IMATA
International Marine Animal Trainers' Association
SOUNDINGS

CURRENTLY VIEWING ISSUE
Volume 49, Number 2 — Second Quarter 2024

CAN I TRAIN A ROBOT LIKE A DOLPHIN? AN OPEN INVITATION FOR THE APPLICATION OF BEHAVIORAL SCIENCE TO ARTIFICIAL INTELLIGENCE

By Maximilian Du — Intelligence Through Robotic Interaction at Scale (IRIS) Lab Department of Computer Science, Stanford University | Volume 49, Number 2 — Second Quarter 2024

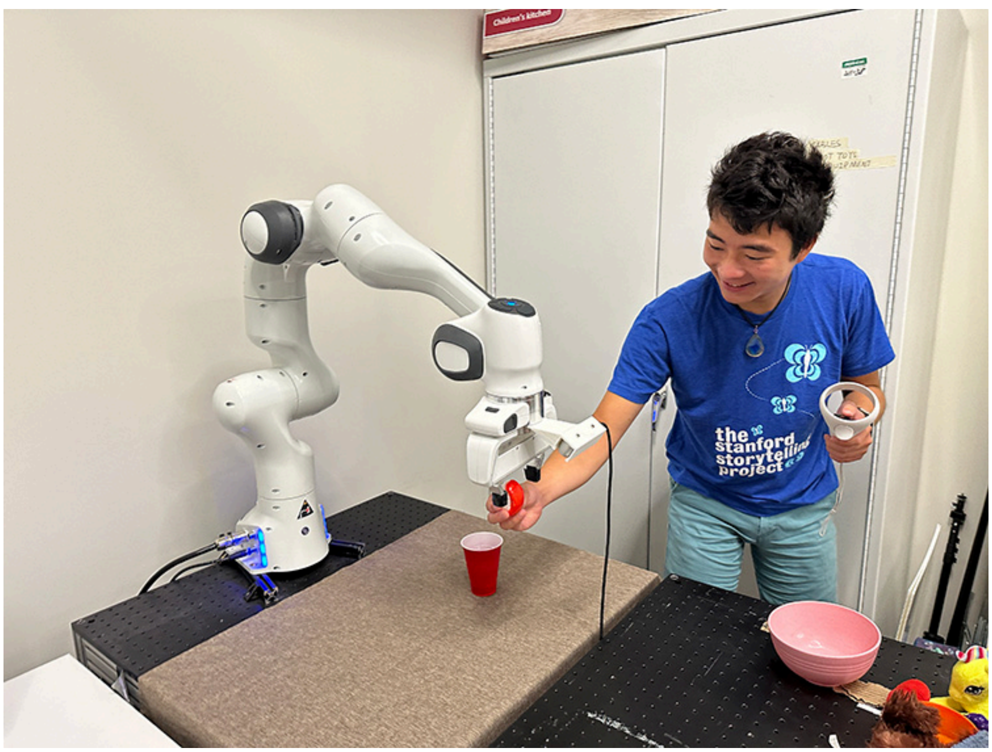
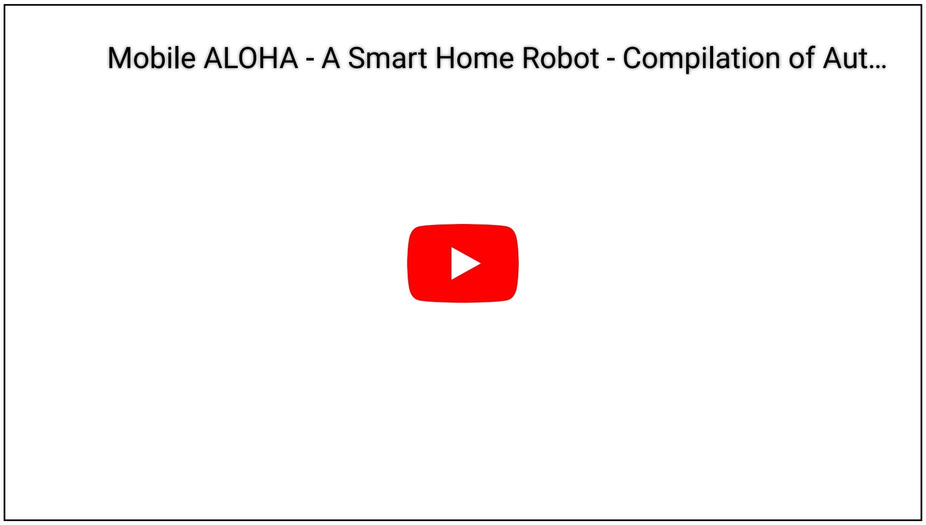


Figure 1: The author teaching manipulation behaviors to a robot arm.

Introduction.

In the field of artificial intelligence (AI), the process of creating knowledge and behaviors is also called "training." AI is a rapidly developing field that gives computers the ability to reason, like us. It has numerous applications, from chatbots to recommender systems and self-driving cars. AI is an emerging technology for conservation efforts that require tracking and identifying wild animals (Wang et al., 2023). AI can also help with animal welfare in zoological populations by monitoring behavioral patterns and even detecting health abnormalities (Winship & Jones, 2023). If executed ethically, the power of AI has the potential to benefit humans, animals, and the world we live in.



Video 1: In the future, autonomous robots like this one could help us do housework and other tasks.

Credit: Zipeng Fu and Tony Zhao, Stanford IRIS Lab

An AI trainer seeks to create productive and complex behaviors in the AI learner, and an animal trainer seeks to do the same with their animals. Perhaps unsurprisingly, these two fields have many related techniques and challenges. Similar to how animal trainers use operant conditioning methods, a widespread AI technique is called *reinforcement learning*, and it encompasses an array of methods that allow AI to learn from rewards (Sutton & Barto, 2018, p. 1). *Reinforcement learning* was inspired by early research in animal behavior and Thorndike's law of effect, which also inspired operant conditioning and Skinner's behaviorism (Kazdin, 2013, p. 16; Sutton & Barto, 2018, p. 15). Critically, beyond a surface-level similarity in the use of reinforcement, the fields of animal training and artificial intelligence often face similar challenges while creating complicated behaviors. In this paper, we will explore some of these intriguing connections between the two fields—and how they can benefit each other.

The Challenge of Credit Assignment.

For an animal to repeat behaviors that are reinforcing, it needs to know the cause of that reward. Otherwise, the animal will not know which behaviors to continue doing. In the field of AI and cognitive science, this is a problem of *causal reasoning*, and it can be quite challenging. Superstitious behaviors arise when an animal assigns an unintended cause to a reward, leading to extraneous behaviors (Ramirez, 1999, p. 552). As an example, a sea lion might incidentally learn to nod its head while offering a bedding behavior.

In AI, the *reinforcement learning* framework relies on the explicit propagation of rewards onto past actions (Sutton & Barto, 2018, p. 119), which leaves plenty of opportunities for superstitions in AI learners as well. As a hypothetical example, consider a robot that fills a glass of water, accidentally spills it, refills the glass, and hands it to a person. If the robot is rewarded at that moment, it may propagate the reward cause to both the full glass of water and the spilling procedure. In the future, it may intentionally spill the water as part of the task.

Scarcity of information is a key contributor of superstitions in AI learners and animals. Marine mammal trainers typically use a singular bridge signal to mark a reward. Likewise, it is more practical to use an all-or-nothing reward with AI because it is easier to detect overall success than to measure the progress of the target behavior (Ibarz et al., 2021). These all-or-nothing rewards are easier to give, but they leave much to interpretation, and superstitions arise from alternative reward interpretations (Ramirez, 1999, p. 552). Researchers are interested in making AI that can learn from as little information as possible, making the marine mammal training setting especially relevant.

Animal trainers have approached superstition challenges in a variety of ways. Most notably, they take advantage of the temporal effect on causality. Events that occur simultaneously tend to be interpreted as causal (Faro et al., 2013). The bridging stimulus ("bridge") can be given at the exact time of desired behavior, leaving very little room for alternative interpretations (Kazdin, 2013, p. 261). Timing and reward structure are also relevant for training AI learners, and given the strong overlap in the problem setup, advanced behavioral techniques used by animal trainers could inspire new AI training techniques that help reduce superstitions in AI.

Animal trainers also use long-term relationships with individual animals for better training outcomes. A familiar trainer can also know the types of superstitions an animal typically develops, which allows them to hold training sessions designed to eliminate superstitious behavior, or even avoid certain training approximations because of known past superstitions. These strategies demonstrate that superstition reduction should be viewed as a proactive process for both animal and AI trainers, leading to a better learning outcome for all. Superstitions can be an undesirable phenomenon, but it exposes a fundamental challenge of learning that transcends the realms of natural and artificial.

Generalization of Learned Behavior.

Animal trainers often hope that their trained behavior will generalize across many different settings, including different times of day, trainers, and locations within the habitat. Research in AI is also dedicated to improving the generalization of learned behavior. Generalizable behavior is critical for the safety of animals (Ramirez, 1999, p. 137), and it is also important for safety-critical procedures for AI learners. For example, a self-driving car should be reliable under any environmental conditions.

If an animal learns a behavior in one situation and it carries over into a new situation, then the behavior is said to have generalized (Chance, 2008, p. 301). Generalization is easier if the new situation resembles the training situations (Chance, 2008, p. 308), which means that behaviors should generally be trained under a variety of circumstances (Ramirez, 1999, p. 157). The added situational variability increases the chances of rehearsing a situation similar to the one needed for the behavior in a new situation. Analogously, a common technique to improve AI generalization is to expose the model to a high diversity of experiences (Ibarz et al., 2021; Perez & Wang, 2017). However, a key question remains: what sort of experiences help the most? Trainers decide the types of situations that they show their animals, and these decisions will impact the success of generalization. AI researchers often have large collections of data that they can expose their AI learners to, but achieving the right data variety and schedules of exposure are open research questions in creating good AI generalization.

A critical challenge for behavior generalization occurs when situations are markedly different from those encountered during training. For example, a facility undergoes construction, or a new trainer is brought onto the team. In these situations, a behavior can be greatly weakened, or it may even stop occurring. Similar failures can happen with more mundane shifts in context. If an animal is only trained for husbandry in a certain pool, it might not comply when asked in a different pool. The learned behavior of a robot or other AI learners is also significantly affected by shifts in context. In some robot experiments, a slight shift to the robot's camera or a change of room lighting can completely destroy a learned behavior (Ibarz et al., 2021). The current challenges of generalization in AI robots means that many robot experiments are conducted within the same room and setup, which is far removed from real-world situations.

Video 2: Some of the robot failures in this video are caused by lack of generalization.

Credit: Zipeng Fu and Tony Zhao, Stanford IRIS Lab

Generalization is important for successful animal behaviors, and it is a critical and unsolved problem for AI learners too, especially for robots. However, if we ever want robots and other AI learners to work safely in the real world, they must demonstrate that they can sufficiently generalize in unpredictable situations.

Successful Exploration.

Skilled animal trainers will do their best to set up their animals for success in their sessions and interactions. In contrast, if a trainer hasn't established explicit behavior criteria, the animal may not know how to offer the desired behavior in its entirety. Analogously, if we always give the same reward to a hypothetical robot chef, it may never learn how to cook our favorite foods. Any learner—natural or artificial—must *explore* the underlying reward structure of an environment in order to effectively *exploit* it (Sutton & Barto, 2018, p. 3). If the exploration is not done correctly or the reward structure is too challenging, then the learner will fail to acquire the desired behavior (Ibarz et al., 2021).

One approach to the exploration challenge is to create the behavior iteratively through *successive approximations* (Kazdin, 2013, p. 54). Critically, if the approximations are close enough together, the animal always knows how to achieve the next step. Sophisticated behavior emerges after a sequence of approximations. In AI, the analogous technique is known as a *curriculum*. For robots, we may start the robot close to the goal and bring it further back as it becomes more adept at the easier task (Florensa et al., 2018). We may also train a second robot that creates harder and harder tasks as the primary robot learns (Luo et al., 2021). However, most curriculums in AI are hand-designed by researchers, and they can be quite primitive. With advanced shaping techniques inspired by behavioral principles, we may be able to create more intricate AI behavior.

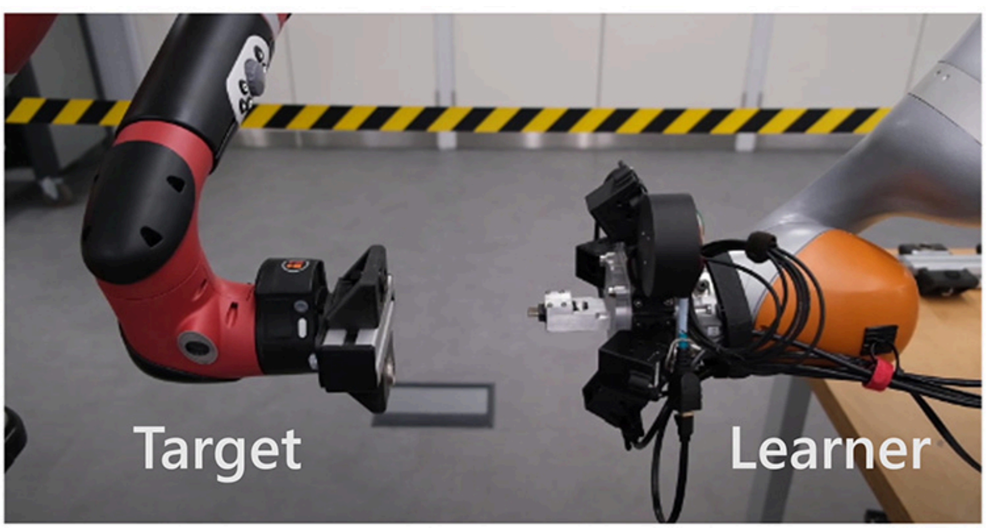


Figure 2: Robots are occasionally trained through successive approximations. Here, the red robot acts like a target for the orange robot.

Credit: Luo, J., Sushkov, O., Pevceviciute, R., Lian, W., Su, C., Vecerik, M., Ye, N., Schaal, S., & Scholz, J. (2021). Robust Multi-Modal Policies for Industrial Assembly via Reinforcement Learning and Demonstrations: A Large-Scale Study. arXiv:2103.11512

In animal training, there are also actions that trainers can take to enhance existing exploration capacities in animals. For example, enrichment can be used to enhance an animal's natural curiosity through free play (Clark, 2013; Clegg et al., 2023), which may make it more likely to engage with the trainer and explore new behaviors. Free play also allows for the spontaneous formation of new behaviors, which can be captured and placed under stimulus control (Pryor, 2000, 2006). In AI—especially for robots—researchers are also trying to get successful exploration through free play (Cui et al., 2022; Lynch et al., 2019). More generally, innate curiosity and self-driven exploration in AI are active fields of research (Schmidhuber, 1991; Ten et al., 2021).

Above all, animal trainers are skilled at identifying the strengths and deficits in an individual animal through detailed observation, and the resulting customized training sessions are highly effective in solving the exploration problem (Pryor, 2006). Likewise, customized feedback from a human evaluator can boost performance of AI learners significantly, and indeed, human feedback is responsible for creating powerful models like ChatGPT (Ouyang et al., 2022; Spencer et al., 2020). There are still many potential future innovations for productive teacher-learner interactions in AI, which may be inspired by the relationships between a trainer and their animal. A good animal trainer sets their animals up for success, and a good AI trainer will do the same.

The Future: Behavior in Artificial Intelligence.

For decades, AI researchers have looked at the natural world for inspiration. Many such works were done in a time when computational resources and architectures were less developed, meaning that the trained models were not comparable to the cognition of natural systems. Now, with the rise of large models like ChatGPT, artificial cognition is finally partially comparable to natural systems. It opens the possibility to re-explore these AI research problems through the eyes of a behaviorist. As we have discussed, animal trainers often have a highly intuitive sense for some of the hardest questions of learning—questions that AI researchers are actively working on. The fields of AI and animal training work on analogous problems, and they have the potential to benefit each other. Now is the time for these two fields to collaborate.

References

Chance, P. (2008). *Learning and Behavior*. Wadsworth Cengage Learning.

Clark, F. E. (2013). Marine mammal cognition and captive care: A proposal for cognitive enrichment in zoos and aquariums. *Journal of Zoo and Aquarium Research*, *1*(1), Article 1. https://doi.org/10.19227/jzar.v1i1.19

Clegg, I. L. K., Domingues, M., Ström, E., & Berggren, L. (2023). Cognitive Foraging Enrichment (but Not Non-Cognitive Enrichment) Improved Several Longer-Term Welfare Indicators in Bottlenose Dolphins. *Animals*, *13*(2), Article 2. https://doi.org/10.3390/ani13020238

Cui, Z. J., Wang, Y., Shafiullah, N. M. M., & Pinto, L. (2022). *From Play to Policy: Conditional Behavior Generation from Uncurated Robot Data* (arXiv:2210.10047). arXiv. http://arxiv.org/abs/2210.10047

Faro, D., McGill, A. L., & Hastie, R. (2013). The influence of perceived causation on judgments of time: An integrative review and implications for decision-making. *Frontiers in Psychology*, *4*, 217. https://doi.org/10.3389/fpsyg.2013.00217

Florensa, C., Held, D., Wulfmeier, M., Zhang, M., & Abbeel, P. (2018, July 23). *Reverse Curriculum Generation for Reinforcement Learning*. https://doi.org/10.48550/arXiv.1707.05300

Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to Train Your Robot with Deep Reinforcement Learning; Lessons We've Learned. *The International Journal of Robotics Research*, *40*(4–5), 698–721. https://doi.org/10.1177/0278364920987859

Kazdin, A. E. (2013). *Behavior modification in applied settings* (7. ed). Waveland Press.

Luo, J., Sushkov, O., Pevceviciute, R., Lian, W., Su, C., Vecerik, M., Ye, N., Schaal, S., & Scholz, J. (2021). Robust Multi-Modal Policies for Industrial Assembly via Reinforcement Learning and Demonstrations: A Large-Scale Study. *arXiv:2103.11512 [Cs]*. http://arxiv.org/abs/2103.11512

Lynch, C., Khansari, M., Xiao, T., Kumar, V., Tompson, J., Levine, S., & Sermanet, P. (2019). *Learning Latent Plans from Play* (arXiv:1903.01973). arXiv. https://doi.org/10.48550/arXiv.1903.01973

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback* (arXiv:2203.02155). arXiv. https://doi.org/10.48550/arXiv.2203.02155

Perez, L., & Wang, J. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning* (arXiv:1712.04621). arXiv. https://doi.org/10.48550/arXiv.1712.04621

Pryor, K. (2000). *Lads before the wind* (Expanded ed., 4th ed). Sunshine Books.

Pryor, K. (2006). *Don't Shoot the Dog! : The New Art of Teaching and Training*. Ringpress Books.

Ramirez, K. (Ed.). (1999). *Animal training: Successful animal management through positive reinforcement*. Shedd Aquarium.

Schmidhuber, J. (1991). *A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers*.

Spencer, J., Choudhury, S., Barnes, M., Schmittle, M., Chiang, M., Ramadge, P., & Srinivasa, S. (2020, July 12). Learning from Interventions: Human-robot interaction as both explicit and implicit feedback. *Robotics: Science and Systems XVI*. Robotics: Science and Systems 2020. https://doi.org/10.15607/RSS.2020.XVI.055

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Ten, A., Gottlieb, J., & Oudeyer, P.-Y. (2021). *Intrinsic Rewards in Human Curiosity-Driven Exploration: An Empirical Study* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/yaxvw

Wang, Z., Pang, Y., Ulus, C., & Zhu, X. (2023). Counting manatee aggregations using deep neural networks and Anisotropic Gaussian Kernel. *Scientific Reports*, *13*(1), 19793. https://doi.org/10.1038/s41598-023-45507-3

Winship, K. A., & Jones, B. L. (2023). Acoustic Monitoring of Professionally Managed Marine Mammals for Health and Welfare Insights. *Animals*, *13*(13), Article 13. https://doi.org/10.3390/ani13132124