RECOVERY RESUMES FRESH IMAGES AT TWO SECONDS. ROBOT STATES CAN DIFFER AT LAUNCH.

Policy	Camera input	Dodge	In area	Success
π_1	Healthy	61/64	64/64	61/64
	Stale	60/64	27/64	23/64
	Stale + recovery	60/64	64/64	60/64
	Blank	57/64	60/64	53/64
	Blank + recovery	57/64	62/64	55/64
π_2	Healthy	51/64	60/64	47/64
	Stale	52/64	58/64	46/64
	Stale + recovery	52/64	57/64	45/64
	Blank	41/64	62/64	39/64
	Blank + recovery	41/64	61/64	38/64

retain per-condition Wilson intervals, camera schedules, physical traces, outcome flags, and matching checks. The original 2,496-trial package includes exploratory cohorts as well as the fixed 192-seed input test. Recovery uses a separate 160-trial pilot and a 640-trial fixed cohort. Exploratory controls and post-hoc scoring analyses are not pooled into that fixed estimate.

APPENDIX B DEMO: CAMERA FAILURES IN SIMULATED HUMANOID DODGEBALL

A. What viewers see

The demo presents synchronized replays of the experiments in this paper. Viewers compare camera conditions, pause at measured events, and follow the robot’s pelvis relative to the 1 m boundary. The examples use the first declared seeds; aggregate results are reported separately so that an illustrative trial is not mistaken for average performance.

The overview video first compares healthy depth with camera freezes at ball launch and 320 ms later. In the illustrated trial, the later freeze allows a dodge but the robot subsequently leaves the area. The video then compares restoration of fresh images, followed by healthy, stale, and blank images for π_2 . The opening slide defines π_1 and π_2 and identifies their sources; separate tables report the 64-seed fixed recovery test

and the 192-seed input comparison. These are two sets of learned weights trained with the same method and recorded settings, not two proposed training methods.

The recovery replay compares continuously stale images with fresh images restored at two or three seconds. All three robots dodge the same throw; only the two-second restoration stays within the area in this example. The three-second example is labeled exploratory. The paper’s aggregate recovery results, rather than this single replay, support the reported effect.

B. Playback and setup

The videos use aligned views and synchronized holds so that the encounter and later movement can be compared. An orange marker remains briefly after the ball’s native removal to make its last recorded position visible. The marker is a display aid, not a simulated rebound or continued ball physics.

The demo consists of recorded simulation, rather than live policy execution. A laptop with a browser is sufficient for presentation. The presenters bring the laptop; a power outlet and an optional external display are useful. Downloaded videos play offline. No physical robot, GPU, or special floor area is needed. Online playback of the YouTube version requires internet access.

C. Video

The overview video is available at <https://youtu.be/MG4ftBPVOY>. Synchronized replays are available on the project website linked below the paper title.

The results concern two fixed policies in one simulated G1 task. They do not establish hardware safety or variation across independent training runs.

D. Code, data and weight availability

Evaluation code and configurations are available in the project’s public repository, linked from the project website. The four recorded-trial archives are prepared but have not been publicly released. The released reference weights can be obtained from their original source; the separately trained weights are not distributed. Recomputing the reported statistics requires the trial archives. Rerunning all simulations also requires the simulator assets and both sets of weights. Public end-to-end reproduction is therefore not currently available.