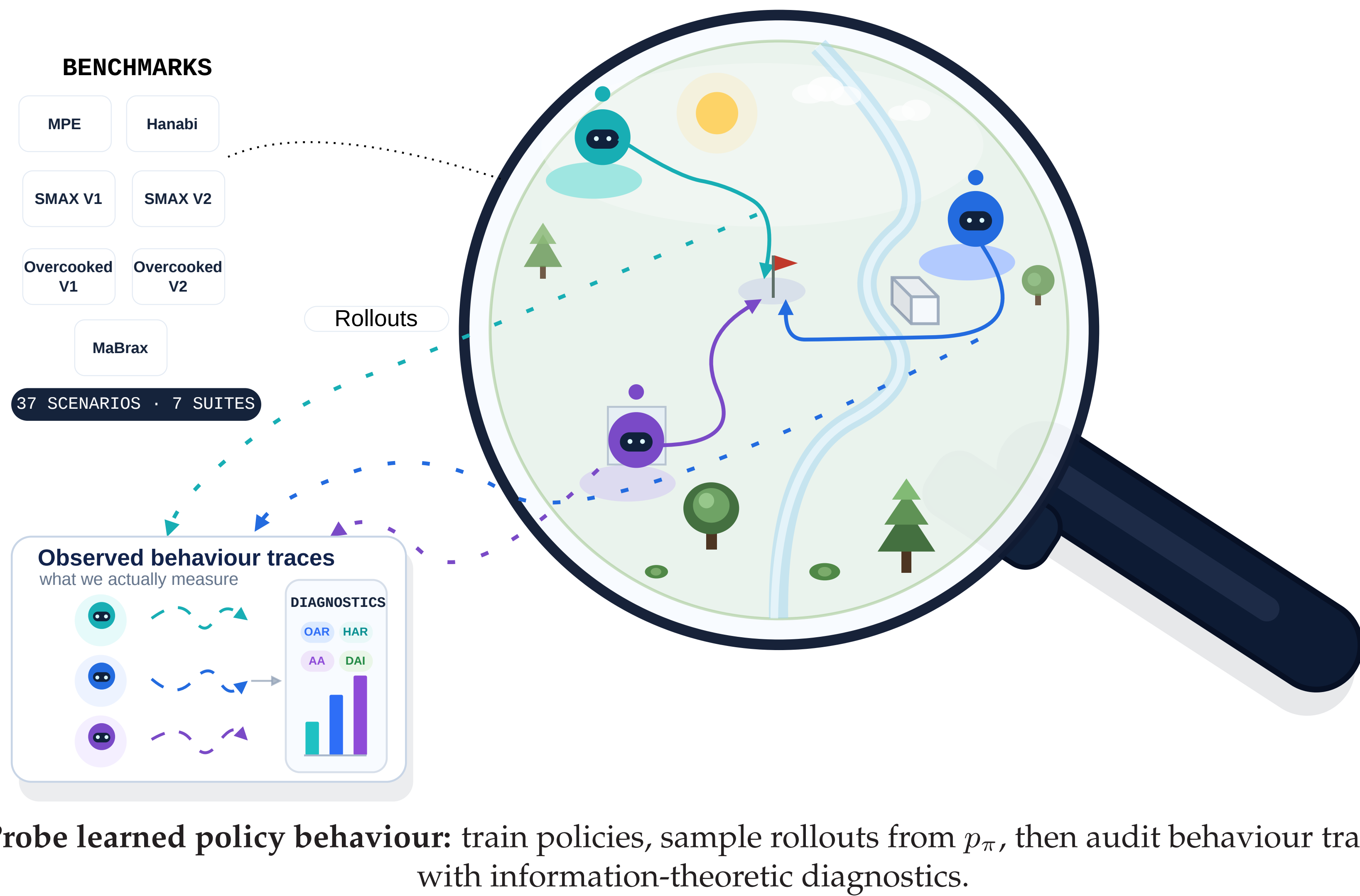




Do modern cooperative MARL environments truly test the Dec-POMDP properties that make these problems hard, or do they permit success via strategies that bypass them?

## 1) PROBING DEC-POMDP REASONING



## 2) TL;DR AND FINDINGS

### TL;DR

Across 37 cooperative MARL scenarios, IPPO/MAPPO policies often succeed without the Dec-POMDP reasoning these tasks are meant to test.

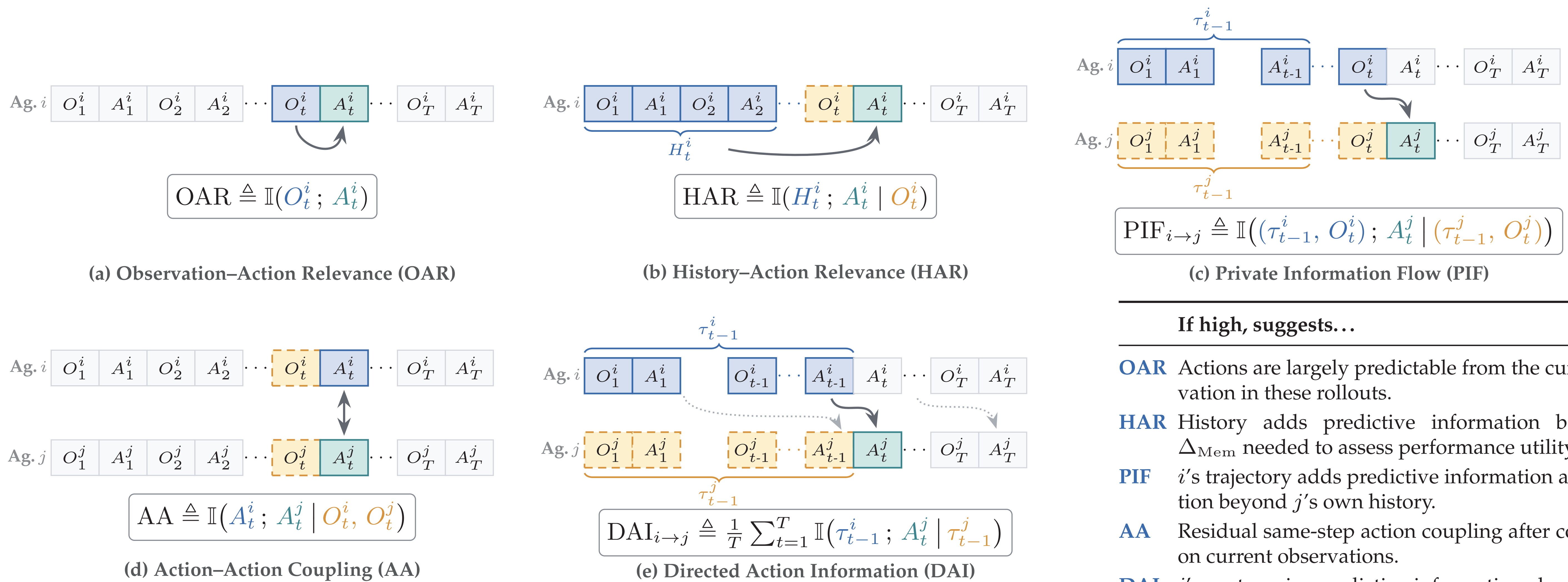
### MAIN FINDINGS

- **History dependence  $\neq$  utility:** all policies show some history dependence, but only 43% need memory for high returns.
- **Hidden state vs. teammate info:** probes disentangle hidden state and hidden teammate information as separate difficulty drivers (e.g., Overcooked V1  $\rightarrow$  V2).
- **Coordination structure varies:** synchronous and temporal mechanisms dissociate across benchmarks.

### TAKEAWAY

Design benchmarks where **partial observability** and **decentralised coordination** are non-optional for success. We provide metrics and an open-source library that can help with this.

## 3) METHODOLOGY: INFORMATION-THEORETIC PROBES



## 4) BENCHMARK AUDIT: 37 SCENARIOS, 7 SUITES

	MPE	SMAX V1	SMAX V2	MaBrax	Hanabi	Overcooked V1	Overcooked V2
Do agents benefit from memory?	100% (3/3)	100% (9/9)	100% (3/3)	20% (1/5)	0% (0/1)	0% (0/5)	0% (0/11)
Do agents use hidden teammate information?	100% (3/3)	67% (6/9)	67% (2/3)	100% (5/5)	0% (0/1)	20% (1/5)	82% (9/11)
Does synchronous coordination emerge?	100% (3/3)	44% (4/9)	0% (0/3)	60% (3/5)	0% (0/1)	100% (5/5)	82% (9/11)
Does temporal coordination emerge?	100% (3/3)	67% (6/9)	67% (2/3)	100% (5/5)	100% (1/1)	40% (2/5)	100% (11/11)

MPE is the only suite where *every* scenario satisfies all four criteria; elsewhere, history dependence and the two coordination modes dissociate across benchmarks.

## FINAL TAKEAWAYS AND CAVEATS

- We measure **how** policies solve tasks, not just **how well**, across 37 cooperative MARL scenarios.
- Under standard training, success often does not require the intended Dec-POMDP reasoning.
- Future benchmarks should make meaningful history use and decentralised coordination *non-optional*.

### LIMITATIONS

- Probes are policy-dependent (expectations under  $p_\pi$ ).
- MI/CMI/DI estimators carry finite-sample bias. Permutation nulls and bootstrap CIs mitigate but do not remove this.