Diverse Experience Leads to Improved Adaptation: An Experiment with a Cognitive Model of Learning

Chase McDonald (chasemcd@cmu.edu)
Erin H. Bugbee (ebugbee@cmu.edu)
Erin McCormick (enmccorm@cmu.edu)
Department of Social and Decision Sciences, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Josh Fiechter (jfiechte@ball.com)
Leslie M. Blaha (leslie.blaha@us.af.mil)
711th Human Performance Wing, Air Force Research Laboratory
Pittsburgh, PA 15213 USA

Christian Lebiere (cl@cmu.edu)
Department of Psychology, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Cleotilde Gonzalez (coty@cmu.edu)
Department of Social and Decision Sciences, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

In dynamic decision tasks, the situations we confront are never the same: the world is constantly changing. Generally, our ability to generalize learned skills depends on the similarity between the learned skills and the situations in which we will apply those skills. However, in dynamic tasks, the situations we are trained in will most likely be different from the situations in which we need to apply skills. For example, in the face of emergencies, one could be trained to handle hypothetical disaster scenarios, but remain unprepared for the emergency that is actually experienced. This raises an important question: how can we best prepare for the unexpected? Cognitive science research suggests that heterogeneity during training helps people adaptat to unexpected situations. However, evidence for a general diversity hypothesis is limited. In this research, we investigate this *Diversity Hypothesis* using a cognitive model of learning and decisions from experience based on Instance-Based Learning (IBL) Theory. We focus on the concept of *decision complexity* to investigate whether confronting decisions of diverse complexities results in improved adaptation to unexpected decision complexities, compared to situations of constant decision complexity. We conduct a simulation experiment using an IBL model in a Gridworld task, and expose agents to various degrees of diversity as they learn; we then observe how these agents transfer their acquired knowledge to a situation of novel decision complexity. Our results support the Diversity Hypothesis and the benefits of diversity on adaptation.

Keywords: transfer of learning; diversity hypothesis; instance-based learning; adaptation; gridworld tasks

Introduction

Most decisions we make in life are dynamic: we evaluate potential alternatives sequentially, determine the values of the options as they develop over time, and select our options in the presence of environmental uncertainties and time constraints (Gonzalez, Lerch, & Lebiere, 2003; Gonzalez, Fakhari, & Busemeyer, 2017). Unfortunately, most research on decision making today involves static situations: decisions are often studied in one-shot choice environments, with no time constraints or high workload and where most information is provided to the decision maker (Gonzalez et al., 2003; Gonzalez, 2013). Notably, research on *heuristics and biases* has dominated behavioral decision research. For example, while demonstrating the explanatory power of Prospect Theory, one of the best known theories of risk, researchers often use monetary gambles (i.e., "prospects") that explicitly state outcomes and associated probabilities. People are presented with a description of the alternatives and are asked to make a choice based on the conditions described (Tversky & Kahneman, 1974).

In dynamic situations, decision making is considered as a learning process, in which individuals must rely on their experience to make decisions (Gonzalez et al., 2003). Importantly, by definition, dynamic situations are unique and constantly evolving. Thus, in dynamic situations, a decision maker never confronts the same exact decision situation more than once—"you cannot step twice into the same stream" (Burnet, 1930). An important research question is therefore: how can decision makers prepare for unexpected and novel situations? This question has been addressed in the learning, skill acquisition, and transfer of skills literatures. For example, it is clear that decision makers can successfully transfer learning when the skills learned during training can be reinstated at transfer, or more generally, when transfer situations share some similarity of the procedures and skills learned during training (Healy, Wohldmann, Parker, & Bourne, 2005; Healy, Wohldmann, Sutton, & Bourne Jr, 2006). While these conditions might be possible in less dynamic situations, they might be more difficult to meet in dynamic conditions of choice.

Schmidt and Bjork (1992) argued that what works best for improving performance during training will not necessarily work well in new conditions of transfer; they suggested

that diverse training might be beneficial. This idea has been tested in some studies in which diverse training appears to be particularly important for adaptation to unexpected situations (Brunstein & Gonzalez, 2011; Gonzalez & Madhavan, 2011). For example, Brunstein and Gonzalez (2011) studied effects of diverse training in a luggage screening task. They prepared targets of various categories (e.g., knifes, guns, etc.) and tested conditions in which people were trained in only one category of objects (e.g., guns) or in diverse categories (e.g., guns, knives, etc.). They observed that those individuals who were trained with diverse categories were able to classify novel items as potentially dangerous in a transfer condition, while those trained with consistent categories of weapons exhibited poor adaptation. Their conclusions suggest a general *Diversity Hypothesis*: Acquiring diverse experiences during learning will result in better adaptation to unexpected situations.

Here, we test the Diversity Hypothesis and investigate the adaptation to novel levels of decision complexity. Decision complexity is defined as in Nguyen and Gonzalez (2020): the trade-off between low-cost, low-value and high-cost, high-value alternatives. When we make decisions, we often have to handle such cost-benefit trade-offs to determine what actions to take. To test this idea, we rely on a Gridworld task developed by Nguyen and Gonzalez (2020), where agents perform a goal-seeking task under uncertainty by navigating a grid. In this situation, we test how the diversity of experienced levels of decision complexity during learning affects adaptation to unexpected levels of decision complexity. This is carried out using a cognitive model based on Instance-Based Learning Theory (IBLT; Gonzalez et al. (2003)), and we discuss the resulting predictions for human adaptation to novel decision situations.

Instance-Based Learning Theory

IBLT is a theory of decisions from experience, derived from the mechanisms proposed in the ACT-R cognitive architecture (Anderson & Lebiere, 1998), developed to explain human learning in dynamic decision environments (Gonzalez et al., 2003). IBLT provides a decision making algorithm and a set of cognitive mechanisms that can be used to implement computational models of human decision making and learning processes. The algorithm involves the recognition and retrieval of past experiences (i.e., instances) according to their relevancy to a current decision situation, the generation of expected utility of the various decision alternatives, and a choice rule that generalizes from experience. An "instance" in IBLT is a memory unit that results from the potential alternatives evaluated. These are memory representations consisting of three elements: a situation (a set of attributes that give a context to the decision, or state S); a decision (the action taken corresponding to an alternative in state S, or action A); and a utility (expected utility or experienced outcome x of the action taken in a state).

An option $k = (S,A)$ is defined by taking action A in state S. At time t, assume that there are $n_{k,t}$ different generated instances $(k, x_{i,k,t})$ for $i = 1, ..., n_{k,t}$, corresponding to selecting k and achieving outcome $x_{i,k,t}$. Each instance i in memory has an *Activation* value, which represents how readily available that information is in memory, and it is determined by similarity to past situations, recency, frequency, and noise (Anderson & Lebiere, 2014).

Here we consider a simplified version of the Activation equation which only captures how recently and frequently instances are activated:

$$Act_{i,k,t} = \ln\left(\sum_{t' \in T_{i,k,t}} (t-t')^{-d}\right) + \sigma \ln \frac{1-\xi_{i,k,t}}{\xi_{i,k,t}} \qquad (1)$$

where d and σ are the decay and noise parameters, respectively, and $T_{i,k,t} \subset \{0, ..., t-1\}$ is the set of the previous timestamps in which the instance i was observed. The rightmost term represents the Gaussian noise for capturing individual variation in activation, and $\xi_{i,k,t}$ is a random number drawn from a uniform distribution $U(0,1)$ at each time step and for each instance and option.

The probability of retrieving an instance i from memory is a function of its activation $Act_{i,k,t}$ relative to the activation of all instances:

$$p_{i,k,t} = \frac{\exp(\frac{Act_{i,k,t}}{\tau})}{\sum_{j=1}^{n_{k,t}} \exp(\frac{Act_{j,k,t}}{\tau})} \qquad (2)$$

where τ is the Boltzmann constant (i.e., the "temperature") in the Boltzmann distribution. For simplicity, τ is often defined as a function of the same σ used in the activation equation $\tau = \sigma\sqrt{2}$. Importantly, the noise and temperature values add stochasticity to the model, ensuring that action selection is non-deterministic. The nature of the model allows for exploration of the option space to reduce over time, and to treat the "explore-exploit tradeoff" without hard coded exploration (e.g., as in ε-greedy reinforcement learning methods (Sutton & Barto, 2018)).

The expected utility of option k is calculated based on a mechanism called *blending* (Lebiere, 1999), using the past experienced outcomes stored in each instance. Here we employ the blending calculation as defined for choice tasks (Gonzalez & Dutt, 2011; Lejarraga, Dutt, & Gonzalez, 2012):

$$V_{k,t} = \sum_{i=1}^{n_{k,t}} p_{i,k,t} x_{i,k,t}. \qquad (3)$$

The blending operation (Eq. 3) is the sum of all past experienced outcomes weighted by their probability of retrieval. The choice rule is to select the option that maximizes the blended value.

Experiment: Knowledge Transfer Across Decision Complexities

Gridworld Goal-Seeking Task

We use the goal-seeking Gridworld environments developed by Nguyen and Gonzalez (2020), implemented in the OpenAI

Gym framework (Brockman et al., 2016). The goal-seeking task is formalized as a Markov Decision Process (MDP), which consists of a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$ and a reward function $\mathcal{R}: \mathcal{S} \rightarrow \mathbb{R}$. We consider a solution to a MDP to be a policy $\pi: \mathcal{S} \rightarrow \mathcal{A}$. In the goal-seeking task at hand, each state $S \in \mathcal{S}$ is a (row, column) coordinate in an 11×11 grid. At each time step $l \in \{1, ..., T\}$, an agent first observes the current state S_l, takes action $A_l \in \mathcal{A}$ corresponding to one of four cardinal directions in the grid, then transitions to S_{l+1} and receives reward R_l.

In the task at hand, an agent must learn to navigate a grid environment to find one of the four outcome goals. The values of the goals are drawn from a Dirichlet distribution such that one goal, which we refer to as the *preferred goal*, is valued higher than the rest of the goals—the *distractor* goals. Interactions with the environment are broken into *episodes*. Each episode is a set of at most 31 steps, and we denote a trajectory $\mathcal{T} = \{(S_l, A_l)\}_{l=1}^{T}$ to be the sequence of state-action pairs in an episode, with $T \leq 31$ being the terminal step. The episode ends when the step limit is reached or the agent finds one of the four goals. Agents receive a penalty of -0.01 for each step taken, -0.05 for walking into walls or obstacles, and the reward of the target if they reach a goal. The optimal policy π^* is always to take the shortest path to the preferred goal. An example of a full grid is shown in Figure 1.

IBL Model in the Goal-Seeking Task

An IBL agent in the Gridworld task stores in memory *instances*, which take the form of a triplet (S, A, x), where x the value assigned to taking action A in state S. Both action and states follow from the definition of the task: the agent observes their state S as their coordinate in the grid and selects actions from the set of four cardinal directions. As previously described, the action selection mechanism dictates that the agents select the action with the highest blended value (Equation 3).

Finally, IBLT suggests a feedback process that uses decision outcomes to update and refine the utility estimates of past options, such that updated instances inform future decisions (Gonzalez et al., 2003). The present task involves rewards that are earned at the end of a task, so we must address the problem of *temporal credit assignment* (Minsky, 1961); that is, how will the agents assign delayed outcomes to their actions over the course of a trajectory $\mathcal{T}$? Here, we use a relatively simple notion of credit assignment, inspired by (Nguyen & Gonzalez, 2020), that disseminates equal credit amongst candidate actions in a sequence if a positive reward is attained, and assign the step-level reward otherwise. Formally, we have that, $\forall l \in 1, ..., T$,

$$x_l = \begin{cases} R_T & R_T > 0 \\ R_l & \text{otherwise} \end{cases} \tag{4}$$

In the context of this task, an agent will update all of its instances in a trajectory with the value of the goal reached R_T.

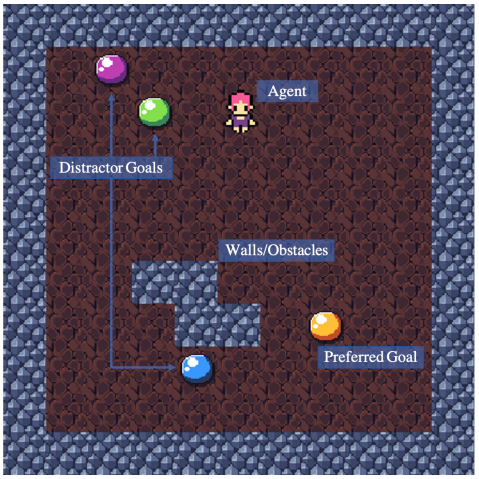

Figure 1: An example Gridworld with a randomized spawn location. The optimal policy is for the agent to reach the preferred goal via the shortest path, avoiding walls and distractor goals.

If a goal is not reached, it will simply update with the step-level cost R_l.

Experimental Simulation Methods

We investigate the diversity hypothesis by looking at within-task adaptation between levels of decision complexity.

Decision complexity is defined as in Nguyen and Gonzalez (2020): the difference between the distance to the nearest distractor d_d and the distance to the highest value goal d_p, formally defined as $\Delta_d = d_p - d_d$. Intuitively, this measure captures the tension between navigating to or discovering the preferred goal versus reaching the distractor: high values of Δ_d correspond to more complex decisions.

We separate an agent's interaction with the environment into two within-task phases: learning and adaptation. In the learning phase, each agent spawns in a specified location in a Gridworld with a predetermined level of decision complexity. The agent then executes the task over 60 episodes. In the adaptation phase, the agent remains in the same grid configuration (the same Gridworld); however, their spawn location is changed to create a different level of decision complexity. The agent then continues for another set of 60 episodes under the new level of complexity. Broadly, the agent carries over the experience from the learning phase to apply it to the adaptation phase.

We defined three decision complexity conditions:

1. High, where the agents' learning phase is in decision complexity $\Delta_d = 5$;
2. Low, with learning decision complexity $\Delta_d = 1$; and
3. Mixed, where the spawn location is randomized at each episode in the learning phase to generate various levels of complexity. The spawn position during learning is never the same as the one in the adaptation phase.

In all three conditions, the agents are required to perform under a new spawn position with decision complexity $\Delta_d = 3$, unexpectedly, after their 60th episode.

We hypothesize that the agents with the most diverse experiences—the agents in the Mixed condition—will perform better during adaptation than the agents in the Low and High complexity conditions. We expect that the agents who have been exposed to more diversity in decision complexity during learning will be able to adapt to a new decision complexity more effectively than those that learned with a consistent level of decision complexity. We also expect that agents in the Low Condition will perform better during learning than agents in the Mixed and High conditions. This is due to the variation in spawn location for the Mixed condition and increased decision complexity for the High condition.

We simulate 100 distinct grid configurations with different goal locations and obstacles. Spawn locations corresponding to the desired levels of decision complexity are generated for each distinct grid.

Our primary dependent measure is *accuracy*, defined as the proportion of episodes where the agent obtains the preferred (i.e., maximum value) goal. Using this metric, we examine agents' performance in the learning and adaptation phases in aggregate, over time, and at the transition between phases.

Results

Overall Accuracy

The average accuracy across 60 learning episodes and 60 adaptation episodes in each condition is shown in Figure 2. The results are aggregated across all 100 grid configurations, with three independent trials in each. We observe that during learning, agents in the Low decision complexity condition perform significantly better than the agents in the High complexity and Mixed complexity conditions. During the adaptation phase, however, agents in the Low complexity condition experience only a slight improvement compared to the learning phase, while agents in the Mixed complexity condition show the largest improvement from the learning phase. Agents in the Mixed condition agent are able to use the diverse experiences acquired in the learning phase to, on average, adapt more successfully than the agents in the other conditions.

Accuracy Over Time

In addition to overall average accuracy, we plotted the learning curves of the agents, which show the average accuracy per episode, grouped by the experimental condition. The results are presented in Figure 3. We observe that although the agents in the Low complexity condition learn to perform accurately very rapidly compared to the High complexity and Mixed conditions, this is the condition where agents appear to have the most difficulty adapting immediately to the new level of complexity (more discussion on this "surprise" effect in the next section). Perhaps the most interesting observation is that agents in the Mixed condition are the only ones that

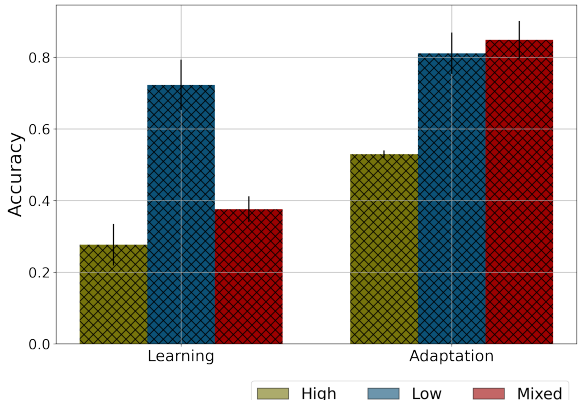

Figure 2: Average accuracy during the learning and adaptation phase for each condition.

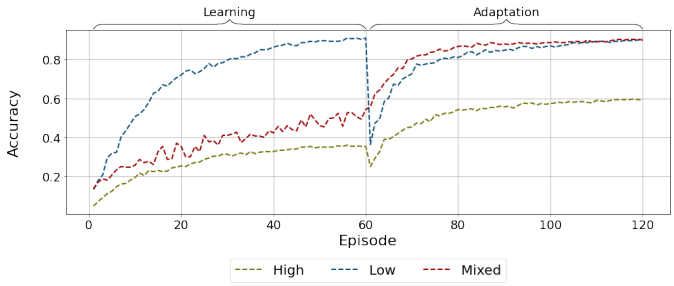

Figure 3: The learning curve for each condition. Agents transfer to the unseen decision complexity $\Delta_d = 3$ at the 60th episode, remaining there until the 120th episode.

continue to improve their performance, without an initial decrease during adaptation. During the 60 adaptation episodes, the Low complexity agents are able to match the performance of the Mixed condition agents. This contrasts with the High complexity agents, which are unable to achieve comparable levels of accuracy.

Surprise Effect

Following on the previous analysis, here we focus on the "surprise" effect per condition, characterized by both the accuracy in the first adaptation episode alone (i.e., Episode 61), as well as by the change in accuracy from the last episode of learning and the first episode of adaptation (i.e., Episodes 60 and 61). Figure 4 presents both of these measures. We observe that the agents in the Mixed complexity condition have the highest average accuracy in Episode 61. Furthermore, the difference in accuracy between Episodes 60 and 61 for the Mixed complexity agents is near zero. That is, their surprise effect is low.

In contrast, the Low complexity condition has a lower performance than the Mixed condition, but has the highest surprise effect, where the accuracy decreased more than in any of the other conditions in Episode 61. The High complexity

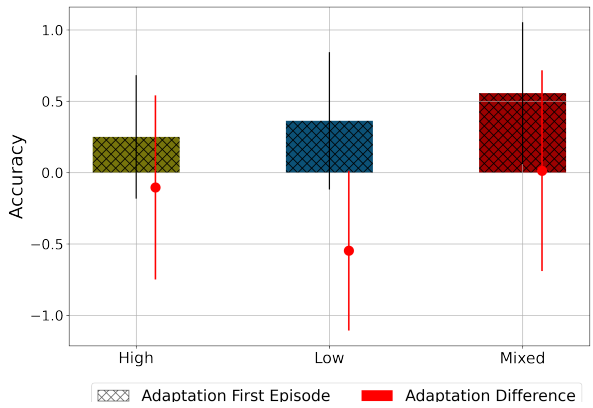

Figure 4: The accuracy in the transfer episode and the difference in accuracy between the first adaptation (Episode 61) episode and final episode of training (Episode 60).

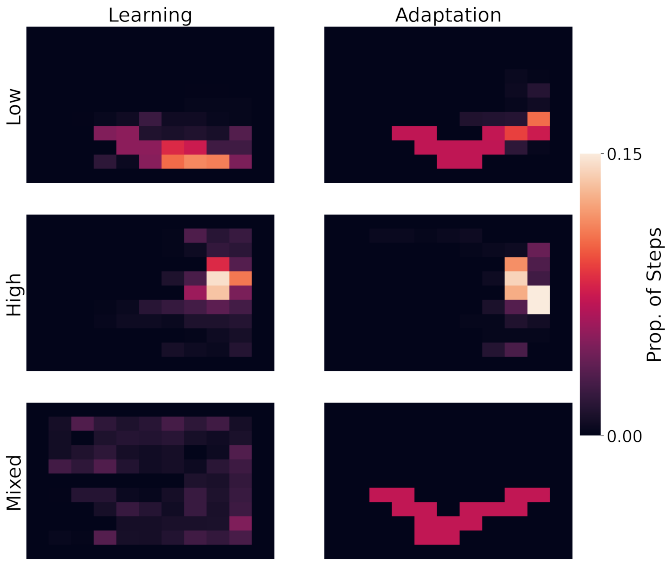

Figure 5: The proportion of times an agent in each condition visited a particular grid cell throughout the learning and adaptation phases. Each condition pictured here represents the same grid configuration.

condition has the lowest accuracy in the first episode of adaptation and a small surprise effect as it transitions to a lower decision complexity.

Explanations for the Benefits of Diversity

In this section, we dive into the mechanisms that may lead to the benefits of diversity for adaptation. A primary explanation is that the likelihood that agents will experience states during learning that are similar—or equivalent—to the states that they will experience in the adaptation phase changes across conditions.

Due to both the nature of the task and the definition of decision complexity, an agent in the High complexity condition is more likely to end up on a shorter path and fail to gain sufficient exposure to the environment to facilitate transfer. To demonstrate this, we simulate a random agent in the same training phase for the Low and High conditions and measure the average number of steps per episode. The High condition with a random agent has, on average, significantly (two-sided T-test, $p < 0.01$) shorter episodes (20.57 ± 0.74 steps) than the Low complexity condition (24.37 ± 0.59 steps). This shows that an agent is more likely to reach a goal earlier (e.g., the nearest distractor) in the High complexity condition, and thus be less exposed to the environment. This lack of exposure during the learning phase makes it more difficult for an agent to apply the instances stored in memory to new situations successfully. The memory instances will be biased towards the previously learned behavior.

As discussed, the Low complexity condition dictates a spawn location that has an increased relative distance to the nearest distractor target. The longer expected episode length in the Low condition—the same value presented above—allows agents an increased opportunity to gather diverse experiences in the environment.

Finally, the Mixed complexity condition results in a higher likelihood of experiencing states that are similar to the complexity experienced during the adaptation phase. In contrast to the above cases, this is because diversity is built into the learning phase. The agent in the Mixed condition has, by definition, a diverse set of experiences. The relative levels of diversity correspond to the relative performance in adaptation.

An illustrative example of the differences in the diversity of experiences during the learning and adaptation phases for each condition is shown in Figure 5. Considering the learning phase, the Mixed complexity condition depicts the highest level of diversity in state visitations, followed by the Low complexity condition; whereas the High complexity condition has a small and focused set of highly visited states.

The behavior in the adaptation phase demonstrates how the behavior during learning translates to the unexpected situation during adaptation phase: in both the Low and Mixed conditions, the agent is able to discover a roughly equivalent policy, whereas the High condition agent fails to learn the location of the preferred goal and a policy that will allow it to reach that position.

Discussion

In past research, the notion of diversity has been applied to motor tasks (Wulf, 1991), visual discrimination tasks (Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992) and classification decisions (Brunstein & Gonzalez, 2011; Gonzalez & Madhavan, 2011). Here, we expand this line of research to demonstrate the diversity of training in the context of decision complexity. We find that agents who learn in consistent decision complexity environments have poorer adaptation to novel and unexpected situations than those that learn with diverse decision complexity.

An interesting observation is that agents that learned in the Low complexity condition performed closely to agents in the Mixed condition during adaptation, while agents that learned in the High complexity condition are very far from reaching the level of performance of the Mixed complexity agents. An explanation we offer in our analyses is that the experiences of the agents in the Low complexity condition are quite diverse during the learning phase. By definition, a Low complexity decision would encourage the agents to navigate the Gridworld to find the target of higher value, because the decision trade-off is easy to resolve (Nguyen & Gonzalez, 2020). In other words, it is a "no brainer" to ignore the temptation of a distractor, because a larger value target is also close to the spawn location. These diverse experiences are thus applicable to a novel level of complexity at transfer, as shown in Figure 5.

In our immediate future work, we plan test both the robustness of the results to changes in environmental parameters, as well as the predictions of these simulations in human experiments. Are humans with diverse experiences in the Gridworld able to adapt more successfully to novel situations? Given that IBL models have been shown to emulate human behavior very closely in many tasks including the Gridworld (Nguyen & Gonzalez, 2021), we expect that the predictions of this paper will hold in human experiments. How far can we stretch the Diversity Hypothesis? That is, how different can the transfer conditions be to take advantage of the diversity of training? Answers to these questions can help us craft diverse training conditions and predict the way these conditions can result in robust decisions under changing and dynamic situations.

Acknowledgments

This research was supported by the US Department of Defense Award Number: POJN1006188. Distribution A: Cleared for Public Release AFRL-2021-1520

References

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought.

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym. *arXiv preprint arXiv:1606.01540*.

Brunstein, A., & Gonzalez, C. (2011). Preparing for novelty with diverse training. *Applied Cognitive Psychology*, *25*(5), 682–691.

Burnet, J. (1930). Early greek philosophy, 1892. *Bywater's no. LVIII*, 137.

Gonzalez, C. (2013). The boundaries of instance-based learning theory for explaining decisions from experience. In *Progress in brain research* (Vol. 202, pp. 73–98). Elsevier.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, *118*(4), 523–51.

Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. *Human factors*, *59*(5), 713–721.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Gonzalez, C., & Madhavan, P. (2011). Diversity during training enhances detection of novel stimuli. *Journal of Cognitive Psychology*, *23*(3), 342–350.

Healy, A. F., Wohldmann, E. L., Parker, J. T., & Bourne, L. E. (2005). Skill training, retention, and transfer: The effects of a concurrent secondary task. *Memory & Cognition*, *33*(8), 1457–1471.

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne Jr, L. E. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 534.

Lebiere, C. (1999). Blending: An act-r mechanism for aggregate retrievals. In *Proceedings of the sixth annual act-r workshop.*

Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, *25*(2), 143–153.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

Nguyen, T. N., & Gonzalez, C. (2020). Effects of decision complexity in goal-seeking gridworlds: A comparison of instance-based learning and reinforcement learning agents. In *Proceedings of the 18th intl. conf. on cognitive modelling.*

Nguyen, T. N., & Gonzalez, C. (2021). Theory of mind from observation in cognitive models and humans. *Topics in Cognitive Sciences TopiCS*.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science*, *3*(4), 207–218.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34.

Wulf, G. (1991). The effect of type of practice on motor learning in children. *Applied Cognitive Psychology*, *5*(2), 123–134.