Title
Name that state: How language affects human reinforcement learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Radulescu, Angela
Vong, Wai Keen
Gureckis, Todd M

Publication Date
2022

Peer reviewed

Name that state: How language affects human reinforcement learning

Angela Radulescu (angela.radulescu@nyu.edu)
Wai Keen Vong (waikeen.vong@nyu.edu)
Todd M. Gureckis (todd.gureckis@nyu.edu)
Center for Data Science and Department of Psychology
New York University

Abstract

We describe two experiments designed to test whether the ease with which people can label features of the environment influences human reinforcement learning. The first experiment presents evidence that people are more efficient at learning to discern relevant features of a task when candidate features are easier to name. The second experiment shows that learning what action to take in a given state is easier when states have more readily nameable verbal labels, an effect that was especially pronounced in environments with more states. The interaction between CLIP, a state-of-the-art AI model trained to map images to natural language concepts, and established human RL algorithms, captures the key effects without the need to specify condition-specific parameters. These results suggest a possible role for language information in how humans represent the environment when learning from trial and error.

Keywords: reinforcement learning, state, task representation, language

Introduction

Adaptive behavior depends on our ability to change our actions based on context. In reinforcement learning (RL), this context dependency is captured by the notion of *state* – features of the current context that are relevant for the task at hand. For many real world tasks, the relevant features are rarely known a priori; instead, we learn them from experience with our environment. For instance, a traveler navigating in a new city might learn that they can decode their location along the North-South axis from their position relative to certain landmarks.

How humans learn to represent states is a topic of active research in cognitive science (Gureckis & Love, 2009; Niv, 2019). Computational accounts have emphasized the importance of abstraction, hierarchy and compression (Ho, Abel, Griffiths, & Littman, 2019; Eckstein & Collins, 2020; Lai & Gershman, 2021). And mechanistic approaches have sought to formalize selective attention as a dynamic process by which agents learn relevant state features from experience (Kruschke, 1992; Le Pelley, Haselgrove, & Esber, 2012; Leong, Radulescu, Daniel, DeWoskin, & Niv, 2017; Braunlich & Love, 2021).

One cognitive function that exhibits many of the properties thought to be important for state representation is language (Tam et al., 2022). Language representations are inherently low-dimensional and flexible at different levels of abstraction (Piantadosi, Tenenbaum, & Goodman, 2016; Antonello, Turek, Vo, & Huth, 2021). Several recent studies in artificial reinforcement learning have built on this insight to demonstrate the benefits of augmenting RL agents with linguistic information (Yao, Rao, Hausknecht, & Narasimhan, 2020; Tuyls, Yao, Kakade, & Narasimhan, 2022).

Yet mechanistic models rooted in the neurobiology of RL do not yet provide a clear account of how language might interface with trial and error learning. In some of these accounts, language is considered a more complex process downstream from simple reinforcement learning and is thus ignored. In this study, we begin to explore the role of language in modulating processes previously thought to reflect simple reinforcement learning. We start from the hypothesis that language may provide a representational basis over which reinforcement learning and action selection can occur. Rather than being downstream, language-mediated processes may provide input to reinforcement learning mechanisms.

We focus on one specific aspect of language, the *nameability* of environmental features. People more easily learn to categorize concepts when features of those concepts are easier to name in language (Zettersten & Lupyan, 2020; Lupyan & Zettersten, 2020), even controlling for lower level perceptual discriminability. Because category learning and reinforcement learning both require mapping multidimensional perceptual observations to states, we expected nameability to also influence trial and error learning of action policies. In particular, if language provides a set of candidate state features for reinforcement learning, people should learn more efficiently when features of the environment are more meaningful and nameable.

We study the effects of nameability in two task environments: one in which selective attention has been shown to constrain reinforcement learning to relevant features (Niv, 2019) (Experiment 1); and another in which learning has been shown to depend on working memory (Collins & Frank, 2012) (Experiment 2). We find that nameability facilitates learning which features of a task are relevant; and it also promotes more efficient learning of action policies, in particular for larger state spaces. Our findings suggest the intriguing possibility that language constrains how people learn to represent goal-directed tasks. We propose a modeling approach leveraging state-of-the art AI models in combination with reinforcement learning algorithms to help explain our results.

Experiment 1

In the first experiment, we tested the effects of language on representation learning. Specifically, we asked whether the extent to which features of the environment are easier to name enables more efficient learning to attend to relevant features of a task (Niv, 2019). An effect of nameability would suggest that selective attention during learning is influenced by language processes related to naming.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Method

Participants 98 participants were recruited via the online platform Prolific. They received as compensation a flat rate of $3 plus a bonus based on their performance on the task.

Operationalizing nameability To select stimuli for our experiments, we followed Zettersten and Lupyan (2020). In their study, the authors quantified nameability by measuring the extent to which a large number of human participants agreed on a common label for a feature. In particular, *Simpson's diversity index* quantifies agreement while accounting for (1) the number of possible labels; (2) the frequency with which people use each label. For both Experiment 1 and Experiment 2, we thus selected features that Zettersten and Lupyan (2020) found to differ in nameability on the basis of Simpson's diversity index.

To independently validate the stimuli, we conducted a post-experiment manipulation check. This check consisted of asking participants "what word would you use to describe" each feature (displayed as an individual image). This allowed us to directly compute Simpson's diversity from language data provided by participants in our experiments.

Procedure Participants learned from trial and error to make repeated choices between three stimuli referred to in the instructions as 'creatures' (Fig. 1A). Each creature was defined by one of three colors and one of three shapes. On each trial, the features within each dimension were reshuffled to form new stimuli (i.e. all six features were always present, but in different stimulus configurations). Participants had 5 seconds to select one of the three creatures. After the participant made their choice, they received a binary reward. Participants were instructed that at any given time, one of the colors or one of the shapes was designated as the "magic feature". Choosing the creature that possessed the magic feature was rewarded with 0.75 probability. Choosing any of the other 2 creatures was rewarded with only 0.25 probability. The magic feature thus defines a correct state representation for the task: participants could maximize reward by distinguishing stimuli only based on the presence or absence of the magic feature, while ignoring features for the other irrelevant dimensions.

We hypothesized that participants would be more efficient at learning the relevant feature when candidate features are easier to name. To test this, we manipulated nameability in a within-subjects design. Within one of the categories (color or shape), three of the features were more nameable, and three were less nameable, yielding two conditions – *high* and *low-nameability* (Fig. 1B). Note that colors were selected such that there were little to no differences in perceptual discriminability as measured in an independent norming study; and shapes were relatively matched on shape complexity (Zettersten & Lupyan, 2020). Each participant experienced 12 blocks in pseudo-random order, 6 in the high-nameability condition and 6 in the low-nameability condition. Each block ("round" to participants) was defined as a period of 18 trials during which the magic feature remained constant.

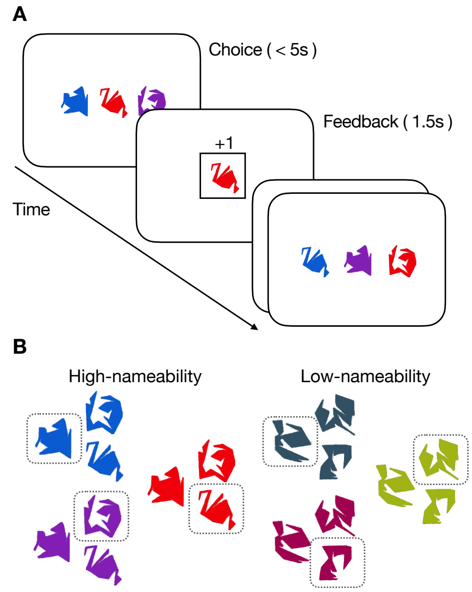

Figure 1: **Experiment 1 procedure.** **A**: Example trial sequence. **B**: The stimulus set in the *high-nameability* condition is shown on the left. In this condition, colors are more aligned with natural categories and shapes are more recognizable than in the *low-nameability* condition (right). Example stimuli that could appear on a single trial are circled.

Computational modeling

Work on learning task representations has proposed attention-weighted reinforcement learning as a candidate mechanism for learning what features are relevant for a task (Kruschke, 1992; Leong et al., 2017). To formalize the hypothesis that selective attention during reinforcement learning is influenced by language, we built a simple reinforcement learning agent that learns to update feature weights through trial and error. As in classic models of associative learning (Pearce & Mackintosh, 2010), we assumed that each feature is updated in proportion to its own learning rate (or "associability"). This learning rate was determined by how nameable each feature is. The full model is specified below.

Feature reinforcement learning To make a choice, the agent computes the value of each stimulus as a linear combination of features, equally weighted by dimensional attention (that is, learning is not biased towards color or shape):

$$V_t(S_i) = \sum_{d=1}^{2} \frac{1}{2} w_t(f_d). \tag{1}$$

Values are passed through a softmax action selection function:

$$p(c_t = S_i | V_t(S_i)) = \frac{e^{V_t(S_i)\beta}}{\sum_j e^{V_t(S_j)\beta}}. \tag{2}$$

Once a choice is made, the weights of the chosen features are updated in proportion to the prediction error, times a learning rate determined by each feature's nameability.

$$w_{t+1}(f_c) = w_t(f_c) + \eta(f)[R_t - V_t(S_c)]. \tag{3}$$

For all simulations, we presented the agent with the same sequence of stimuli that participants experienced in the actual experiment. Feature weights were initialized at 0. Choice softmax temperature β was fixed at 10.

A computational index of nameability Classic models of reinforcement learning typically treat features as uniformly learnable, and do not readily accommodate humans' priors for how states should be structured around language. As a first step in trying to address this limitation, we combined feature reinforcement learning with CLIP (Radford et al., 2021), a recent deep neural network architecture trained on 400M image-caption pairs sourced from the internet. CLIP is trained by trying to match a given image with its corresponding caption, and we hypothesized that the multimodal representational space learned by CLIP would be a good candidate for encoding language priors for two main reasons. First, CLIP contains knowledge about a very large number of multimodal concepts (Goh et al., 2021), and second, CLIP can flexibly encode an arbitrary set of natural language descriptions to use for zero-shot image classification (Radford et al., 2021).

To extract a measure of nameability using CLIP, we passed each (image) feature through the vision encoder of a pretrained CLIP RN50x16 network, obtaining a set of feature embeddings. Separately, we passed the most frequent human-generated natural language label for each feature (which we denote as $l \in L$), through the corresponding text encoder of CLIP to obtain a set of label embeddings. For each image feature, we used CLIP's zero-shot classification procedure to classify it amongst the set of modal labels, by computing the dot product between a given feature embedding and all possible label embeddings, followed by applying a softmax along the label dimension. The resulting probability distribution represents how likely each label matched a given feature. We reasoned that more nameable features should also be more separable in the representational space learned by CLIP; and that this separability will be indexed by a feature-specific classification entropy:

$$H(f) = -\Sigma_l p(l) log(p(l)). \tag{4}$$

.

To obtain an index of nameability that can be mapped onto $(0,1)$ bounded learning rates in a reinforcement learning setting, we passed classification entropy $H(f)$ through a reverse-sigmoid transform:

$$\eta(f) = \frac{1}{1+e^{H(f)}}. \tag{5}$$

Intuitively, this procedure captures a naming process by which humans label the features before any learning has taken place (i.e. how likely is this visual concept to be "blue", "purple" etc). Features that are easier to name should result in a lower classification entropy, while features that are harder to name should result in a higher classification entropy.

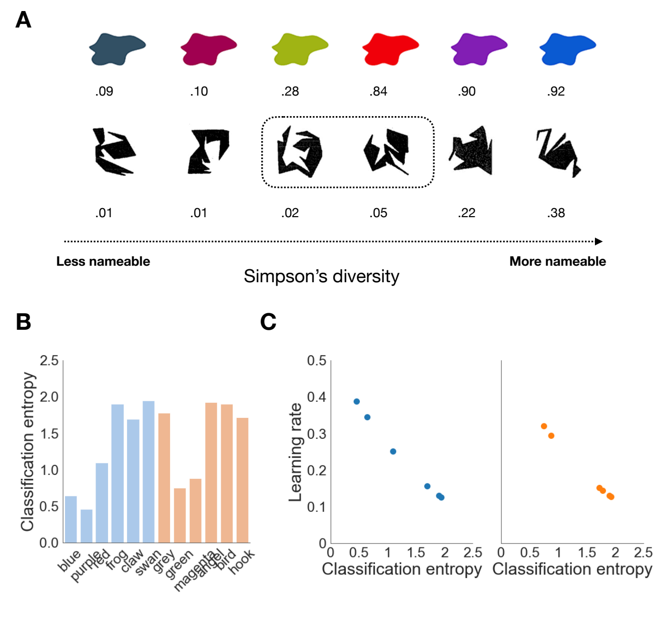

Figure 2: **Experiment 1 nameability index**. **A**: Manipulation check. Colors (top) and shapes (bottom) ordered by Simpson's diversity. For two of shapes (circled), condition labels were flipped based on participants' language data. **B**: Classification of visual features was performed over the labels that humans most frequently used to describe each image. Each feature (e.g. "red", "frog", etc.) was provided to CLIP as a single image. Images that were easier to classify had lower classification entropy. **C**: Learning rates in the high- and low-nameability conditions for Experiment 1, obtained by passing CLIP classification entropy through a reverse-sigmoid transform.

Results

Manipulation check To verify that participants judged features in the high-nameability condition to be more nameable, we computed Simpson's diversity based on the labels that participants used to describe each of the features. In general, participants were more likely to use consistent labels for stimuli in the high-nameability condition (Fig. 2A). We also found that our computational index of nameability derived from CLIP broadly tracks these differences, such that more nameable features and shapes had lower classification entropy ($M_{high} = 1.42$, $M_{low} = 1.92$, Fig. 2B). Computing feature-level learning rates by passing the classification entropy through a reverse-sigmoid resulted in lower learning rates on average for features in the low-nameability condition (Fig. 2C).

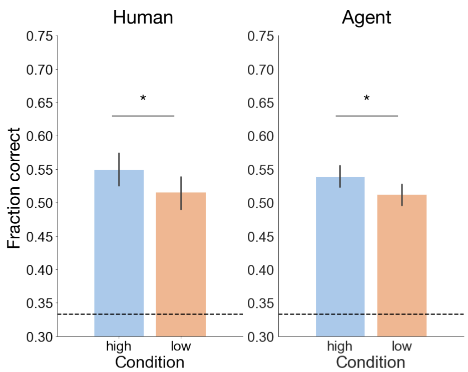

Figure 3: **Experiment 1 results**. Empirical (left) and simulated (right) fraction correct. Error bars denote 95% confidence intervals.

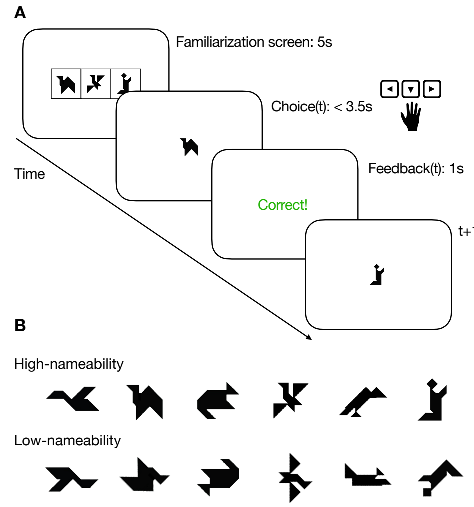

Figure 4: **Experiment 2 procedure. A**: On every trial, participants observed a symbol and selected one of three actions. If the action was correctly matched to the symbol, they received feedback that their action was correct (and incorrect otherwise). **B**: Stimuli in the *high-nameability* condition (top) and their rotated counterparts in the *low-nameability* condition (bottom).

Nameability facilitates representation learning We found that, in line in with our hypothesis that learning should be easier when features of the environment are more nameable, participants were more accurate in learning the task-relevant feature in the high-nameability condition (paired-sample t-test, $t(97) = 2.30, p < 0.05$, Fig. 3, left).

We reproduced these effects by simulating behavior from a feature reinforcement learning model that learns to update features weights in proportion to nameability (Fig. 3 right). Notably, the performance difference emerged only by allowing each feature's learning rate to vary as a function of nameability. That is, learning rates were *independently* determined by a language model that operated on raw pixels, without specifying any a priori differences between conditions in model parameters.

Discussion

In Experiment 1, participants had to learn from trial and error which features of a multidimensional environment are relevant for predicting reward. Participants were more efficient at learning the correct state representation when features of the environment were easier to name. These results lend support to the idea that attentional selection during representation learning is at least partly determined by language-based constraints on information processing (Lupyan, Rahman, Boroditsky, & Clark, 2020).

Modeling results also suggest that CLIP's multimodal representational space captures behaviorally relevant structure between concepts in natural language and images. Endowing reinforcement learning agents with such structure suggests a mechanism by which language might mediate learning through attentional selection, enhancing both the activation and separability of perceptual features during learning.

Experiment 2

In the second experiment, we tested the effects of language in an environment in which reinforcement learning has been shown to depend on working memory (Collins & Frank, 2012; Collins, Brown, Gold, Waltz, & Frank, 2014). In the original study by Collins and Frank (2012), participants learned to associate individual stimuli with three possible actions. A set-size manipulation of the number of stimuli revealed that performance drops as more stimuli are included in the state space, suggesting a contribution of working memory processes to reinforcement learning. Here, we were interested in whether more nameable features can mitigate the drop in learning performance for larger state spaces.

Method

Participants 63 participants were recruited via the online platform Prolific. They received as compensation a flat rate of \$5 plus a bonus based on their performance on the task.

Procedure Participants learned from trial and error to match symbols with one of three possible actions (Fig. 4). On each trial, participants had 3.5 seconds to select an action using the left, down and right arrow keys. After each choice, they received feedback about whether their action was correct or incorrect. The correct action was randomly and independently assigned to each stimulus. Participants were instructed that the correct action for each symbol was independent of the correct action for all other stimuli.

The task was divided into multiple blocks ('rounds' to participants). At the beginning of each round, participants were presented with all the symbols in the current round and asked to take 5 seconds to familiarize themselves with the symbols. Rounds varied in how nameable the symbols were (nameability manipulation), as well as the number of symbols participants had to learn about (set-size manipulation). Each set size (2-6) occurred once per participant, once in the high-nameability and once in the low-nameability condition.

Importantly, the sequences of stimuli in the high vs. low-nameability rounds were yoked, such that symbols were presented in the exact same order and had the same correct action associated to them, but only differed in nameability. For example, if the first symbol in Fig. 4B appeared on the first trial of a high-nameability block, its rotated counterpart would appear in the same position and be associated with the same correct action in a low-nameability block. As in the previous experiment, nameability was manipulated within participants. Each participant performed 10 blocks of the task, 5 in the high-nameability and 5 in the low-nameability condition. Within conditions, each symbol was presented 15 times in pseudorandom order. The order of blocks was also randomized within participants. We hypothesized that when features comprising the state space (i.e. the symbols in a round) are more nameable, learning would be more efficient, and that this facilitation would be especially apparent in larger state spaces.

Computational modeling

Collins and Frank (2012) proposed *Reinforcement learning with working memory* (RLWM) as a computational framework that can account for the influence of remembered experience on current decisions. A recent study based on RLWM has suggested that linguistic information primarily affects the RL component of decision-making, via enhanced learning rates when language information is available to distinguish perceptual information (Yoo, Keglovits, & Collins, 2022). Here we ask whether such an effect could be explained by the influence of nameability on reinforcement learning, as we found to be the case in Experiment 1.

The full formulation of the RLWM model that we used can be found in Yoo et al. (2022). In brief, the algorithm consists of two modules. The RL module updates the Q-value of each state-action pair with learning rate determined by each state's nameability (computed here at the stimulus level rather than per feature as in Experiment 1):

$$Q_{t+1}(s,a) = Q_t(s,a) + \eta(s)[R_t - Q_t(s,a)]. \tag{6}$$

And the the WM module updates the contents of working memory for a particular state-action pair as follows:

$$W_{t+1}(s,a) \leftarrow r_t. \tag{7}$$

Working memory for all state-action pairs is also assumed to decay on every trial with rate λ:

$$WM_{t+1}(s,a) \leftarrow (1-\lambda)WM_t(s,a) + \lambda \frac{1}{N_a} \tag{8}$$

Response probabilities associated with each module are calculated independently using softmax, with an added perseveration parameter ϕ that captures agents' tendency to repeat previous actions regardless of the current stimulus and reward.

The final response policy is assumed to be a weighted sum of the contribution from the RL and WM modules, with mixing weight ω controlling the weight of the contribution of the working memory module, and an additional parameter ε controlling random responding.

To compute nameability, we took the same approach as in Experiment 1. For all simulations, we set the stimulus-specific learning rate parameter η in the RL module to the reverse sigmoid-transformed CLIP classification entropy, multiplied by a constant scale factor. Other parameters were fixed at $\lambda = 0.2$, $\phi = 0.08$, $\beta = 100$, $\varepsilon = 0.001$ and $\omega = 0.4$.

Results

Manipulation check As for Experiment 1, we computed Simpson's diversity based on the labels that each participant provided in the manipulation check that followed the main task. Participants were more likely to use consistent labels for stimuli that had been pre-assigned to the high-nameability condition (Fig. 5A). We again observed (small) condition-level differences in CLIP classification entropy (Fig. 5B), yielding on average higher learning rates for the high- nameability condition (Fig. 5C).

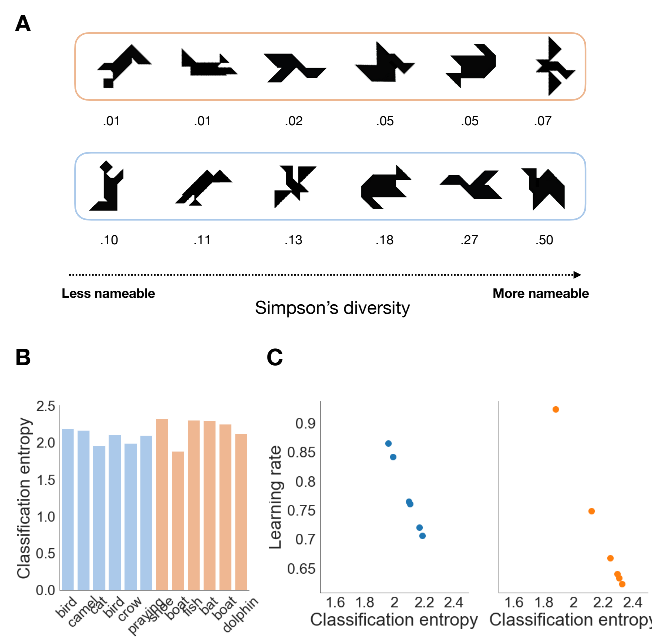

Figure 5: **Experiment 2 nameability index**. **A**: Manipulation check. Symbols ordered by Simpson's diversity. Blue denotes symbols pre-assigned to the high-nameability condition, and orange denotes symbols pre-assigned to the low-nameability condition. **B**: As for Experiment 1, classification of visual features was performed over the labels that humans most frequently used to describe each image. Each symbol and its rotated counterpart was provided to CLIP as a single image. Images that were easier to classify had lower classification entropy. **C**: Learning rates in the high- and low-nameability conditions.

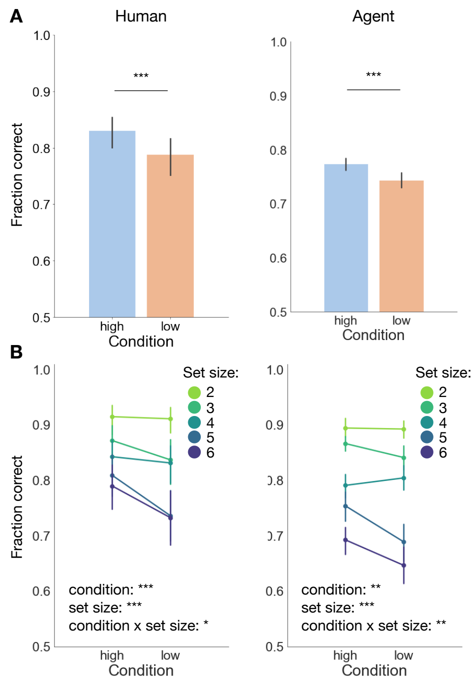

Figure 6: **Experiment 2 results**. **A**: Fraction correct as a function of nameability condition. **B**: Fraction correct as a function of condition and set size. Summary statistics from empirical and simulated data are shown on the left and right respectively. Error bars denote 95% confidence intervals.

Nameability interacts with state space complexity Our results showed that participants were generally more accurate when symbols were easier to name (Fig. 6A left, paired-sample t-test, $t(51) = 3.92, p < 0.001$). Strikingly, this effect was driven by differences in environments in which the set-size was larger (Fig. 6B left, repeated measures ANOVA, condition by set-size interaction, $F(4,204) = 2.84, p < 0.05$). In other words, as the size of the state space increased, more nameable states contributed to more efficient learning.

As was the case in Experiment 1, a simple mechanism by which nameability affects how quickly Q-values are updated was sufficient to reproduce both effects in the context of the RLWM framework (Fig. 6, right).

Discussion

In Experiment 2, we tested the hypothesis that when learning in large state spaces, humans can learn more efficiently when they can potentially utilize language to represent states. Our findings supported this hypothesis: people were more accurate when states were easier to name; and this effect was amplified for larger state spaces. Simulation results suggest a potential mechanism by which nameability specifically enhances learning from trial and error.

General Discussion

In this paper, we ask whether nameability modulates state representation in reinforcement learning. We find that people are more accurate in learning which features of a task are relevant for predicting reward when candidate features are easier to name. We also find that people are more efficient at learning and maintaining action selection policies when states are easier to name. And this facilitation by nameability is more pronounced in larger state spaces. All three findings are consistent with a computational account that assumes more nameable features are prioritized during value updating.

Our study leaves open the exact mechanism by which nameability interfaces with RL. We proposed a preliminary explanation that features of the environment that are more easily named are selectively amplified during learning, perhaps by language-related processes. However, it remains unclear if this is a language-specific enhancement or related to semantic elaboration and/or frequency or prevalance of nameable objects. In either case, such an effect could arise either via top-down processes such as selective attention or working memory; or it might reflect a bottom-up prior for which states should be learned about in the first place (Yoo et al., 2022).

If language were to influence learning via a top-down process, what might such a process consist of? Here we modeled nameability effects as arising offline due to prior representations that are shaped by language. An alternative possibility is that nameability effects could also arise in an online fashion, as participants generate candidate language labels as representations of the task (Ballard, Miller, Piantadosi, Goodman, & McClure, 2018; Radulescu, Niv, & Ballard, 2019; Vong & Lake, 2022). One avenue for future research is to determine how to use CLIP to generate candidate labels for each feature over the course of learning without human guidance.

Finally, the idea that language-based representations can ground learning has recently gained significant traction in the study of artificial agents (Andreas, Klein, & Levine, 2017; Wang & Narasimhan, 2021; Hill, Mokra, Wong, & Harley, 2020). Here we show that humans might employ similar strategies to optimize learning in complex environments. And we demonstrate the potential utility of combining modern approaches to natural language processing of raw images with reinforcement learning to capture internal representations that are otherwise difficult to formalize.

Acknowledgments

The authors thank the Gureckis lab for helpful feedback.

References

Andreas, J., Klein, D., & Levine, S. (2017). Learning with latent language. *arXiv preprint arXiv:1711.00482*.

Antonello, R., Turek, J. S., Vo, V., & Huth, A. (2021). Low-dimensional structure in the space of language representations is reflected in brain responses. *Advances in Neural Information Processing Systems*, *34*.

Ballard, I., Miller, E. M., Piantadosi, S. T., Goodman, N. D., & McClure, S. M. (2018). Beyond reward prediction errors: Human striatum updates rule values during learning. *Cerebral Cortex*, *28*(11), 3965–3975.

Braunlich, K., & Love, B. C. (2021). Bidirectional influences of information sampling and concept learning. *Psychological review*.

Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *Journal of Neuroscience*, *34*(41), 13747–13756.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035.

Eckstein, M. K., & Collins, A. G. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, *117*(47), 29381–29389.

Goh, G., Cammarata, N., Voss, C., Carter, S., Petrov, M., Schubert, L., ... Olah, C. (2021). Multimodal neurons in artificial neural networks. *Distill*, *6*(3), e30.

Gureckis, T., & Love, B. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition*, *113*(3), 293–313.

Hill, F., Mokra, S., Wong, N., & Harley, T. (2020). Human instruction-following with deep reinforcement learning via transfer-learning from text. *arXiv preprint arXiv:2005.09382*.

Ho, M. K., Abel, D., Griffiths, T. L., & Littman, M. L. (2019). The value of abstraction. *Current opinion in behavioral sciences*, *29*.

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. *The Psychology of Learning and Motivation*, 195.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.

Le Pelley, M., Haselgrove, M., & Esber, G. R. (2012). Modeling attention in associative learning: Two processes or one? *Learning & behavior*, *40*(3), 292–304.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in cognitive sciences*.

Lupyan, G., & Zettersten, M. (2020). Does vocabulary help structure the mind?

Niv, Y. (2019). Learning task-state representations. *Nature neuroscience*, *22*(10), 1544–1553.

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. *Attention and associative learning: From brain to behaviour*, 11–39.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological review*, *123*(4), 392.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... others (2021). Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*.

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: the role of structure and attention. *Trends in cognitive sciences*, *23*(4), 278–292.

Tam, A. C., Rabinowitz, N. C., Lampinen, A. K., Roy, N. A., Chan, S. C., Strouse, D., ... Hill, F. (2022). Semantic exploration from language abstractions and pretrained representations. *arXiv preprint arXiv:2204.05080*.

Tuyls, J., Yao, S., Kakade, S., & Narasimhan, K. (2022). Multi-stage episodic control for strategic exploration in text games. *arXiv preprint arXiv:2201.01251*.

Vong, W. K., & Lake, B. M. (2022). Few-shot image classification by generating natural language rules. In *First Workshop on Learning from Natural Language Supervision, ACL*.

Wang, H. J. A., & Narasimhan, K. (2021). Grounding language to entities and dynamics for generalization in reinforcement learning. *CoRR*, *abs/2101.07393*. Retrieved from https://arxiv.org/abs/2101.07393

Yao, S., Rao, R., Hausknecht, M., & Narasimhan, K. (2020). Keep calm and explore: Language models for action generation in text-based games. *arXiv preprint arXiv:2010.02903*.

Yoo, A. H., Keglovits, H., & Collins, A. (2022). The importance of linguistic information in human reinforcement learning.

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition*, *196*, 104135.