Dynamic self-efficacy updating leads to optimistic overgeneralization

Jing Li (jing.li@icahn.mssm.edu)
Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai
New York, NY 10027, USA

Angela Radulescu (angela.radulescu@mssm.edu)
Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai
New York, NY 10027, USA

Abstract

Humans often need to make predictions about future reward by generalizing from similar past experiences. Positive overgeneralization occurs when a rewarding experience is attributed to multiple states of the environment. In this work, we show that a model-free RL agent that dynamically updates its self-efficacy beliefs as it approaches a goal learns optimistic overgeneralized value representations. We suggest that dynamic self-efficacy beliefs may underlie the human tendency to overgeneralize from positive outcomes and test the predictions of our model in a novel behavioral paradigm designed to measure positive overgeneralization across a 1D perceptual state space. We find preliminary evidence that participants with higher self-reported general self-efficacy overgeneralize from rewarding outcomes, and place greater valuation on rewarded trials than those with lower self-efficacy beliefs, as reflected in faster reaction times.

Keywords: self-efficacy; reinforcement learning; positive overgeneralization; mania; computational psychiatry

Introduction

Self-efficacy, defined as one's belief in the capacity to execute actions that achieve desired outcomes, is an adaptive trait that is continuously shaped through performance accomplishments, vicarious experience, social persuasion and physiological signals (Bandura, 1997). Positive overgeneralization (POG) refers to overgeneralizing rewarding outcomes from one aspect of life to others (Stange et al., 2012). For individuals at risk for mania, an initial success often induces a positive appraisal of self, which in turn leads to higher expectations of future successes and can manifest in POG tendencies (Johnson, 2005). However, this link between self-efficacy beliefs and POG has not been formally tested.

Here, we propose *dynamic self-efficacy* as a cognitive mechanism for generalizing rewarding outcomes from one experience to another, and ask whether high self-efficacy update rates lead to overgeneralization of rewarding outcomes. We introduce a learning rule that dynamically updates self-efficacy in appraisal of moving closer to a goal. In our model, self-efficacy — defined as the belief that achieving a goal now is more likely to lead to achieving goals in the future — is a dynamic attribute, continuously shaped by action outcomes.

Methods

Model: Q-Learning with dynamic self-efficacy We implemented a model-free Q-learning agent that learns in sequential grid-world environments (Zorowitz, Momennejad, & Daw, 2020). This set-up enabled us to study how changes in self-efficacy affect reward backpropagation across states as the agent interacts with the environment. Crucially, we aim to provide a mechanism through which the agent can update its self-efficacy based on the outcomes of its actions. In our model, reward prediction errors (RPEs) — the discrepancy between expected and actual action outcomes — serve as direct feedback for the appraisal of performance accomplishment. This aligns with the insight that enactive mastery of experiences is a critical source of efficacy information because it is the most direct way for the agent to learn self-efficacy beliefs based on its successes and failures (Bandura, 1997).

After the agent takes an action using a softmax policy (inverse temperature = 1), the reward prediction error (RPE) is computed as:

$$\delta_t = R_{t+1} + \gamma \cdot w_t \cdot \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \quad (1)$$

The *self-efficacy belief*, w_t, scales the highest possible future expected reward contingent on action, reflecting the agent's belief that it can successfully select the best action in the immediate future. Reward prediction errors serve as the critical signal for updating both the action values and the self-efficacy parameter:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta_t \quad (2)$$

$$w_{t+1} = w_t + w_{LR+} \cdot \delta_t \quad (3)$$

The *reward learning rate* α indexes how quickly value estimates should be updated based on error. And the *self-efficacy learning rate*, w_{LR+}, quantifies how sensitive the agent's beliefs about its own self-efficacy are to positive RPEs, which in a sequential setting can be interpreted as a signal of successfully approaching a goal.

Simulation environment. In the grid-world simulation environment (Fig. 1A and B), the agent starts at the top left location (0,0) and must reach a rewarding terminal state at the bottom right position (9,9). In each state, the agent can either move up, down, left or right. The reward is +1 at the terminal state and 0 everywhere else in the grid-world. A penalty of -0.5 was imposed to discourage the agent from straying off the grid. The agent's training consisted of 200 episodes, each with up to 200 steps, unless the terminal state was reached sooner, concluding the episode.

The "Clock Task". We developed a behavioral task in which human participants were trained to associate different clock hand orientations with binary reward outcomes (Fig. 1C). During the training phase, participants were shown stimuli featuring two directions of clock hands that were 90 degrees apart. One direction was associated with a positive monetary reward, while the other one was not. During the testing phase, participants were presented with new stimuli featuring directions of clock hands that were intermediate to the two shown in the training phase. Participants were asked to predict whether each test stimulus will lead to a reward based on how similar they are to the two previously shown, and only received pseudofeedback. A total of 24 online participants from Prolific completed a pilot study, and data from 17 participants were included in the analyses after they met the inclusion criteria for learning the task.

Participants also completed the **Positive Overgeneralization (POG)** scale, an 18-item self-report questionnaire where participants are asked to respond to items by indicating if they agree or disagree with the statements on the tendency to

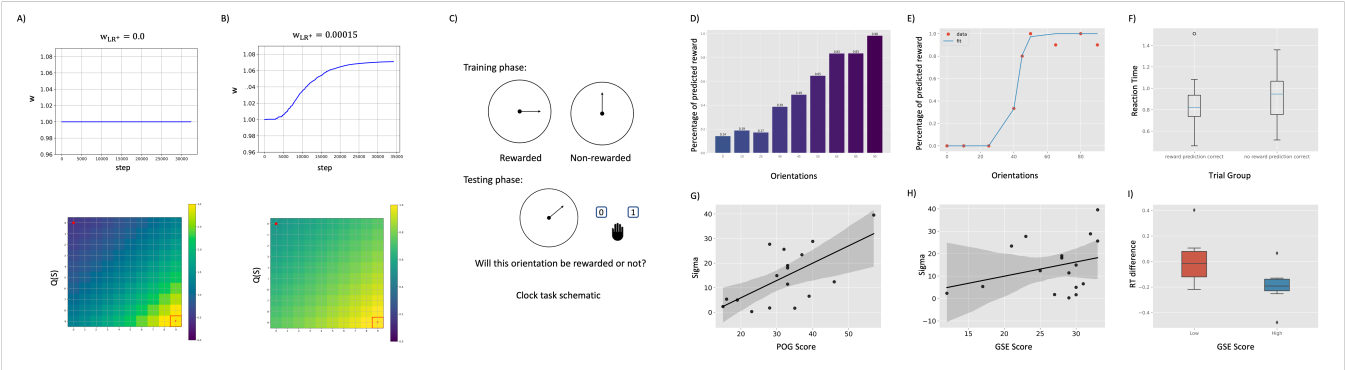

Figure 1: A) Agent with static self-efficacy ($w_{LR+} = 0.0$). B) The agent with dynamic self-efficacy ($w_{LR+} = 0.00015$) learns to place greater values for states closer to the start position. C) Schematic of the "Clock Task". D) Percentage of correct predicted reward trials for each clock hand orientation on the group level. E) Fitted psychometric curve for one participant. F) Reaction time difference for correct rewarding trials versus non-rewarding trials at the group level. G) Correlation between POG score and fitted sigma parameter. I) Comparison of RT differences between correct rewarding and non-rewarding trials for low vs. high self-efficacy individuals.

overgeneralize from a given successful experience (Stange et al., 2012); and the **The General Self-Efficacy (GSE)** scale, a 10-item self-report questionnaire that reflects participants' optimistic self-beliefs to perform difficult tasks (Schwarzer & Jerusalem, 1995).

Results

Simulation results. We first tested how different update rates for self-efficacy impact the underlying value representation. We repeated the simulation for two different agents with the self-efficacy learning rate w_{LR+} fixed at 0.0 and 0.00015 respectively. As expected, the two agents acquired different levels of self-efficacy at the end of training, with higher learning rates leading to higher self-efficacy levels (Fig. 1A and B). We also found that for agents that update self-efficacy more quickly, the reward at the terminal states backpropagates to states closer to the start state, and the overall value of all states is higher. In other words, agents with more sensitive self-efficacy beliefs develop *optimistic overgeneralized future reward expectations*.

Behavioral results. To test our model predictions that higher self-efficacy levels lead to overgeneralized reward expectations, we collected behavioral data from N=17 participants and analyzed their reward prediction responses and reaction times (RT) on the Clock Task. Participants effectively learned to predict rewards based on the directions of the clock hands in the test phase (Fig. 1D). For each participant, we fit a sigmoid psychometric curve based on their reward prediction responses (Fig. 1E) and extracted the sigma parameter, which controls the steepness of the curve, and provides a measure of positive overgeneralization. A paired-sample t-test revealed a trend where participants responded more quickly to stimuli that would lead to a reward (Fig. 1F, $t(16) = -1.68, p = .11$), consistent with higher valuation of rewards.

Interestingly, at the individual level, the POG score was significantly correlated with the sigma parameter (Fig. 1G, $r(15) = .64, p < .01$). Individuals who reported overgeneralization from positive experience also exhibited a more gradual transition in predicting rewarding stimuli, validating the Clock Task as a computational assay of the tendency to overgeneralize from positive outcomes. Furthermore, the pilot study results indicated a trending positive correlation between self-efficacy levels, as measured by the GSE scale, and the sigma parameter (Fig. 1H, $r(15) = .31, p = .22$). Although this trend is not significant in the pilot group, it suggests that individuals with higher self-efficacy levels may also tend to overgeneralize more, in line with our model prediction that higher self-efficacy levels will lead to overgeneralized reward expectations.

Finally, when we split participants by self-reported self-efficacy, an independent t-test showed that the high self-efficacy group exhibited a significantly larger difference in RT between rewarding trials and non-rewarding trials (Fig. 1I, $t(16) = 2.23, p < .05$). This result suggests that individuals with higher self-efficacy place greater valuation on rewarding trials compared to those with low self-efficacy, and aligns with our model prediction that the overall value of rewarding states is higher when self-efficacy levels are elevated.

Conclusion and future directions

This study aimed to formalize the hypothesis that a higher sensitivity of self-efficacy beliefs to goal-directed feedback could provide a mechanism for positive overgeneralization. We proposed a model that augments model-free RL agents with dynamic self-efficacy beliefs that are updated based on reward prediction errors (RPEs), which provide a direct signal of goal attainment. We showed that in sequential learning settings, this simple mechanism is enough to give rise to optimistic overgeneralization of reward expectations. We also developed a behavioral task paradigm and showed in a pilot sample that individuals with higher self-efficacy levels exhibited reward overgeneralization and placed greater value on rewarding trials than those with lower self-efficacy. Further research with a larger cohort will validate the preliminary findings.

References

Bandura, A. (1997). *Self efficacy: the exercise of control*. New York (N.Y.): W. H. Freeman.

Johnson, S. (2005). Mania and dysregulation in goal pursuit: a review. *Clinical Psychology Review*, *25*(2), 241–262.

Schwarzer, R., & Jerusalem, M. (1995). *General Self-Efficacy Scale*.

Stange, J. P., Molz, A. R., Black, C. L., Shapero, B. G., Bacelli, J. M., Abramson, L. Y., & Alloy, L. B. (2012). Positive overgeneralization and Behavioral Approach System (BAS) sensitivity interact to predict prospective increases in hypomanic symptoms: A behavioral high-risk design. *Behaviour Research and Therapy*, *50*(4), 231–239.

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, Avoidance, and Sequential Evaluation. *Computational Psychiatry*, *4*(0), 1.