aRtificiaL death: learning from stories of failure

Marcin Korecki[1,2,†], Cesare Carissimo[1,†] and Tanner Lund[2]

[1]Computational Social Science, ETH Zurich, Stampfenbachstr. 48, 8006 Zürich, Switzerland
[2]Ikegami Lab, Department of General Systems Studies, University of Tokyo, Komaba, Tokyo, Japan
[†]Equal Contributions

Abstract

Sharing stories, particularly about death, is an important part of many cultures. In light of these known cases of inter-generational knowledge transmission in biological systems, we explore such learning through sharing information ("stories") about death. A simulated environment with novelty-seeking Q-learning agents allows us to explore the effects of different types of information sharing on the lifespans of individual agents and the ability of inter-generational chains to maximize novelty via exploration. We find that sharing information about death provides a significantly better learning signal than sharing information about random states in the environment. Moreover, sharing shorter stories appears better than sharing longer ones. Sharing stories promotes survival and exploration in subsequent generations. This provides a foundation upon which further exploration of story sharing dynamics between agents can be explored.

Introduction

The certainty of death has been mentioned and investigated over the ages in countless works of philosophy, art, science and religion. Entire movements and schools of thought have been built around the inescapable demise of all living things. Conversely, many approaches that attempted to deny death's dominion and bring everlasting life to mankind have also been synthesized. In essence, death (whether its extrinsic possibility or intrinsic certainty) can be considered a defining property of life as we know it (Froese, 2017).

While death certainly remains a mysterious phenomenon in biological systems, it is perhaps even more unclear what it entails for artificial systems, especially those that attempt to mimic biological life. Clearly, the definition of death will be closely related to the definition of life (Gershenson, 2013). Presently, we do not attempt to provide a concrete definition but rather consider methodological implementations of death in artificial systems.

For instance: in some evolutionary algorithms, death is a probabilistic feature that might allow the optimisation process to escape local minima (Werfel et al., 2015). In many others agents 'die' when they are replaced by other agents due to computer memory limitations, and death is not modelled explicitly. However, we find that there is much more at play when death is considered at the individual rather than evolutionary level. An individual aware of its mortality will engage differently with risk taking behavior (Rosenbloom, 2003). The level of extrinsic mortality in a given environment will affect willingness to reproduce and the rate of reproduction (Quinlan, 2010). Moreover, these effects at the individual level are likely to translate into evolutionary effects as they directly affect reproduction and selection.

At a strictly subjective level the awareness of death has been linked with a certain state of anxiety in humans (Neimeyer and Van Brunt, 2018). On the other hand, positive perspectives on death from an individual, subjective point of view are also possible (Jonas, 1992) and have been present in most religions from Christianity to Buddhism and in philosophy (Schopenhauer, 1969). Regardless of its positive or negative framing it is clear that the possibility of death and interactions with it (death of kin etc.) has a strong influence on the behavior and development of individuals. However, a study of a strictly subjective perspective on death in artificial systems is a daunting challenge that perhaps is more adequately addressed by art (Greenfield and Cao, 2021). This is partly due to the fact that at present artificial systems have not been shown convincingly to possess a form of individuality or subjective perspective.

Nevertheless, what can be studied, perhaps as a kind of proxy, are the interactions between mortal, learning agents[1] - interactions induced by death in artificial systems. These include, mirroring biological life: learning from the death of others, learning about danger, and learning not to die. The ability to learn the aforementioned is of paramount importance to the sustenance of individual's life and so again is a prerequisite for evolution to take place (if an organism cannot survive sufficiently long, it will not reproduce). Learning from death can be essentially framed as a form of purposeful learning from failure (Sinapayen, 2021). It has been shown

This work was supported by an ETH Zurich Doc.Mobility Fellowship. The authors also wish to thank Dr. Michael Kaisers for the supervision while creating the first version of the code.

[1]Here we assume that an agent which learns can be seen as an individual, which can approximate the subjectivity of a perspective.

that indeed such learning occurs in dangerous conditions or as response to death in many species (Lindström et al., 2016; Griffin and Boyce, 2009). One needs only to think about humanity's discernment of edible from poisonous foods, so crucial to our survival (Turreira-García et al., 2015). Conceivably, millennia ago some early hominids died as a result of a trial and error approach to foraging, but their sacrifices were immortalised and the knowledge of discernment passed over time. It is not implausible that early cultures learned about danger in this manner. It is perhaps of special interest that in humans (but not exclusively so (Dawson and Chittka, 2014)) this learning relies on language and an intricate mesh-work of cultural traditions, stories and rituals (Anderson et al., 2018). Having evolved in interaction with a complex world and as members of inter-generational evolution, humans know that the choice of stories to be shared between generations is important Harari (2014).

The main focus of this paper is to investigate ways in which artificial agents can learn from death. With this approach we intend to bring a perspective on death as a crucial aspect of life into the artificial life community. Existing methods from reinforcement learning are used as models of individual behaviour: agents are greedy Q-learning agents that collect intrinsic novelty rewards in an environment that can 'kill' them. The novelty of our approach is that agents do not 're-spawn' upon death, but some experiences from their life can be shared with subsequent agents. We believe that this set-up captures some of life's drive for exploration and discovery and includes the risks associated with the termination of experience. In the following sections we present some related literature and give details of the implementation of the system that we investigate. Lastly, we present our results and discuss them.

Background

In this section we present the literature related to the topics of death as it has already been researched in the domain of Evolutionary Algorithms. We also look into research that investigates learning about/from danger in living systems that serves as an inspiration to our implementation and experiments. Lastly, we refer to the literature on Q-learning and Novelty Search that form the backbone of our methodology.

Death in Evolutionary Algorithms:

The concept of death has often been studied within the domain of Evolutionary Algorithms (EA). Initially, a death of an organism would occur when it would be replaced by a newly generated organism. Usually there would be a limit to the number of organisms that can be operated on at one time. This, however, would be more due to computational or memory constraints and not an intention to model or simulate death. An intrinsic death would also be introduced in EA that included a notion of energy, an organism whose energy level would drop beyond a given threshold would be removed. Deliberate introduction of an explicit death operator came later, to address issues caused by organisms evolving to escape the energy depletion death and thus stalling evolution (Todd et al., 1993). A model for death - referred to as "programmed self-decomposition" was introduced at a similar time (Oohashi et al., 1995).

Death has also been studied in Evolutionary Simulations, which were able to show that a population of initially immortal agents can evolve into limited lifespans. This suggests that there are in fact evolutionary benefits to a limited lifespan on some level (Oohashi et al., 2014). These conclusions have been supported by other publications, where it was shown that in spatial models, with local reproduction, programmed deaths robustly resulted in long-term benefits to a lineage (Werfel et al., 2015). Moreover, it has been shown that intrinsic mortality can be beneficial for the evolvability of a population (Veenstra et al., 2020).

As can be seen from the mentioned work, in EA death is mostly seen from a perspective of the collective (ie. species) and the main interest lies in its effects on evolution. We focus on a more subjective, individualistic aspect of death (and associated knowledge transfer) that manifests itself on a smaller time-scale than evolution.

Learning about Danger in Living Systems:

Learning about danger and effectively avoiding extrinsic death is a key feature at the individual level of living systems that allows for their continuous evolution. For instance: in humans, the discernment of edible from poisonous relies on inter-generational knowledge transmission and cultural behavior. While there are concerns about this knowledge decreasing, it is clear that it had been continuously passed on for many generations (Turreira-García et al., 2015).

Similar behavior can be observed in non-human animals, like the bird - common mynah, which has been shown to learn about dangerous places by observing the fate of others (Griffin and Boyce, 2009). Similarly, social animals, such as bumblebees, use social information as an indicator of safety in dangerous environments (Dawson and Chittka, 2014).

Generally, two main types of behavioral adaptation to danger have been identified. A genetic inclination to avoid certain stimuli or actions and social learning (Lindström et al., 2016). In this work our interest lies with the latter, while the former has to some extent been studied by the field of EA.

Novelty Search and Q-Learning:

The methodology that we employ to model the agents in our simulations is Q-Learning (Sutton and Barto, 2018). It is an often used and well researched model of individual learning behavior that allows us to draw parallels with inter-generational knowledge transfer. A key element in the implementation of Q-Learning agents is the reward that they seek to optimise. We choose novelty search because

it is a good model for exploratory behavior (and perhaps also a good model for life-like behavior). Since exploration can be considered to be associated with uncertainty and increased precariousness (which is a setting which we are interested in), this is a fitting framework for us. The novelty reward is also able to capture the open-endedness of evolution, which in itself fits with our aim of modelling a living system (Lehman et al., 2008). The novelty seeking behavior has also been shown to occur in animals at an individual level (Wood-Gush and Vestergaard, 1991). Novelty search has also been explicitly combined with an evolutionary simulations further reinforcing its fit for modelling life-like behavior (Lehman and Stanley, 2011).

Novelty rewards are intrinsic to an agent, in that they are not generated from the environment (extrinsic) but depend on the previous experience of an agent. For example, the novelty of an experience A may depend on the number of times A was experienced previously. Intrinsic rewards are typically applied as reward bonuses: $r = r^e + r^i$, an extrinsic reward plus an intrinsic reward. The intrinsic rewards thus provide exploration bonuses in initial phases of Q-learning and are designed to decay over time. For context, a few approaches that have been tried are estimations of learning progress (Lopes et al., 2012), and count-based exploration suitable for tabular environments and extended to continuous environments (Bellemare et al., 2016; Ostrovski et al., 2017). When only 'novelty rewards' (tabular counts, or non-tabular pseudo-counts) are considered, it is analogous to pure exploration approaches investigated in Multi-Armed Bandits (Bubeck et al., 2009), which have also been extended to *Markov Decision Processes* (Ménard et al., 2021). Their results suggest that for best-policy identification (one possible objective of pure exploration) $1/n$ reward bonuses scale better than $1/n^2$, where n is the number of times an event was experienced. It has been shown that intrinsic motivation only can be sufficient to learn many useful skills in reinforcement learning tasks (Eysenbach et al., 2018; Burda et al., 2018) in fully unsupervised learning without a reward function. Since crafting reward functions can be a consuming, possibly infeasible task for large complex environments, intrinsic rewards are promising future directions for reinforcement learning.

Methods

As we have stated, our intention is to study the effects of death on learning in an artificial system. We employ tabular novelty Q-Learning agents in a gridworld with death states. The agents explore the gridworld until they encounter a death state. Once a death state is encountered the agent dies and is removed. A new *child* agent is then deployed into the same gridworld and some of the previous *parent* agent's memory is passed on to the child. This sharing of memories is intended to model an inter-generational sharing of experiences, which in human culture commonly takes the form of stories, written and oral. In this section we provide details of our methodological set up, though the methods we employ can entirely be found in the referenced literature on reinforcement learning. Our contribution is a novel interpretation of reinforcement learning methods as inter-generational transmissions.

Q-Learning

Our learning agents are using tabular Q-Learning, which is based on *Markov Decision Process* (MDP), with the tuple $\langle \mathcal{S}, \mathcal{A}, R, P \rangle$. $\mathcal{S} \subseteq \mathbb{R}^n$ represents the set of all possible states of the environment. $\mathcal{A} \subseteq \mathbb{R}^m$ is the m-dimensional action space. $R \in (\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}$ is the reward function determining the "reward" for state s' given to the agent after selecting action a in state s. P is the transition probability function. Maximizing the cumulative reward function makes it possible for an agent to learn the action a to take, in a given state s (Sutton and Barto, 2018).

The approach to solving the MDP that we employ is called Q-learning (Watkins and Dayan, 1992). It uses a function $Q : s \times a \to \mathbb{R}$ to map state and action pairs to the reward space. The Q-function estimates the expected cumulative sum of discounted future rewards given a state s and a greedy policy π which picks the highest value action in each state. As an off-policy dynamic programming learning method, the Q-values are updated online with the Bellman update rule:

$$Q^{\text{new}}(s,a) = (1-\alpha)Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q\left(s', a'\right) \right], \tag{1}$$

where α is the learning rate, which weights the importance of new experiences and affects the speed of convergence, and γ is a discount factor that weights the importance of future rewards relative to immediate rewards.

The agent has access to a Q-function, where the Q-values for each state and action combination are stored. Following most implementations of deep Q-learning, the agent collects experiences (s, a, r, s') in a *memory buffer*, and at each iteration a random sample is drawn from the memory buffer and used to update the Q-function. Random samples from the memory buffer are an attempt to make up for violated i.i.d. assumptions needed for convergence of stochastic gradient. Though stochastic gradient descent is not used in our method, we will use the memory buffer as storage to share stories between agents.

Stories

We model inter-generational sharing by sharing experiences in memory buffers as visualized in Figure 1. An experience is a single tuple (s, a, s'). A death experience is a tuple where the state s' is a death state. When a parent dies it can share a collection of experiences with the child which form *stories*. Stories are a representation of what individuals

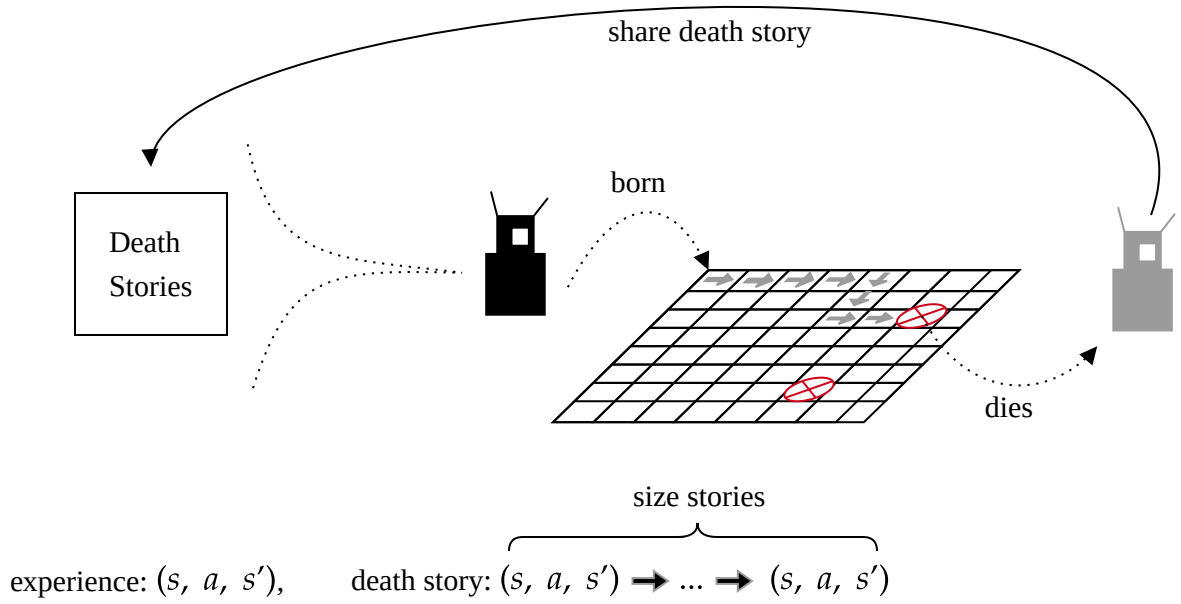

Figure 1: An illustration of death story sharing. An agent is born, inheriting whatever death stories have accumulated from its predecessors and exploring the world in search of novelty. If and when it encounters a death state, the story of where it died is added to the death stories repository for the next generation, while the Q-table and V-table are reset.

tell one another about their experiences in an environment. A single story has a *story size*, which is the number of experiences that make up the story. A story size of 1 means that only the final experience is shared. Larger story sizes include more experiences leading up to the final experience.

Then, a *death story* is a sequence of experiences that led to the death experience, e.g.

$$(s, a, s') \rightarrow (s', a, d) \rightarrow (d, a, d'),$$

where d' is a death state, and the above story has size 3.

Environment

The environment is a 15x15 grid world, seeded with randomly located *death states* as seen in Figure 2. When an agent reaches a death state, its experience terminates and it can no longer continue or collect any more reward. Thus, these death states work like 'strict' terminal states, and have added repercussions on the agents that are explained in the next section. The number of learning steps that an agent experiences before reaching a death state (also known as the length of a trajectory) is conceptualized as the *lifetime* of an agent.

The start position of each agent is fixed at the (0,0) state. In this gridworld states s are the grid tiles the agents can move to $S = \{s|s = (i, j), i, j \in [0, ..., 14]\}$, the actions a are up, down, left, right $\mathcal{A} = \{u, d, l, r\}]$, and the reward r is *novelty*. Since the environment is a discrete and small gridworld, our agents use a table, the Q-table, to update the expected cumulative sum of discounted future *novelty* rewards while they explore the environment.

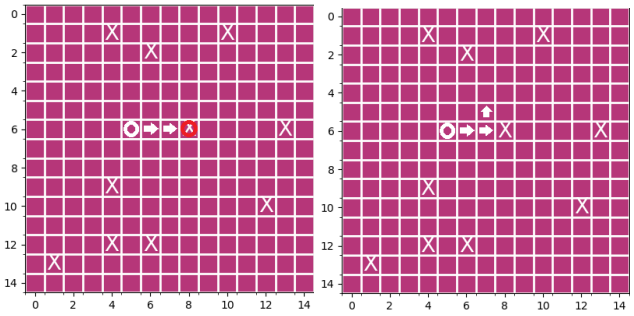

Figure 2: The 15x15 gridworld used in the experiments. Death states are marked with an "X". (Left) An agent that enters a death state ends its run. (Right) An agent may avoid the death state if their Q-values for actions that lead to other states are higher, perhaps because a parent agent died there and shared that death story.

Novelty Seeking Agents

Our agents use a greedy policy $\pi(s) = arg\max_a Q(s, a)$ while gathering intrinsic *novelty* rewards. The agents are thus *greedy explorers* that always pick the action they think will lead to the most novelty. We choose a simple and intu-

itive definition for count-based *novelty*:

$$N(s) = \frac{1}{V(s)+1}, \tag{2}$$

where $N(s)$ denotes the *novelty* of state s and $V(s)$ is the number of times state s has been visited. Since the environment is a discrete and small gridworld, our agents use a table, the V-table, to count the number of times that states are visited. They update their Q-table with novelty rewards given an experience tuple $(s, a, N(s'), s')$ using the Bellman update rule in Equation 1.

These *greedy explorers* continuously explore the environment and in it they may encounter death states. The novelty of a death state is fixed at 0, and reaching a death state terminates the exploration for an agent thus impeding the agent from further novel experiences. Whenever an agent learns from a death state that has been added to the replay buffer, the novelty estimates of the actions from the death state are all set to 0, encoding the fact that no further actions are possible from the death state. This interacts with the discount factor γ: higher values of γ increase the estimates of value of non-death states from which future actions are possible.

After the death of a parent, a child is re-initialized at the starting state with both the Q-table and the V-table reset. The Q-table is reset to random values between 0 and 1 uniformly for all (s, a) pairs, and the V-table is reset to 0 for all states. The child receives a new replay buffer which may contain some of the experiences of the parent. Thus the child will start its learning with some experiences in its replay buffer. When it samples an experience (s, a, s') from the buffer and uses it to update the Q-table, the agent will use his own V-table to calculate the novelty reward for the experience, rather than the novelty score at the time of the original experience. This is justified for two reasons:
1) The novelty is a subjective, experience-dependent quantity, and changes every time a state is visited. Thus old novelty rewards are no longer meaningful to agents in the future which may have visited a state more times since the original experience.
2) When a child is born, his V-table is reset, and therefore all experiences are now maximally novel. Then, the child samples from a replay buffer with some experiences shared from the parent and updates his Q-table using his own V-table, which is consistent with the idea that a child will find things that it has never done before to be novel.

Finally, novelty rewards monotonically decrease as a function of visits. Thus, there are progressively fewer novelty rewards to gather as agents explore the environment. When learning from death experiences (which have novelty 0), the Q-values for the novelty of actions that lead to the death state will approach 0 at a speed determined by α, so the novelty of those experiences during learning will be greater than 0. Therefore, an agent that has exhausted the novelty rewards in the states surrounding a death state may find the Q-value of an action leading to a death state to be greater than the Q-values of all other actions. Such an agent would greedily pick the action that leads to the death state and terminate his learning. To avoid such situations, agents must have sufficiently many death experiences in their replay buffer, and sufficiently long lifetimes, to be able to learn from the death experiences fast enough that the expected novelty rewards of those experiences can never be greater than the expected novelties of non-death experiences.

Experiments

We run a set of experiments to estimate the impact of story size on the lifetimes of agents, and their ability to explore the gridworld environment[2] We vary story size in the range 0 to 15, and compare three different types of replay memory that is shared between generations for which we measure and report the lifetimes of agents, their cumulative rewards, and the states they visit:

- **death:** all death stories are shared,
- **random:** shared stories are created from random samples of the replay memory,
- **both:** all death stories are shared, and the agents also get samples of random stories, thus getting *twice* the amount of stories: death + random.

The agents have the remaining parameters which will be fixed throughout all simulations: $\alpha = 0.1$, $\gamma = 0.2$, batch-size of 10, buffer-size of 1000, a randomly initialized Q-table with values between 0 and 1 uniformly, and a V-table initialized to all 0's. The maximum number of steps for an inter-generational simulation is fixed at 10000, and there is no upper bound for the lifetime of an agent meaning that a successful agent is allowed to live for the entirety of 10000 steps.

Results

In this section we report the results of our experiments. As we are interested in what the agents can learn from death of their predecessors we report the effects that sharing experiences between generations (transitions s, a, r, s') has on the average lifespan of successor. Moreover, we provide results on the degree to which the environment gets explored over generations. It is expected that these results would be correlated: maximising the lifespan should have a positive effect on the ability of an agent to explore, and maximising exploration should allow future generations to avoid most death states.

In Figure 3 we report the time series plots of the average lifespan of agents averaged over 10 inter-generational runs. A difference between top and middle (or bottom and

[2]The code allowing for replication of our experiments is available here: `https://github.com/CCarissimo/RapidRL`.

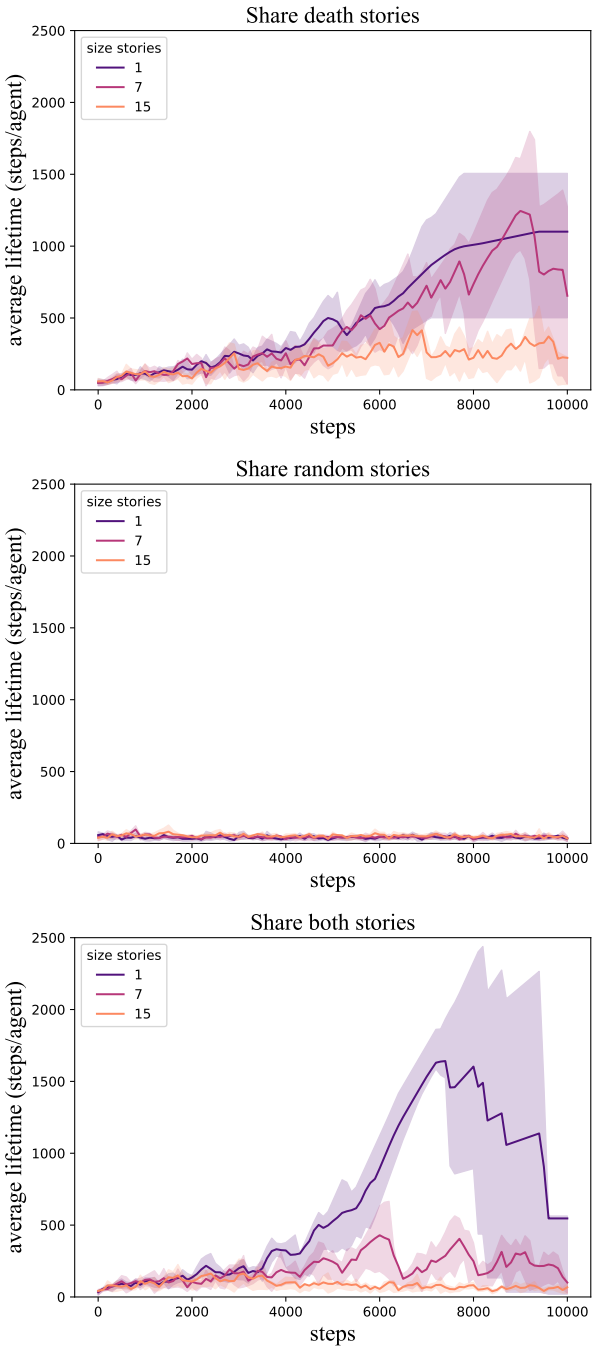

Figure 3: Average lifespan of agents over time, averaged over 10 inter-generational runs. Error bars are calculated with the 25th and 75th quantile. The rows separate three shared memory settings, top) agents share death states, middle) agents share random states, and bottom) agents share both (death + random). Each plot has three curves for different sizes of shared stories (1, 7, and 15).

middle) shows that sharing death stories has a positive effect on lifetimes, while sharing random stories has no effect. Another noticeable result is that sharing longer stories (size 7 and 15 in top and bottom) does not uniquely lead to longer lifetimes. This is explained by an over-crowding effect, whereby important death experiences are diluted in the memory buffer by less important experiences, and is greater for bottom where twice the number of experiences are shared, fewer of which are death experiences.

It is worth noting that while for the 'death' condition of size 7 and the 'both' condition of size 1 the lifetime achieves high values and then decreases rapidly. The only consistently high lifetime result is achieved by 'death' condition of size 1. This is consistent with the reasoning that at some point, while passing more information, the memory buffer becomes filled and most important information is diluted.

We also note that lifespan is correlated strongly with the cumulative reward of our agents: the longer the agent lives, the longer it is able to traverse the environment, which will yield higher rewards over time.

In Figure 4 we present the heat-maps of visits to all the states that make up the environment. We report the visits for three conditions of sharing stories of size 1, namely: for sharing the death stories, random stories and both death and random. We pick stories of size 1 since they were the most beneficial story size in Figure 3. For each of these conditions we show the visits for the first agents, agents in the middle and agents at the last agent of an inter-generational run. As can be seen, only a portion of the environment is explored at the start, with many areas not being visited at all. This is to be expected at the early stage of learning when the agents might die quickly. For top and bottom rows there is less overlap of the death experiences (marked as white X's) compared between the first and middle, and middle and last heat-maps. This indicates that the death experience shared between generations are likely to enable successor agents to avoid repeating the death experiences of predecessors. The middle row instead (for random stories) does not show this effect. At the end for the death sharing group, (top row) all of the environment has been explored (each state is visited at least once) and many of the death states have been avoided. Sharing a mix of random and death states (bottom row) presents similar results. Sharing random stories does not greatly benefit exploration (middle row): most of the world is unexplored and the agent reaches a death state early in its exploration.

Discussion

Based on our results we are able to draw some conclusions about what can be learned from death by novelty seeking Q-Learning agents at an individual level. Furthermore, we discuss these results in the light of known cases of inter-generational knowledge transmission in biological systems.

As indicated by the results, sharing information about death specifically provides a significantly better learning signal as opposed to sharing information about random states.

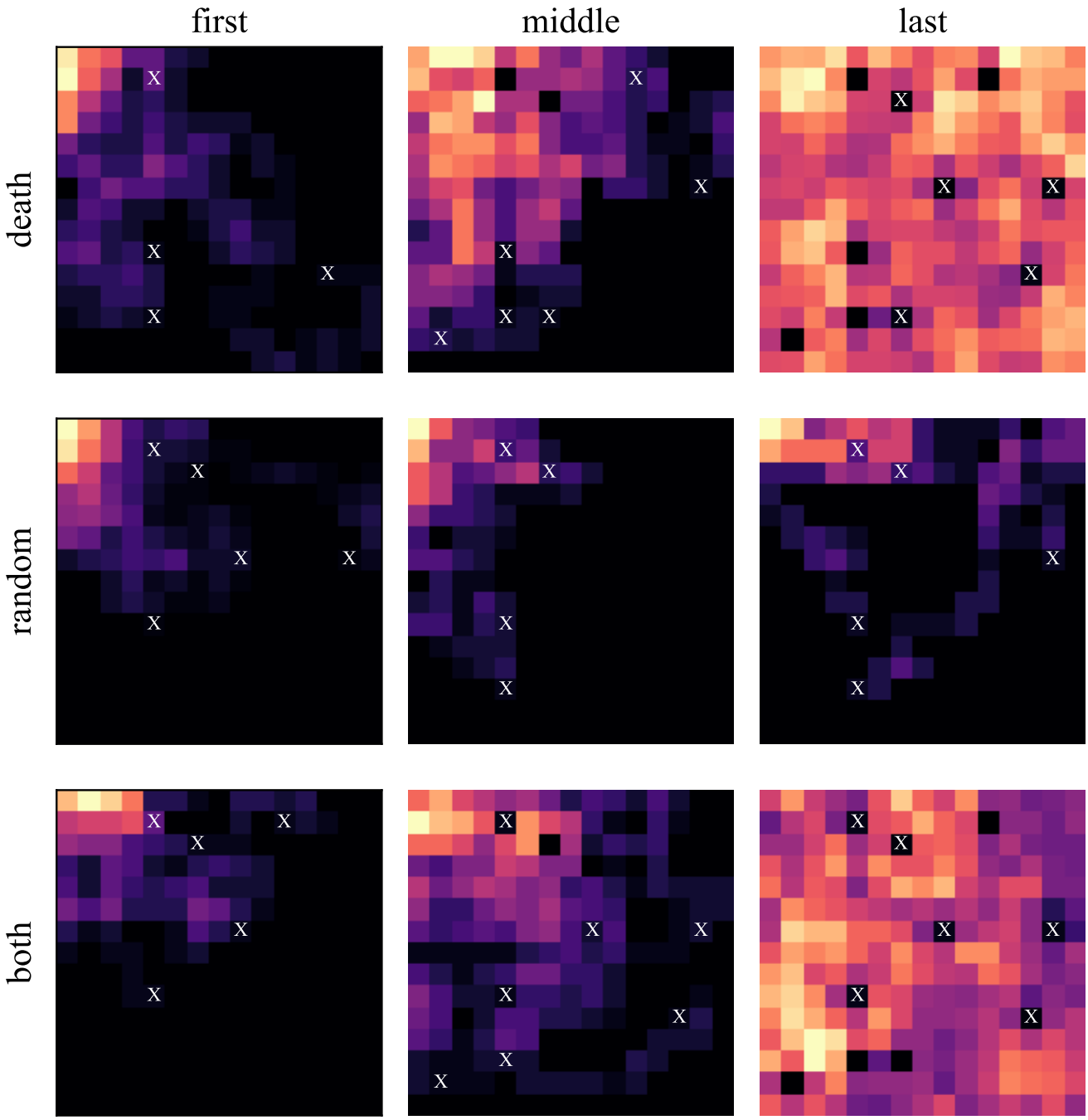

Figure 4: Heatmap colors map: high values are bright, and low values are dark, relative to a lifetime. Values are the number of visits of each state, averaged over 10 runs. Snapshots in time are taken in three periods during the inter-generational learning (columns, from left to right): the first agent, the middle agent, and the last agent. Which stories are shared is also represented (rows, from top to bottom): death, random, and both. For all runs, the story size is fixed at 1, as Figure 3 showed story sizes of 1 to be most influential in both "death" and "both" scenarios. White X's mark the states where agents died.

This is not surprising and follows the general understanding of Q-learning. In the environment that we study the death states are most important as they affect the agent's cumulative reward most and (crucially) in an irreversible way, since agents that avoid death states survive for longer and are able to accumulate more novelty rewards. What is interesting and

perhaps unexpected is that sharing the death story of size one (only death) appears better than sharing a death story of a larger size (including states leading to death). It appears that the reason for that is that the leading states do not provide as much information while drowning the essential death state information in 'noise'. This effect is very much akin to limited attention and information processing mechanisms in biological systems. For our agents a limiting factor is basically the size of their memory buffer as well as the batch size they use for updating their Q-tables. Thus, stories that are long in size may need to overwrite old memories which may overwrite death experiences. Additionally, longer stories lead to more non-death experiences in a buffer which reduce the chances of sampling death experiences from the buffer during training, thus reducing the chances that a child will successfully learn from the death experiences of its parents. In this gridworld, where the death information is much more important than any other, it is not beneficial to share much more than just the information about the death state.

We can draw parallels between this effect and cultural transmission in human societies. Indeed, it appears that most societies would focus on prioritising to identify and share the most pertinent dangers and ways of handling them rather than sharing random information. Furthermore, sharing the key information as opposed to all the little details is also a feature of much of cultural transmission (that perhaps becomes overloaded over time as more information gets accumulated). Intuitively, 'Don't eat red berries.' is a more effective meme than 'Don't go north in this forest for 20 steps, crouch next to the tree by the big bush and eat the red berries.'. Sharing only the death state information as opposed to death and leading state is akin to information distillation that also occurs in human cultures. It is perhaps also worth noting here that a given culture might not be able to immediately discern the reason for the death of their kin and so would include all the potential causes, which would then be distilled over time.

Nevertheless, the phenomenon of redundancy of information in certain cultural transmissions is present in most cultures. One could consider superstitions or folk tales to be a good example of memes that were initially likely intended to posses practical, educational value but have actually lost value and relevancy over time (this can occur also due to the failure of contemporaries to discern the intended value). Our simple model is able to capture some of these complex mechanism of cultural inter-generational knowledge transmission.

Some level of cultural reset might be valuable once a culture has accumulated too much information or the information has become irrelevant due to a changing environment. In our model this could be implemented on a large scale with a full reset, on a smaller level with a sliding memory window, or with some other forgetfulness mechanism. In any case, extending the size of the world and the length of the generational chain would allow for the "need" for forgetting or resets to manifest, if indeed it would help.

Future Work

We believe our work presents a convincing model of learning from death at an individual level that is also able to affect subsequent generations of agents, with cascading impact. The main intention of this work is to introduce the model and its features. There is, however, much interesting work to be done.

A follow-up study could observe the effects of the size of the memory buffer on knowledge transmission and its effects. The number of past experiences (corresponding to previous generations) could also be explicitly limited. Such study could identify the optimal number of 'relevant' past generations that should be remembered. As our study indicates it is likely that storing all the past generations might at some large sizes become detrimental. The memories could also be weighted, for example, with the more frequent deaths being saved differently or given more importance.

Furthermore, a limitation of our study is that the environment that we are studying is static. This allows us to focus on the knowledge transmission aspect without confounding effects of changing dynamics. Nevertheless, a more realistic, future research could consider dynamic environments that change over time or due to the agents' actions. In a similar vein, since we have shown good results for passing the death state across generations, it would be interesting to consider passing other experiences. This could be achieved by adding states with other interesting (and relevant to agents' rewards) behavior to the environment and tracking the effect of the stories told about them. It is also worth noting that reinforcement learning, which is essentially an optimisation, has some limitations in terms of its ability to fully model human behavior, which has been argued not to be reducible to optimisation (Carissimo and Korecki, 2023). While, these limitations do not disqualify it as a model of certain features of human behavior, such as the one presented in this work, it should be kept in mind that it might not be appropriate to model the entirety of human behavior with a reinforcement learning model.

We point out that generation like learning has been successfully applied in state of the art reinforcement learning research (Adaptive Agent Team et al., 2023), where the first agent trained in an environment, who tends to converge to a poor local optimum, is used as a teacher for a next-generation agent which surpasses the teacher in performance. Clearly there can be a benefit in inter-generational learning, whereby the learned shortcomings of a teacher (parent) are not passed on to a student (child). Understanding better what constitutes an experience worth sharing in artificial life is an open question. We posit that the intuition we have as humans may serve as a useful guide in answering these questions.

References

Adaptive Agent Team, Bauer, J., Baumli, K., Baveja, S., Behbahani, F., Bhoopchand, A., Bradley-Schmieg, N., Chang, M., Clay, N., Collister, A., Dasagi, V., Gonzalez, L., Gregor, K., Hughes, E., Kashem, S., Loks-Thompson, M., Openshaw, H., Parker-Holder, J., Pathak, S., Perez-Nieves, N., Rakicevic, N., Rocktäschel, T., Schroecker, Y., Sygnowski, J., Tuyls, K., York, S., Zacherl, A., and Zhang, L. (2023). Human-timescale adaptation in an open-ended task space.

Anderson, J. R., Biro, D., and Pettitt, P. (2018). Evolutionary thanatology.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., and Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, 29.

Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory: 20th International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings 20*, pages 23–37. Springer.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., and Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.

Carissimo, C. and Korecki, M. (2023). Limits of optimization. *Minds and Machines*, pages 1–21.

Dawson, E. H. and Chittka, L. (2014). Bumblebees (bombus terrestris) use social information as an indicator of safety in dangerous environments. *Proceedings of the Royal Society B: Biological Sciences*, 281(1785):20133174.

Eysenbach, B., Gupta, A., Ibarz, J., and Levine, S. (2018). Diversity is all you need: Learning skills without a reward function.

Froese, T. (2017). Life is precious because it is precarious: Individuality, mortality and the problem of meaning. *Representation and reality in humans, other living organisms and intelligent machines*, pages 33–50.

Gershenson, C. (2013). What does artificial life tell us about death? In *Investigations into Living Systems, Artificial Life, and Real-World Solutions*, pages 17–22. IGI Global.

Greenfield, R. and Cao, S. (2021). The edge of life-as-we-know-it: Aesthetics of decay within artificial life and art. *Technoetic Arts: A Journal of Speculative Research*, 19(1-2):185–201.

Griffin, A. S. and Boyce, H. M. (2009). Indian mynahs, acridotheres tristis, learn about dangerous places by observing the fate of others. *Animal Behaviour*, 78(1):79–84.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.

Jonas, H. (1992). The burden and blessing of mortality. *The Hastings Center Report*, 22(1):34–40.

Lehman, J. and Stanley, K. O. (2011). Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 211–218.

Lehman, J., Stanley, K. O., et al. (2008). Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE*, pages 329–336.

Lindström, B., Selbing, I., and Olsson, A. (2016). Co-evolution of social learning and evolutionary preparedness in dangerous environments. *PloS one*, 11(8):e0160245.

Lopes, M., Lang, T., Toussaint, M., and Oudeyer, P.-Y. (2012). Exploration in model-based reinforcement learning by empirically estimating learning progress. *Advances in neural information processing systems*, 25.

Ménard, P., Domingues, O. D., Jonsson, A., Kaufmann, E., Leurent, E., and Valko, M. (2021). Fast active learning for pure exploration in reinforcement learning. In *International Conference on Machine Learning*, pages 7599–7608. PMLR.

Neimeyer, R. A. and Van Brunt, D. (2018). Death anxiety. *Dying: Facing the facts*, pages 49–88.

Oohashi, T., Maekawa, T., Ueno, O., Kawai, N., Nishina, E., and Honda, M. (2014). Evolutionary acquisition of a mortal genetic program: The origin of an altruistic gene. *Artificial Life*, 20(1):95–110.

Oohashi, T., Sayama, H., Ueno, O., and Maekawa, T. (1995). Programed self-decomposition model and artificial life. In *Proceedings of the 1995 International Workshop on Biologically Inspired Evolutionary Systems*, pages 85–92. Citeseer.

Ostrovski, G., Bellemare, M. G., Oord, A., and Munos, R. (2017). Count-based exploration with neural density models. In *International conference on machine learning*, pages 2721–2730. PMLR.

Quinlan, R. J. (2010). Extrinsic mortality effects on reproductive strategies in a caribbean community. *Human Nature*, 21:124–139.

Rosenbloom, T. (2003). Sensation seeking and risk taking in mortality salience. *Personality and Individual Differences*, 35(8):1809–1819.

Schopenhauer, A. (1969). *The world as will and representation: In two volumes*. Dover Publ.

Sinapayen, L. (2021). Perspective: Purposeful failure in artificial life and artificial intelligence. *arXiv preprint arXiv:2102.12076*.

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

Todd, P. M. et al. (1993). Artificial death. In *Proceedings of the Second European Conference on Artificial Life (ECAL93)*, volume 2, pages 1048–1059.

Turreira-García, N., Theilade, I., Meilby, H., and Sørensen, M. (2015). Wild edible plant knowledge, distribution and transmission: a case study of the achí mayans of guatemala. *Journal of Ethnobiology and Ethnomedicine*, 11:1–17.

Veenstra, F., de Prado Salas, P. G., Stoy, K., Bongard, J., and Risi, S. (2020). Death and progress: How evolvability is influenced by intrinsic mortality. *Artificial life*, 26(1):90–111.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Werfel, J., Ingber, D. E., and Bar-Yam, Y. (2015). Programed death is favored by natural selection in spatial systems. *Physical review letters*, 114(23):238103.

Wood-Gush, D. and Vestergaard, K. (1991). The seeking of novelty and its relation to play. *Animal Behaviour*, 42(4):599–606.