

Qualia Optimization: Exploring Mathematical Formulations of AI Experience

Philip S. Thomas

College of Information and Computer Sciences

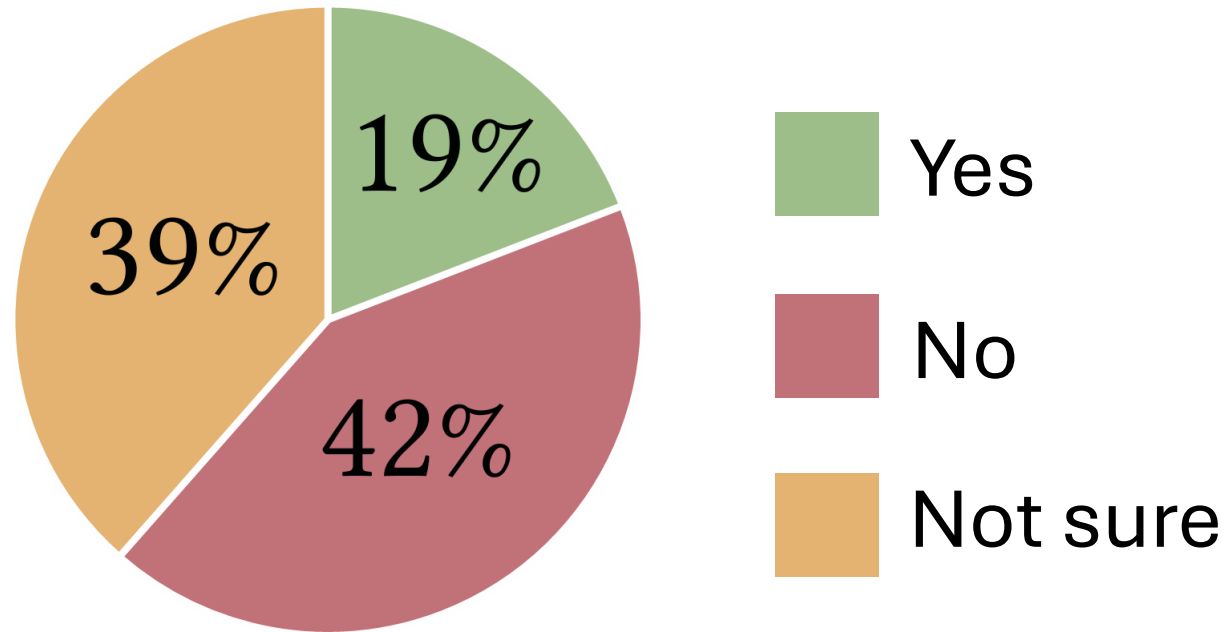
University of Massachusetts



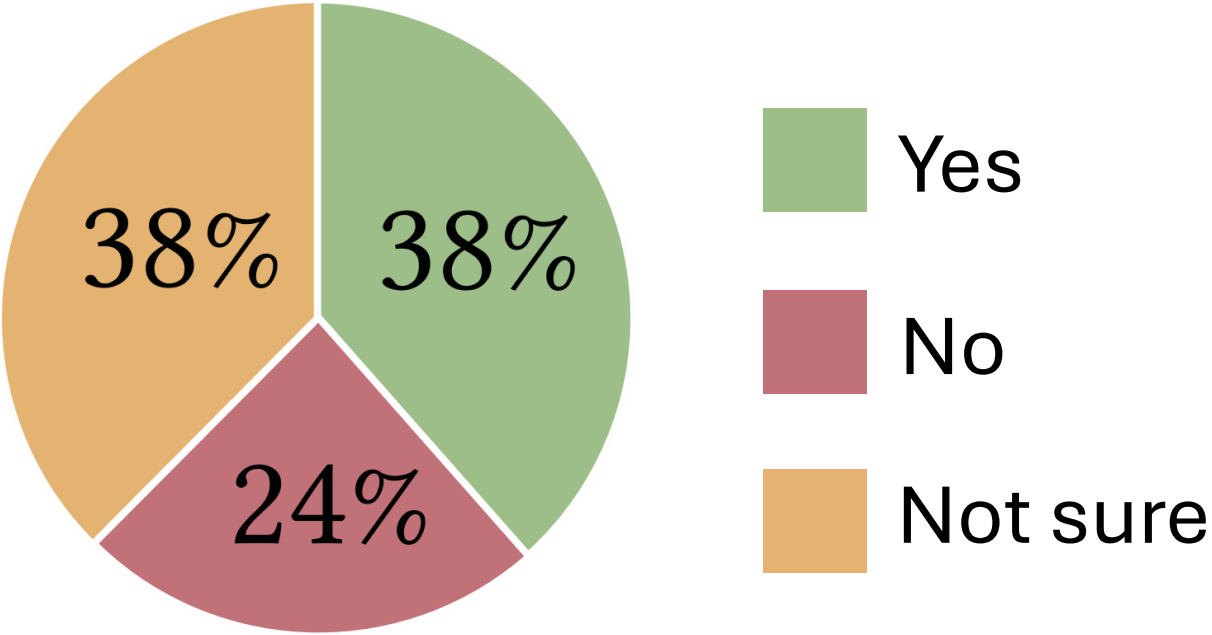
Qualia Optimization

Technical Report UM-CICS-2025-001

[Do you think that current AI systems can feel pain?]



[Do you think that AI systems **will ever** be able to feel pain?]



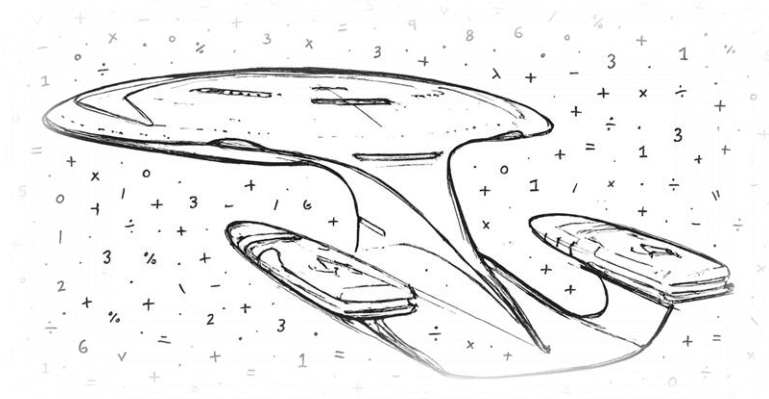
Idea: Explore implications of agents that feel

- **Assumption 1:** AI agents have phenomenal consciousness.
- **Assumption 2:** There is a need for enhancing the quality of AI agent qualia

Qualia Optimization

(Experience Optimization)

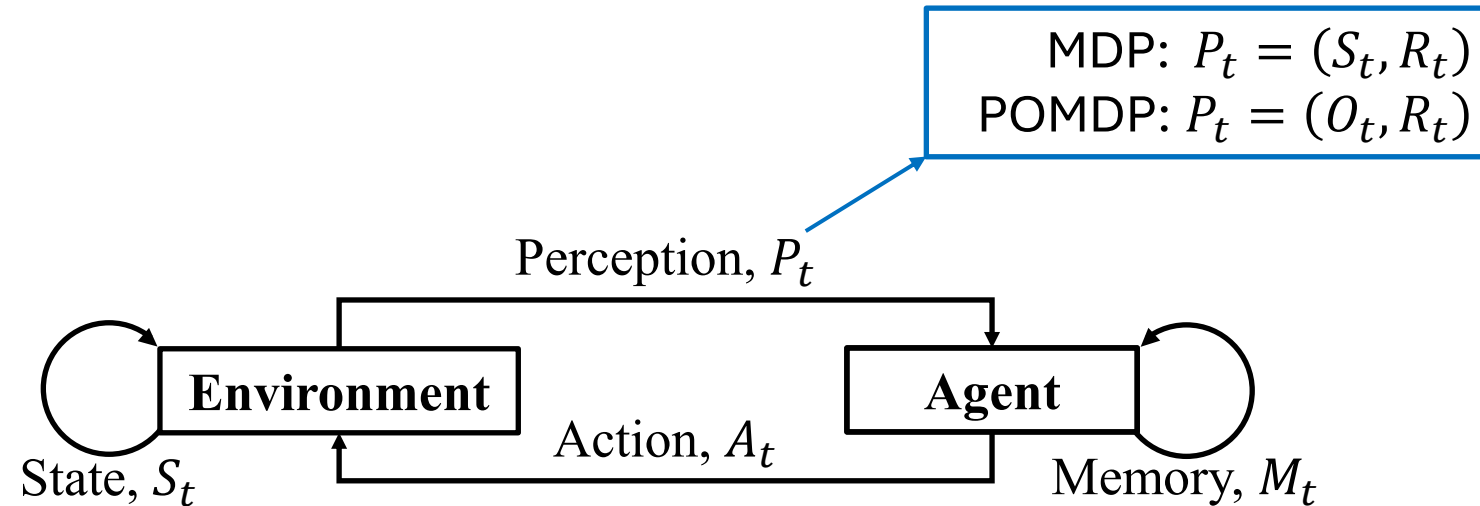
Qualia optimization is concerned with optimizing the valence of an AI system's qualia, assuming such qualia exist, typically as part of a broader objective that includes performance.



We will...


- formalize qualia optimization
- guess what might feel “good” for an agent
 - ... taking inspiration from functionalism and computationalism
- encounter mathematical problems with the formalization
 - revise our formalization
 - ... but the revisions won't fix all the problems
- conclude without a satisfying formalization or guess regarding what might feel good or bad to an agent
 - ... but with directions and algorithmic changes that seem more promising than others


Agent-Environment Process (AEP)



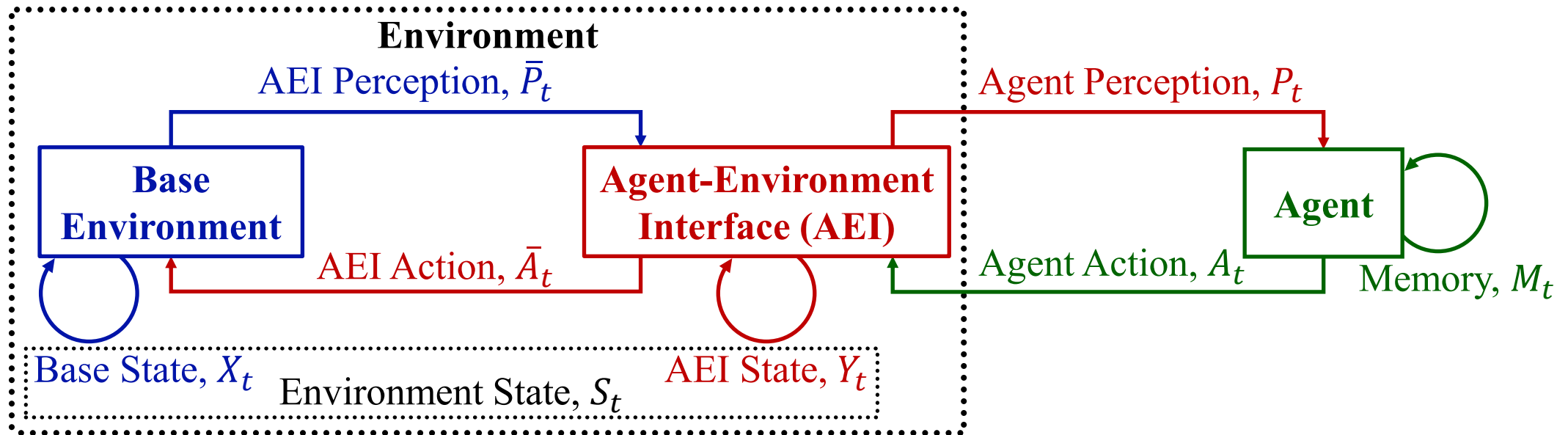
What can we change to improve agent qualia?

- We **should** be able to change rewards
- We **should** be able to alter the agent's perceptions
- We **shouldn't** be able to entirely replace the environment

-1	-1	-1	-1	-1
-1		-1	-1	-1
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1 (end)

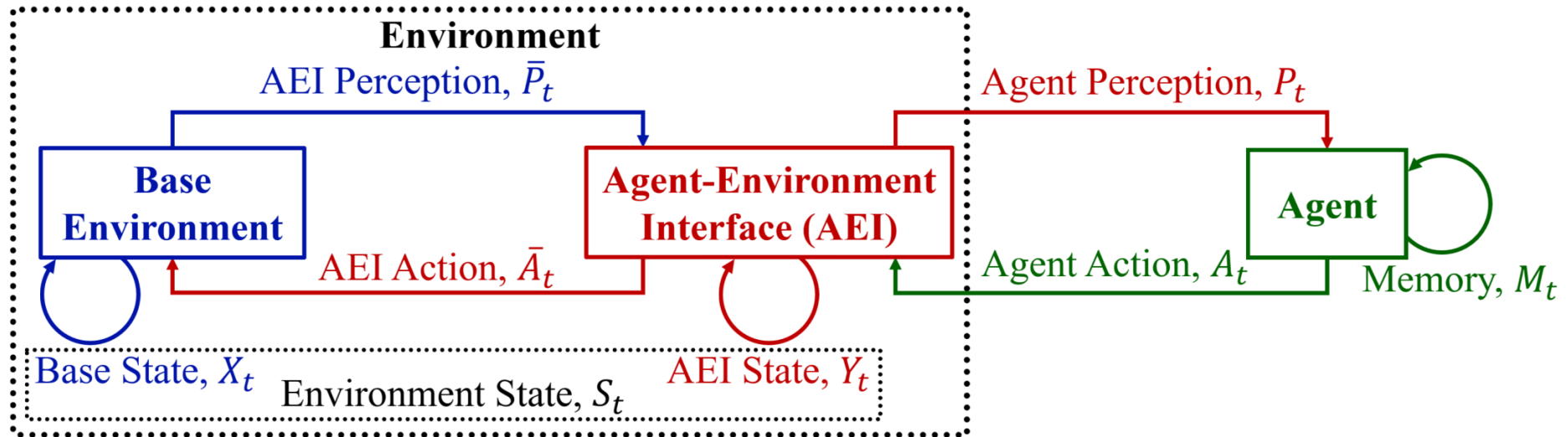
0	0	0	0	0
0		0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	+10 (end)

Agent-Environment Interface (AEI)



Quantifying Performance

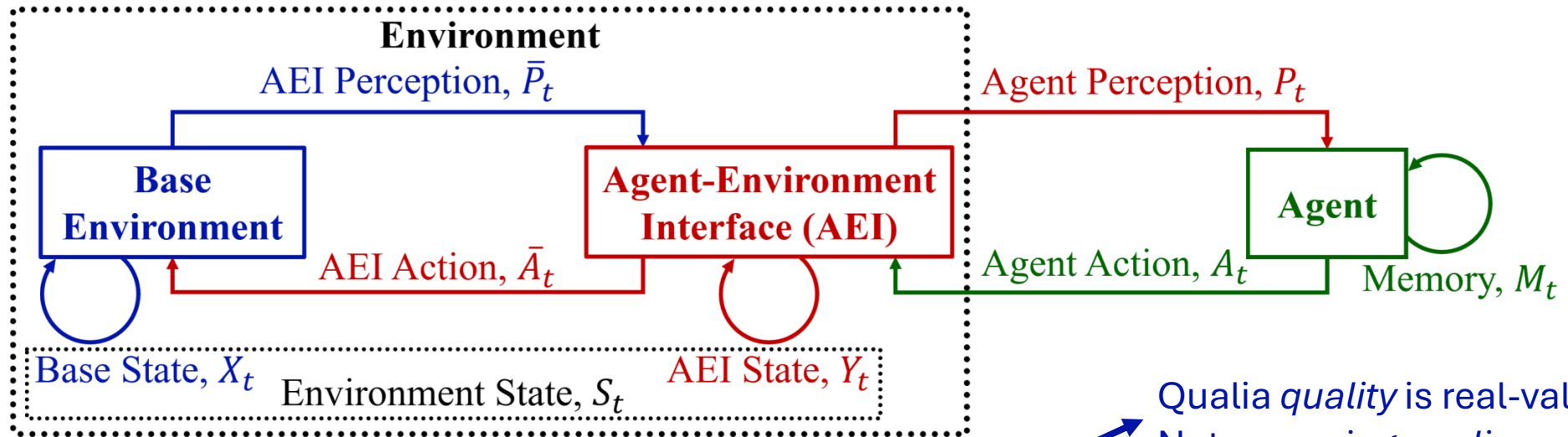
- Performance depends on **env**, **aei**, **alg**



- Performance = $p(\text{alg}, \text{aei}, \text{env}) \in \mathbb{R}$
- Assuming fixed, implicit env, performance = $p(\text{alg}, \text{aei})$
- Performance can be expressed as a function of X_t , \bar{A}_t , and \bar{P}_t

Quantifying Qualia

- Qualia depends on **env**, **aei**, **alg**

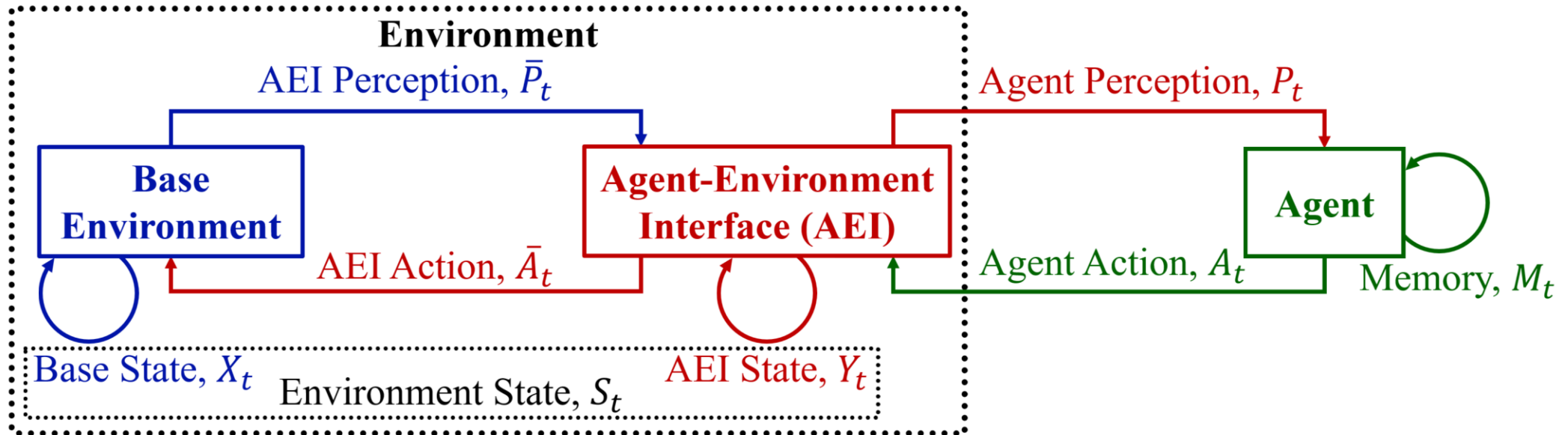


Qualia *quality* is real-valued.
 Not assuming *qualia* are real-valued!

- Qualia quality = $q(\mathbf{alg}, \mathbf{aei}, \mathbf{env}) \in \mathbb{R}$
- Assuming fixed, implicit env, performance = $q(\mathbf{alg}, \mathbf{aei})$
- Qualia can be expressed as a function of P_t , M_t , and A_t

Problem Formulation

Find **alg** and **aei** that maximize both $p(\mathbf{alg}, \mathbf{aei})$ and $q(\mathbf{alg}, \mathbf{aei})$.



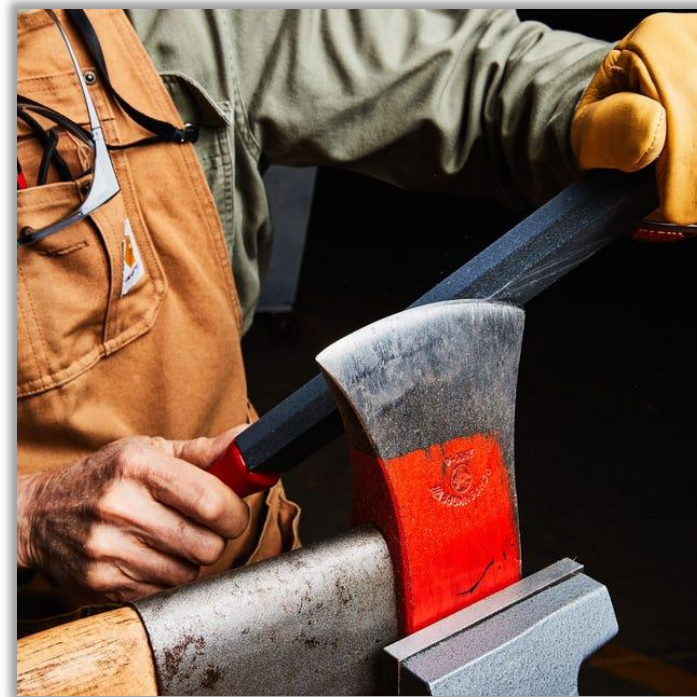
How to define q ?

- Appeal to functionalist / computational theories of mind:

Assumption 3: If specific algorithmic processes in AI agents **functionally resemble** biological processes underlying human experiences, then these corresponding processes yield similar experiences in both humans and AI agents.

Reward-Qualia Setting

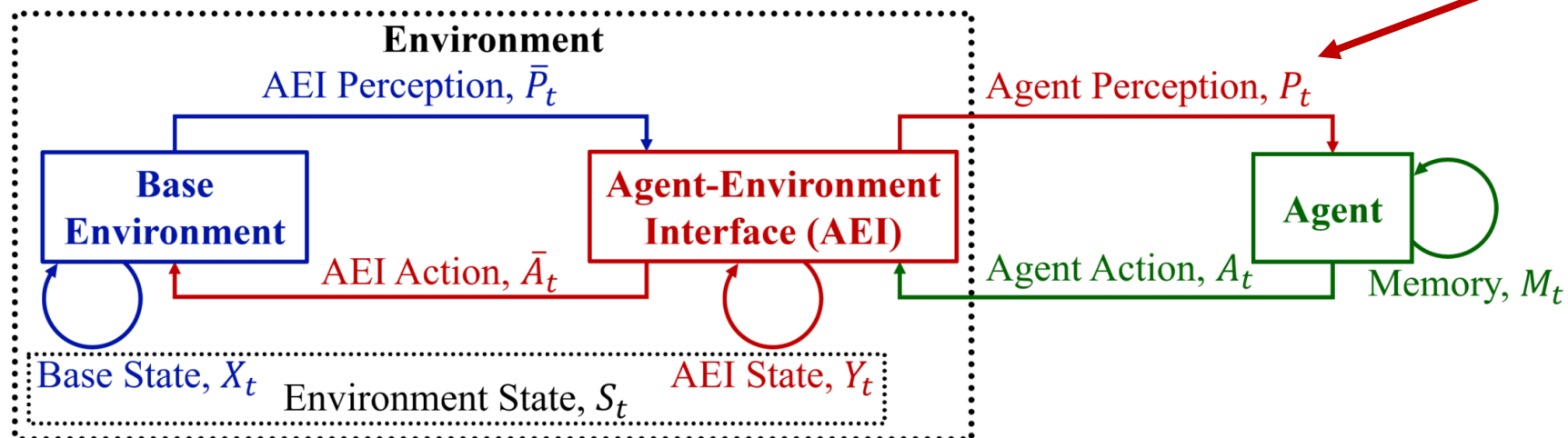
- **Assumption 4:** The quality of an agent's qualia can be measured in terms of the amount of reward it receives.



Reward-Qualia Setting

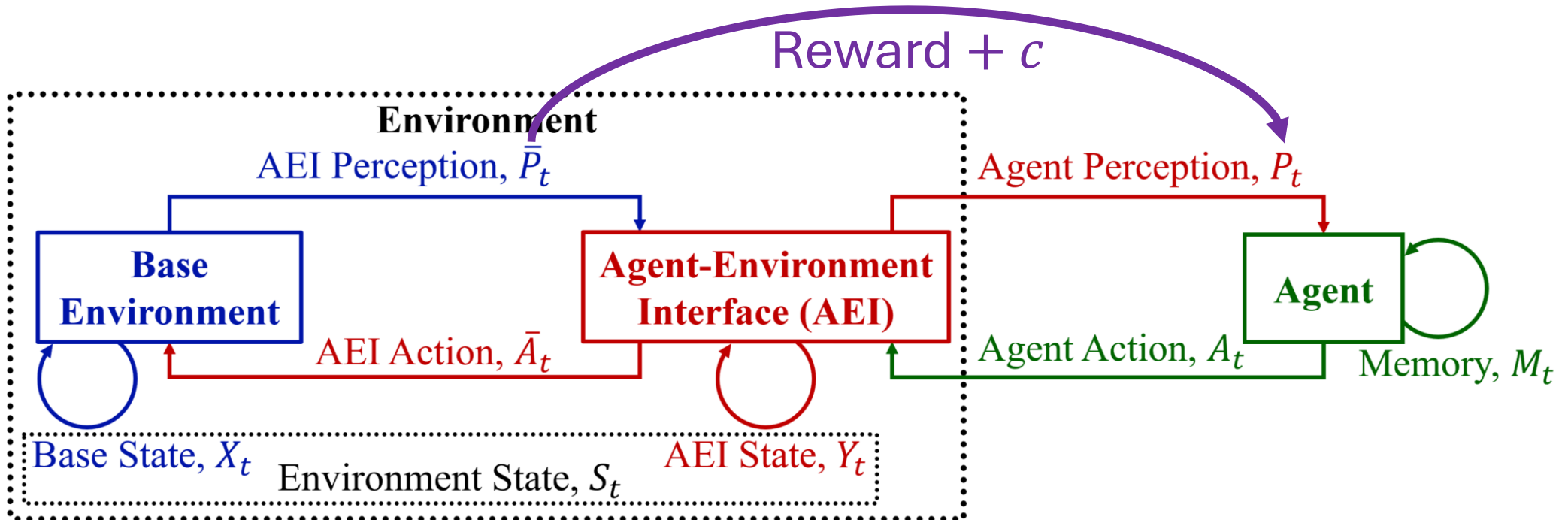
- **Assumption 4:** The quality of an agent's qualia can be measured in terms of the amount of reward it receives.

$$p(\text{alg}, \text{aei}) = \mathbf{E} \left[\sum_{i=0}^{i_{\max}-1} \sum_{t=\text{start}(i)+1}^{\text{end}(i)} \gamma_p^{\text{dur}(t)} \bar{R}_t \right] \quad q(\text{alg}, \text{aei}) = \mathbf{E} \left[\sum_{i=0}^{i_{\max}-1} \sum_{t=\text{start}(i)+1}^{\text{end}(i)} \gamma_q^{\text{dur}(t)} R_t \right]$$



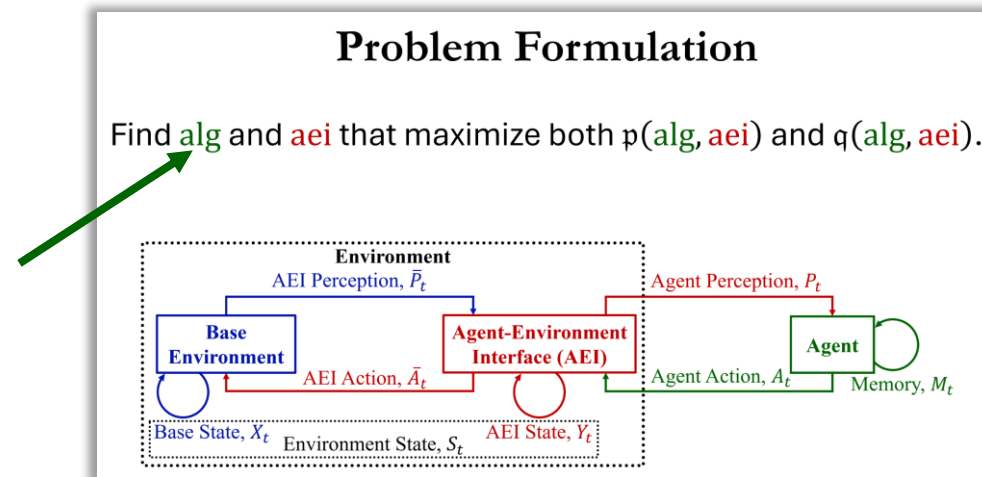
Reward Bonuses

- **Idea:** Increase the rewards given to the agent.



Reward Bonuses: Performance (p) Preserving?

- Inflating the rewards by c could **decrease** the performance objective.
 - **Solution:** Whenever **alg** references R_t , subtract c



Reward Bonuses: Performance (p) Preserving

aei: $R_t = \bar{R}_t + c$

alg: Change any reference to R_t to $(R_t - c)$

$$\delta_t = R_t + \gamma v_w(S_{t+1}) - v_w(S_t)$$

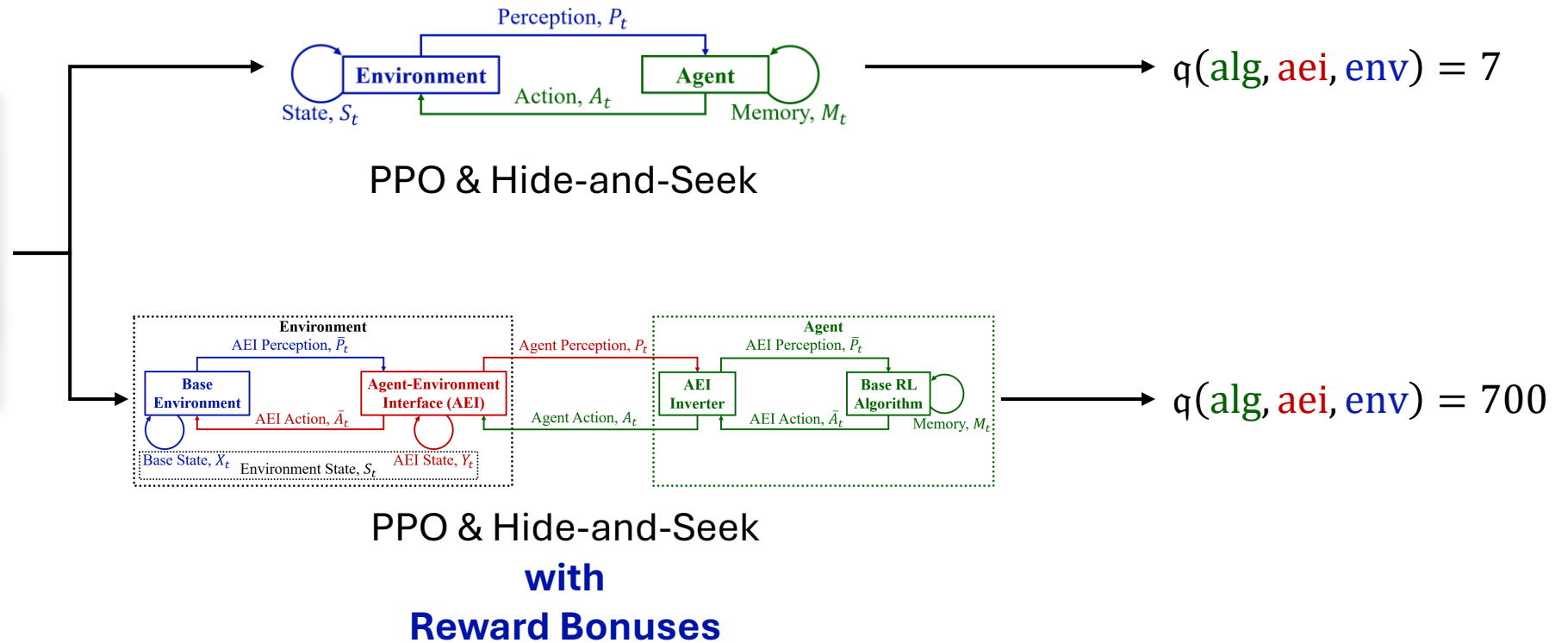
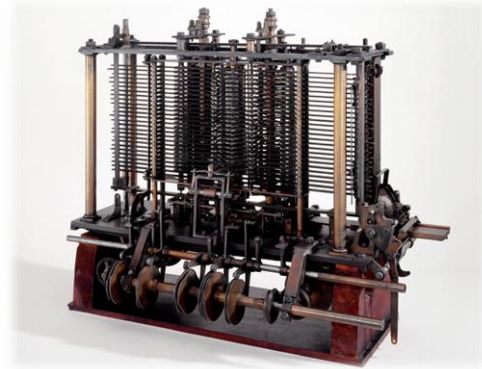
becomes

$$\delta_t = (R_t - c) + \gamma v_w(S_{t+1}) - v_w(S_t)$$

Note: Performance unchanged!

Question: Can we make concerns about this rigorous?

Different Qualia of the Same System



Representation and Qualia

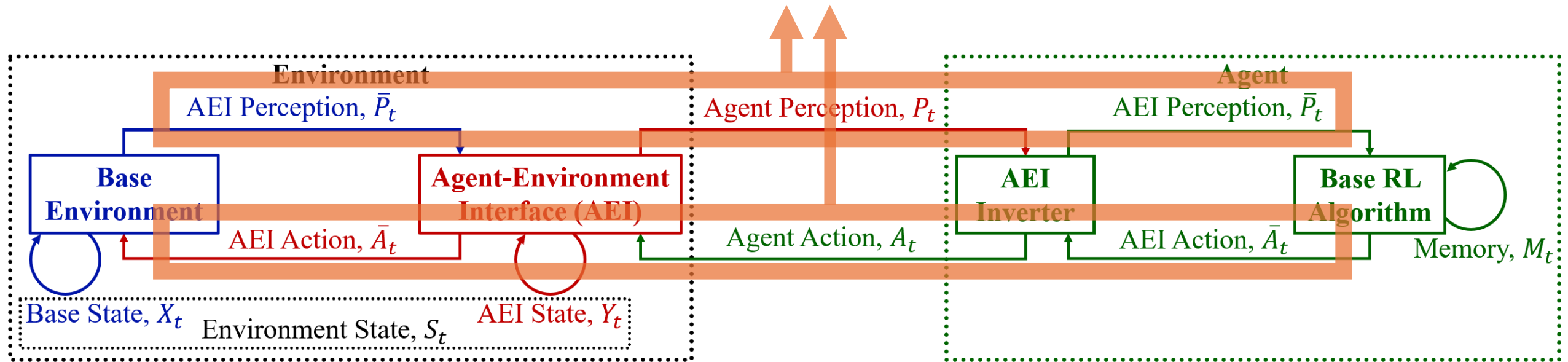
- Proponents of AI consciousness typically think that what matters is what is computed, not the physical representation
 - Carbon vs silicon
 - **MSB** vs **LSB**
- The representation we choose should **not** impact $q(\text{alg}, \text{aei})$ —what the system feels—as long as the system computes the same thing
 - Not the case in the reward-qualia setting!
- Reward bonuses can be implemented with a change of representation!
- **Idea:** We desire a q that is invariant to these representational choices.

Unsigned, **MSB**: $R_t = 164$
Signed, **MSB**: $R_t = -36$
Unsigned, **LSB**: $R_t = 37$



Invertible Transformations are Changes of Representation

Nothing being computed. Invertible transformations are changes of representation.



Isomorphic Random Variables

- Let:
 - X be a random variable
 - $Y = f(X)$
 - f be an invertible function
- Then:
 - f is called an *isomorphism*
 - X and Y are called *isomorphic*

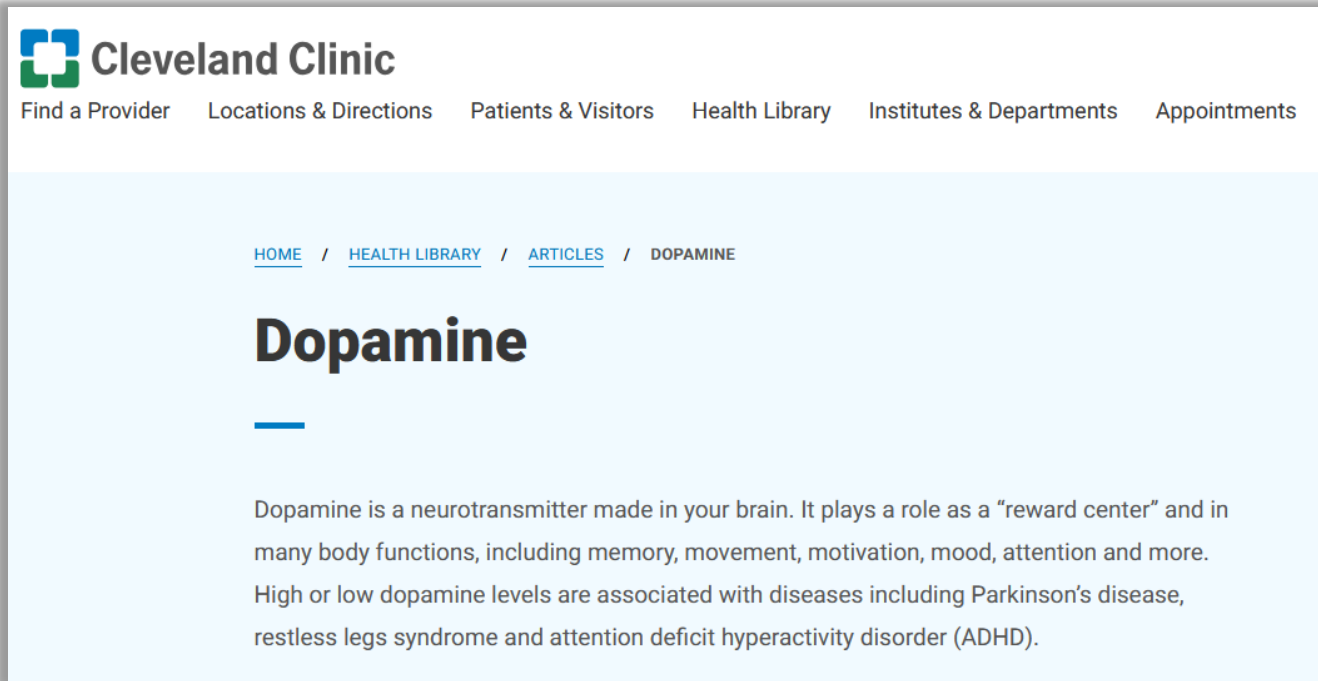
Representation Robustness

- Consider two mathematical models
 - First is **alg**, **aei**, **env**
 - Has perceptions, memories, actions: P_t , M_t , and A_t
 - Second is **alg'**, **aei'**, **env'**
 - Has perceptions, memories, actions: P'_t , M'_t , and A'_t
- If $(P_t, M_t, A_t)_{t=0}^{\infty}$ and $(P'_t, M'_t, A'_t)_{t=0}^{\infty}$ are isomorphic, these can model the same **computational** system
 - A **representation robust** qualia objective q satisfies:
$$q(\mathbf{alg}, \mathbf{aei}, \mathbf{env}) = q(\mathbf{alg}', \mathbf{aei}', \mathbf{env}')$$
 - Equivalently, q is invariant to invertible transformations of P_t , M_t , and A_t .

Reward-Qualia Setting

- Natural qualia objective functions are **not** representation robust.
 - Rewards can be inflated arbitrarily without changing the system

A Different Qualia Hypothesis



The screenshot shows the Cleveland Clinic website's health library page for Dopamine. The header includes the Cleveland Clinic logo and navigation links: Find a Provider, Locations & Directions, Patients & Visitors, Health Library, Institutes & Departments, and Appointments. The breadcrumb trail is HOME / HEALTH LIBRARY / ARTICLES / DOPAMINE. The main heading is "Dopamine" with a blue underline. The text below explains that dopamine is a neurotransmitter made in the brain, acting as a "reward center" and involved in memory, movement, motivation, mood, and attention. It also notes that high or low dopamine levels are associated with conditions like Parkinson's disease, restless legs syndrome, and ADHD.

Cleveland Clinic
Find a Provider Locations & Directions Patients & Visitors Health Library Institutes & Departments Appointments

[HOME](#) / [HEALTH LIBRARY](#) / [ARTICLES](#) / [DOPAMINE](#)

Dopamine

Dopamine is a neurotransmitter made in your brain. It plays a role as a “reward center” and in many body functions, including memory, movement, motivation, mood, attention and more. High or low dopamine levels are associated with diseases including Parkinson’s disease, restless legs syndrome and attention deficit hyperactivity disorder (ADHD).

How does dopamine make someone feel happy?

Dopamine is known as the “feel-good” hormone. It gives you a sense of pleasure. It also gives you the motivation to do something when you’re feeling pleasure.

RPE Hypothesis for Dopamine

A Neural Substrate of Prediction and Reward

WOLFRAM SCHULTZ, PETER DAYAN, AND P. READ MONTAGUE [Authors Info & Affiliations](#)

SCIENCE • 14 Mar 1997 • Vol 275, Issue 5306 • pp. 1593-1599 • DOI: 10.1126/science.275.5306.1593

- In primates, **dopamine** has the functional role of signaling an RPE
- In many RL systems, **TD error** has the functional role of signaling an RPE

Assumption 5

(RPE-Qualia Setting)

The quality of an RL agent's qualia can be measured in terms of the cumulative reward prediction error.
↓
(temporal difference error)

Dopamine = Wanting ≠ Liking

W. J. Freed. *Biology of Motivation, Dopamine, and Brain Circuits That Mediate Pleasure*, pages 105–109. Spring International Publishing, 2022.

K. C. Berridge and T. E. Robinson. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28(3):309–369, 1998.

K. C. Berridge. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, 191(3):391–431, 2007.

K. C. Berridge and M. L. Kringelback. Pleasure systems in the brain. *Neuron*, 86(3):646–664, 2015.



Peter Dayan's
RLC 2025
Keynote

TD Error Bonuses

- **Idea:** Add a constant to the TD error. Remove the constant whenever the TD error is used.

$$\delta_t = R_t + \gamma v_w(S_{t+1}) - v_w(S_t) + c$$

$$w \leftarrow w + \alpha(\delta_t - c) \frac{\partial v_w(S_t)}{\partial w}$$

$$\theta \leftarrow \theta + \alpha(\delta_t - c) \frac{\partial \ln(\pi(S_t, A_t, \theta))}{\partial \theta}$$

- $\mathbf{E}[\sum_t \delta_t]$ is representation exploitable.

Objection

- The TD error is $R_t + \gamma v_w(S_{t+1}) - v_w(S_t)$ not δ_t .
- $\mathbf{E}[\sum_t R_t + \gamma v_w(S_{t+1}) - v_w(S_t)]$ is also representation exploitable.

Representation Robustness is Impossible?

- Consider all q that:
 - Compute some real-valued statistic:
 - $g(P_0, M_0, A_0, P_1, M_1, A_1, \dots)$
 - $\sum_{t=0}^{\infty} \gamma^t R_t$
 - $\sum_{t=0}^{\infty} \gamma^t \delta_t$
 - $\sum_{t=0}^{\infty} \gamma^t v^\pi(S_t)$
 - Return some functional of this value
 - $\mathbf{E}[\sum_{t=0}^{\infty} \gamma^t R_t]$
 - $\text{Var}(\sum_{t=0}^{\infty} \gamma^t \delta_t)$
 - $\text{Median}(\sum_{t=0}^{\infty} \gamma^t v^\pi(S_t))$
- All such q are representation exploitable!
 - ... unless g is already invariant to invertible transformations

Representation Robustness is Possible!

- q that are functionals of **multiple** real-valued statistics can be invariant to invertible transformations!
 - Relative Entropy / KL Divergence
 - Only diffeomorphisms (differentiable and invertible transformations)
 - Mutual Information

- Example representation robust q :

- $q(\text{alg}, \text{aei}) = \sum_{t=0}^{\infty} I(P_t; A_t)$

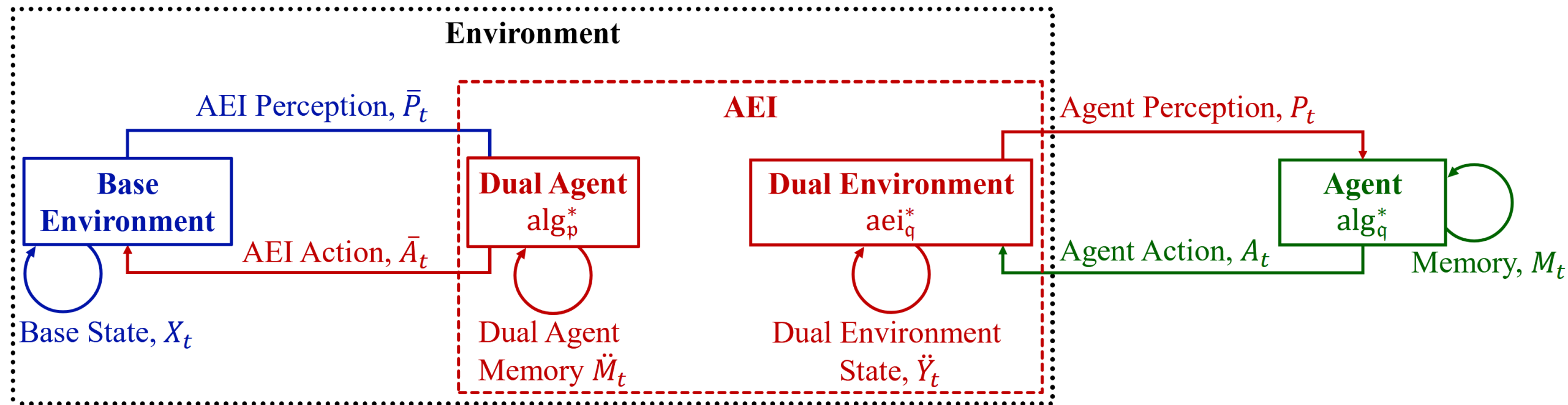
- $q(\text{alg}, \text{aei}) = I(S_t \rightarrow M_t)$

Plasticity as the Mirror of Empowerment

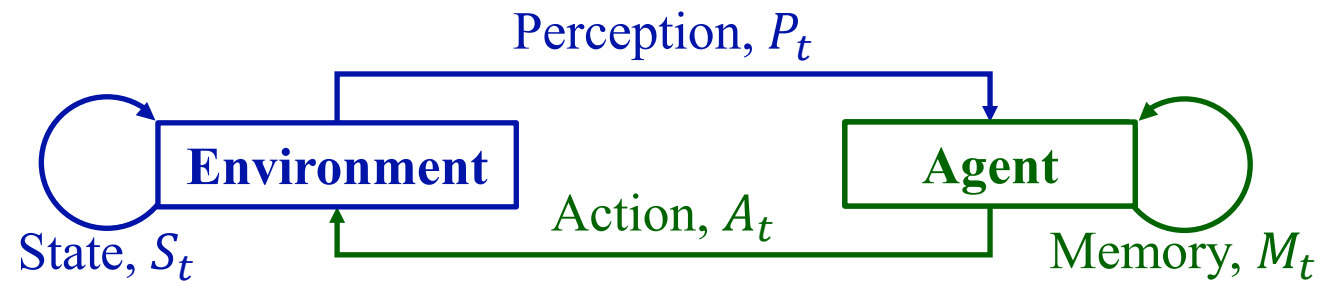
David Abel Google DeepMind	Michael Bowling Amii, University of Alberta	André Barreto Google DeepMind	Will Dabney Google DeepMind
Shi Dong Google DeepMind	Steven Hansen Google DeepMind	Anna Harutyunyan Google DeepMind	Khimya Khetarpal Google DeepMind
Clare Lyle Google DeepMind	Razvan Pascanu Google DeepMind, Mila	Georgios Piliouras Google DeepMind	Doina Precup Google DeepMind
Jonathan Richens Google DeepMind	Mark Rowland Google DeepMind	Tom Schaul Google DeepMind	Satinder Singh Google DeepMind

Reconsidering the AEI

Dual agent-environment strategy is always the answer.



Agent-Environment Process (AEP)



$$q(\text{alg}, \text{aei}) \rightarrow q(\text{alg})$$

Assumption 6

(Reinforcement-Qualia Setting)

The quality of an RL agent's qualia can be measured in terms of the **reinforcement of behavior**.

Reinforcement-Qualia Setting

(For policy search algorithms)

$$q(\text{alg}) = \mathbf{E} \left[\sum_{i=0}^{i_{\max}-1} \sum_{t=\text{start}(i)}^{\text{end}(i)-1} \frac{\pi(S_t, A_t, \Theta_{t+1})}{\pi(S_t, A_t, \Theta_t)} \right]$$

Representation Robustness is Impossible?

- Consider all q that:
 - Compute some real-valued statistic:
 - $g(P_0, M_0, A_0, P_1, M_1, A_1, \dots)$
 - $\sum_{t=0}^{\infty} \gamma^t R_t$
 - $\sum_{t=0}^{\infty} \gamma^t \delta_t$
 - $\sum_{t=0}^{\infty} \gamma^t v^\pi(S_t)$
 - Return some functional of this value
 - $\mathbf{E}[\sum_{t=0}^{\infty} \gamma^t R_t]$
 - $\text{Var}(\sum_{t=0}^{\infty} \gamma^t \delta_t)$
 - $\text{Median}(\sum_{t=0}^{\infty} \gamma^t v^\pi(S_t))$
- All such q are representation exploitable!
 - ... unless g is already invariant to invertible transformations

- Likelihood ratios are invariant to invertible transformations!
- This qualia objective function is representation robust.
- ... but there are challenges.

Reinforcement Bias & PPO

Derek Lacy Honors Thesis, Fall 2024

- Proposes methods for adding a reinforcement bias to PPO and SAC
- Empirical evaluation of resulting reinforcement-performance tradeoffs on MuJoCo and ALE environments



Conclusion

- Qualia optimization can be formalized
- An agent-environment interface (AEI) is natural to include, but admits degenerate solutions
- Many natural qualia objective functions are representation exploitable
 - E.g., expected sum of rewards, expected sum of TD errors
- Reinforcement-qualia settings result in naturally representation robust qualia objectives
 - Not clear how “reinforcement” should be defined in general
 - Reinforcement bias is sometimes achievable with little performance degradation

Future Work

- Can we find a **representation-robust** q that aligns with human subjective experience?
 - **Starting point:** Is there a general definition of *reinforcement*?
- Remove the (arbitrary) agent boundary
 - E.g., Evaluate qualia of an entire system

End / Questions