Modeling intrinsic motivation

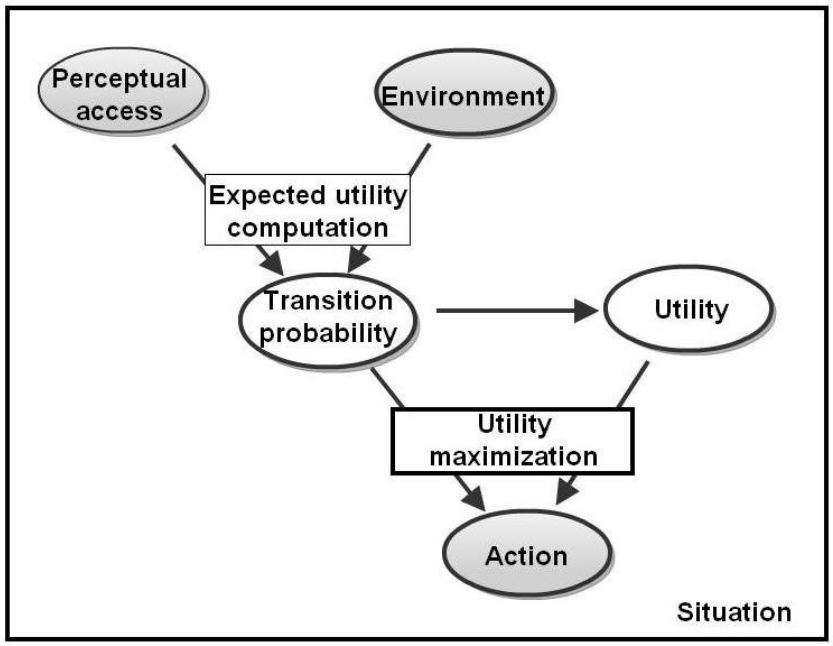
CS786

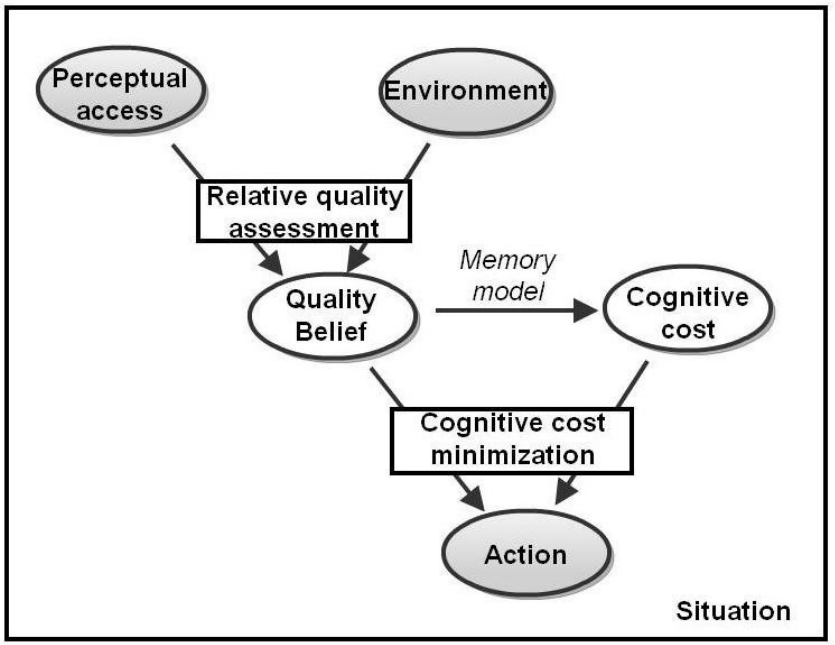
10th September 2024

Introduction

- Real agents don't respond reliably to multiple presentations of identical environmental stimulus
    - Important aspects of agents' motivations are intrinsic
- Intrinsic motivation currently modeled explicitly as a drive
    - Novelty, surprise (Barto, Singh)
    - Need to learn, curiosity (Oudeyer, Schmidhuber)
- Intrinsic motivation must emerge naturally from a realistic choice model
- Once such a choice model is specified, the nature of intrinsic motivation will become clear
- We have created a realistic choice model
- This paper shows that our choice model leads to an endogenous and positive theory of intrinsic motivation

A new definition of rationality





Old school: A rational agent acts in ways that maximize its perceived outcome utility

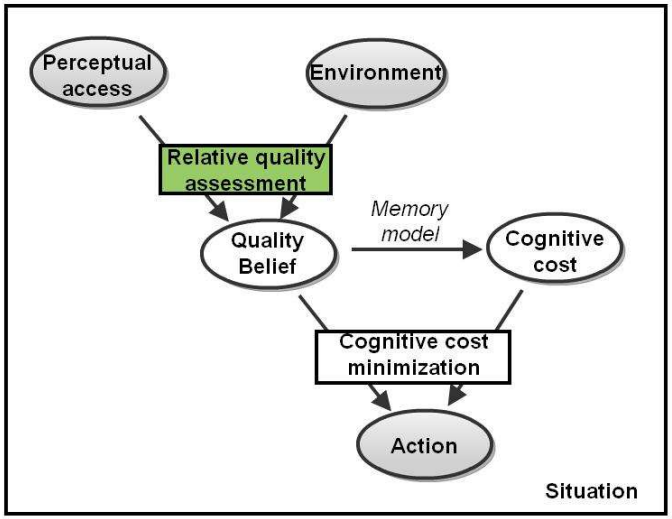
Our idea: A rational agent acts in ways that minimize its cognitive effort in deciding how to act

Relative Quality Assessment

- Environment changes
- Agent's theory about the environment becomes obsolete
- Agent's predictions about the environment result in poor results
- Agent experiences regret
- We define a measure of regret after arriving at a new quality belief x_a with respect to an old one x_b,

$$R(x_a, xb) = \sum_{j=1}^{|S|} x^j{}_a \log \frac{x^j{}_a}{x^j{}_b}$$

- Using a KL divergence gives us a cardinal quantity in the form of a difference between distributions reflecting different relative quality-beliefs
- Captures the intuition that all valuations are made from an agent's current belief's vantage point, i.e., preference valuations are affine and relative



Memory Model

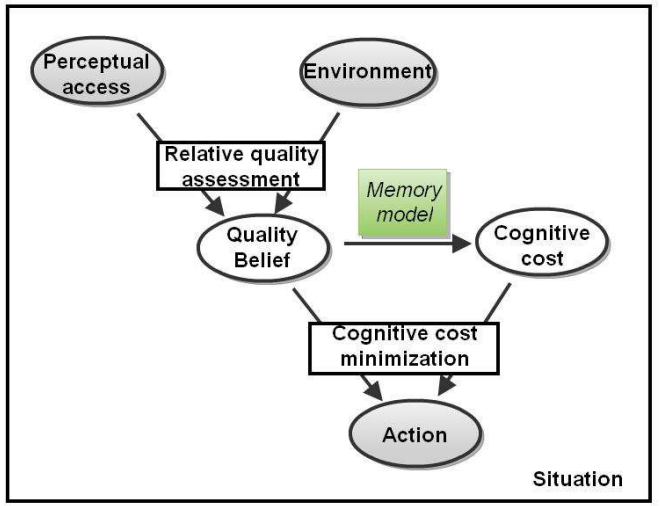
- Exceptionality of a past quality belief is measured as the degree to which it is surprising with respect to the current quality belief,

$$A(\mathrm{x}_{old}) = |\mathrm{R}(\mathrm{x}, \mathrm{xold}) - \bar{R}|,$$

capturing the intuition that both highly surprising and highly unsurprising events are exceptional.

- The more exceptional a past quality belief, the more available and hence, less costly, it will be to recall. Assuming a nominal recall cost of unity, the total cognitive cost *T* of populating an active memory *M'* from all past experiences is,

$$T = \sum_{x_i \in M\prime} A^{-1}(xi)$$



Memory Model

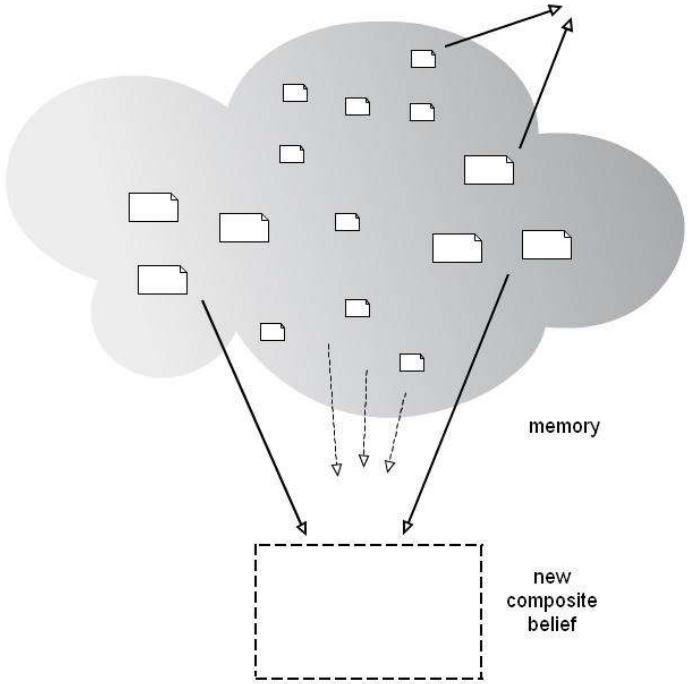
- Some subset of old beliefs, recalled from memory, additively combine to create a new prospective belief
- Which subset is chosen depends on how well the agent is able to predict using the obtained belief
- Cognitive cost T trades off against the reward predicting utility of the quality belief the cognitive process of memory recall generates. We quantify this using a measure of prediction confidence,

$$C = \frac{1}{C_{max}} \frac{\log|x| - H(x)}{\sum_{x_{old} \in M,} R(x, xold)},$$

where H(x) is a measure of information entropy.

- Reward-inference r updates an agent's current belief x whenever observed via a convex sum weighted by confidence,

$$x' = Cx + (1 - C)r$$

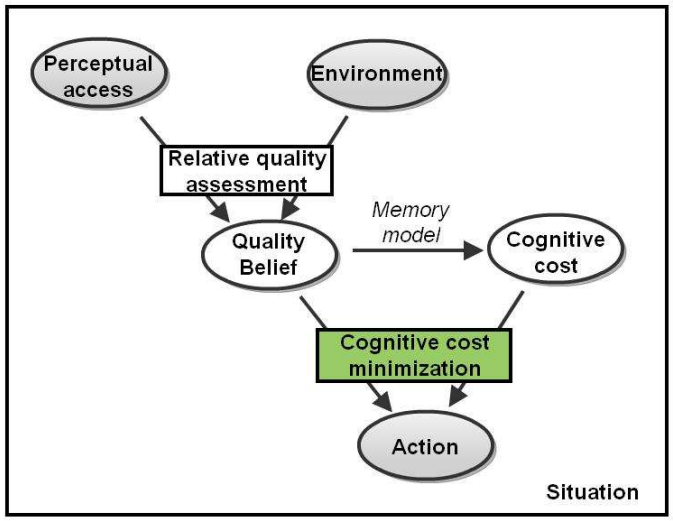


Cognitive Cost Minimization

- Our decision model yields a novel principle of rational action: need satisficing cognitive cost minimization, or,

$$arg\min_{x} T,$$
$$C_{new} \geq Col_d$$

- Intuitive objective function
  - An agent that remains confident about its predictive ability retains its existing belief about outcomes
  - Agent makes cognitive effort to recall and decide only if its confidence in its ability to predict the environment is reduced
- Computationally, we solve this as a combinatorial optimization problem of identifying which prior beliefs to recall into active memory.
- Once this subset is identified, the agent's new experience expectation is obtained by averaging over these recalled beliefs.



Key differences

Traditional rational models

- Assume outcome valuation can be well-modeled as absolute reward
- Optimizing expected reward/cost is the optimal decision strategy
- Expected reward is computed by combining rewards from all past experiences with a multiplicative forgetting function
- Intrinsic motivation terms have to be added separately to this framework
- Existing models of intrinsic motivation are thus *normative*

Self-motivated learning

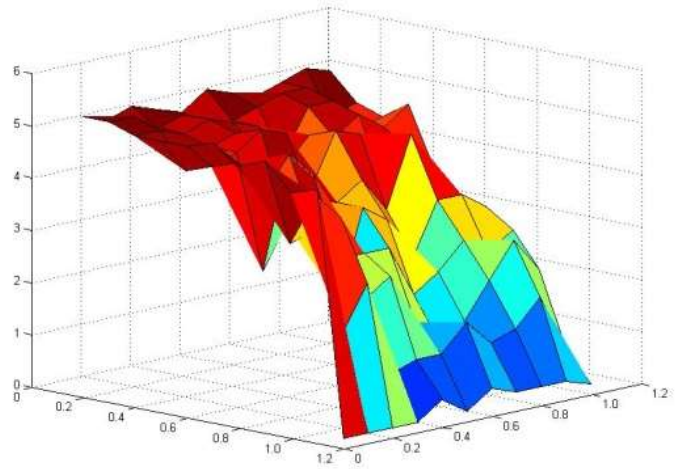
- Assumes outcome valuation creates preferences relative to all other outcomes
- Can't optimize relative quantities, have to look elsewhere!
- We assume that preferences are updated using a particular memory model that gives us a non-trivial forgetting function
- Operating this memory model requires cognitive effort, which we quantify.
- Minimizing cognitive cost gives a tractable choice theory + a *positive* theory of intrinsic motivation

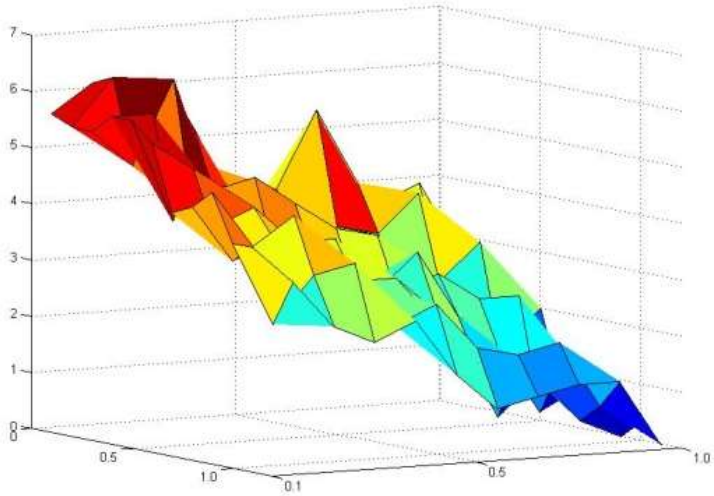
Overview of results

- Three experiments
  - Experiment 1: comparison of self-motivated learning with Gittins index solution to multi-arm bandit problem
    - Shows self-motivated learning to be a reasonable learning algorithm
  - Experiment 2: comparison of self-motivated learning against Robust IAC on an abstract signal tracking task
    - Shows that curiosity emerges from our model in ways congruent with IAC predictions
  - Experiment 3: demonstration of learning cooperative strategies in prisoner's dilemma settings
    - Shows additional dimensionality (altruism/cooperation) of intrinsic motivation (other than curiosity) emerging from our model

Learned altruism in iterated prisoner's dilemma

- PD represents a good test of social realism for rational actors
- Iterated PD games allow strategies to be learned depending on opponent's behavior.
- Mimics real-world interactions at an abstract level.
- Bottom two axes in results plots show a grid of opponents following mixed strategies [p1 p2], where
    - p1 is probability that opponent defects given the player cooperated on the previous trial
    - p2 is the probability that the opponent defects given the player defected on the previous trial
- Tit-for-tat is a robust strategy, also seen in the natural world.
- Traditional rational agents fail to learn realistic strategies
- Our agent learns an approximately *mean* tit-for-tat strategy , i.e., reciprocates, unless opponent never defects, in which case it exploits.
- Potential explanation of emergence of indirect altruistic behavior - strategies that requires minimal changes in behavior is promoted via cognitive cost criterion
- <u>Take home message:</u> individual utility, including intrinsic motivations, is sufficient to account for cooperative behavior! No social utility needed.





Discussion

- Making alternative assumptions about the nature of valuation forces us to turn inward to find a coherent criterion for rationality
- Cognitive effort in recalling prior experiences to construct beliefs is found to be a quantity of interest
- Minimizing cognitive effort gives us a learning algorithm that behaves as if it is intrinsically motivated
- Captures key aspects of curiosity identified by Oudeyer in IAC
    - Testable differentiation between the two theories: exposure to intermediate complexity after exposure to high complexity
        - IAC adapts to the new environment quickly
        - Our model fails to learn for an indefinitely long period of time
- Congruent with entropy reduction ideas promoted by Schmidhuber as fundamental to intrinsic motivation
    - Differentiating factor is cognitive effort
        - Schmidhuber's agents will attempt to reduce uncertainty at all costs
- Congruent with expectation violation premises used by Barto, Singh et al
    - Differential prediction is dyadic nature of intrinsic motivation
        - Barto et al predict motivation as an increasing function of prediction error
        - Most realistic theories of curiosity and intrinsic motivation predict motivation will drop for high prediction errors