To Appear In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2014), Vancouver, BC, Canada, July 2014.

Overcoming Deception in Evolution of Cognitive Behaviors

Joel Lehman
University of Texas at Austin
1 Inner Campus Drive
Austin, TX 78712 USA
joel@cs.utexas.edu

Risto Miikkulainen
University of Texas at Austin
1 Inner Campus Drive
Austin, TX 78712 USA
risto@cs.utexas.edu

ABSTRACT

When scaling neuroevolution to complex behaviors, cognitive capabilities such as learning, communication, and memory become increasingly important. However, successfully evolving such cognitive abilities remains difficult. This paper argues that a main cause for such difficulty is deception, i.e. evolution converges to a behavior unrelated to the desired solution. More specifically, cognitive behaviors often require accumulating neural structure that provides no immediate fitness benefit, and evolution often thus converges to *non-cognitive* solutions. To investigate this hypothesis, a common evolutionary robotics T-Maze domain is adapted in three separate ways to require agents to communicate, remember, and learn. Indicative of deception, evolution driven by objective-based fitness often converges upon simple non-cognitive behaviors. In contrast, evolution driven to explore novel behaviors, i.e. novelty search, often evolves the desired cognitive behaviors. The conclusion is that open-ended methods of evolution may better recognize and reward the stepping stones that are necessary for cognitive behavior to emerge.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*connectionism and neural nets, concept learning*

General Terms

Algorithms

Keywords

Diversity Maintenance; Evolutionary Robotics; Cognition; Deception

1. INTRODUCTION

A goal of neuroevolution research is to evolve artificial neural networks (ANNs) able to solve increasingly complex tasks. So far, most tasks have been largely *reactive*, i.e. behaviors that can result from reacting to current sensory input without persistent state or computation. It is still challenging to solve tasks that require cognitive behaviors such as communication, remembering previous states, or learning from experience. When such behaviors have been successfully evolved, artificial domain constraints [19, 26], specialized domain knowledge in the reward scheme driving evolution [13, 18], or specialized encodings [28] were often needed. Yet such special knowledge is difficult to come by and a more general solution would be desirable.

While it is easy to design fitness functions that *recognize* a successful instance of a cognitive behavior, it is more difficult to craft a fitness function that *paves* the path towards it [10]. In particular, to realize a cognitive behavior through an ANN often requires significant scaffolding structure, i.e. additional neurons and connections adapted to provide cognitive functionality. Yet building only one part of such scaffolding may provide no immediate fitness benefit, particularly if fitness is measured narrowly by how closely a behavior resembles evolution's target. As a result, evolution may often converge to locally optimal reactive policies, and there may be no performance gradient connecting such a reactive policy to the desired cognitive one. In such cases, evolution may remain deceived by non-cognitive local optima. This observation leads to the main hypothesis explored in this paper: deception is a significant factor making evolving cognitive behaviors difficult.

Previous studies have explored evolving cognitive behaviors such as communication [19, 26], memory [13, 28], and learning [17, 27]; yet overall, evolving sophisticated cognitive behaviors remains difficult. Importantly, some such studies support the idea that evolving cognitive behaviors may be deceptive [13, 17] and further suggest that methods that are based on novelty rather than fitness may be beneficial [17]. In particular, novelty search [8] rewards behaviors relative only to how different they are from those previously encountered during the search. Risi et al. [17] shows that novelty search succeeds significantly more often than objective-based search in evolving learning behavior.

This paper builds upon such previous results to evolve three separate cognitive behaviors. A T-Maze domain, similar to those used in previous evolutionary robotics (ER) studies [13, 17, 18, 28], is varied to form three separate tasks that require evolving communication, short-term memory, and learning from experience. Across such tasks, the performance of a variety of objective-based methods is compared to that of novelty search. To probe how such methods scale

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2576768.2598300.

with increased complexity, a more difficult extended T-Maze domain is also tested. The goal is to verify that deception is problematic across a spectrum of cognitive behaviors, and determine how it is exacerbated by increased problem complexity.

The results support the hypothesis that deception is a significant factor in cognitive tasks. The conclusion is that the stepping stones towards cognitive behaviors may not be identified by traditional objective-based fitness functions. Instead, methods that open-endedly discover structure in a domain, thereby potentially accumulating such stepping stones, are needed.

2. BACKGROUND

This section first reviews previous approaches to evolving cognitive behaviors, then describes the NEAT and novelty search methods applied in this paper's experiments.

2.1 Evolving Cognitive Behaviors

The evolution of various cognitive behaviors has been studied in artificial life and ER [12, 13, 17, 19, 26–28]. It is helpful first to review mechanisms applied in these experiments to make evolution more tractable.

First, domains are often artificially restricted to isolate and encourage cognitive behavior's evolution [19, 26]. Without such restrictions, evolution may often converge to simpler non-cognitive policies that either fail to solve the task or solve the task in trivial or unintended ways [19]. For example, Werner and Dyer [26] introduced an experimental setup in which female animals are stationary but can produce signals, whereas male animals can sense signals from females but cannot produce them nor can they perceive the environment. The advantage of such an unrealistic setup is that communication offers the only way to solve the task, thereby reducing the potential for convergence to simpler non-communicative solutions.

Other experiments demonstrated that injecting additional knowledge can improve performance in cognitive domains. Such approaches include applying incremental evolution [21], refined fitness functions [18], more complicated neural models [28] or additional helper objectives [12, 13]. Ideally, however, it would not be necessary to uncover technical domain knowledge specific to new applications.

Similarly, the benefit of encouraging higher-level (e.g. behavioral) diversity has been demonstrated in several cognitive tasks [12, 13, 17, 25]. In particular, Risi et al. [17] demonstrated that driving search *only* towards behavioral novelty can sometimes more effectively evolve learning policies than evolution driven by an objective-based fitness function. By directly comparing diversity-driven and objective-driven search in this way, deception can be effectively isolated [8, 17]; the same approach is applied in this paper.

Overall, previous results demonstrate that it is possible to evolve cognitive behaviors, but it is difficult to do so without domain knowledge or specialized approaches. Furthermore, some such studies directly implicate deception as the main obstacle [13, 17].

2.2 NEAT

In experiments described in this paper, behaviors are evolved for robots that are controlled by artificial neural networks (ANNs). Thus a neuroevolution (NE) method is needed to underpin these experiments. The NEAT method is appropriate because it is widely applied [1, 8, 17, 22–24] and well understood.

The NEAT method was originally developed to evolve ANNs to solve difficult control and sequential decision tasks [22–24]. Evolved ANNs control agents that select actions based on their sensory inputs. Like the SAGA method [5] before it, NEAT begins evolution with a population of small, simple networks and *complexifies* the network topology into diverse species over generations, leading to increasingly sophisticated behavior. A similar process of gradually adding new genes has been documented in natural evolution [11]. This section briefly reviews the NEAT method; for comprehensive introductions see e.g. [23, 24].

To keep track of which gene is which while new genes are added, a historical marking is uniquely assigned to each new structural component. During crossover, genes with the same historical markings are aligned, producing meaningful offspring efficiently. Speciation in NEAT protects structural innovations by reducing competition among differing structures and network complexities, thereby giving newer, more complex structures room to adjust. Networks are assigned to species based on the extent to which they share historical markings. Complexification, which resembles how genes are added over the course of natural evolution [11], is thus supported by both historical markings and speciation, allowing NEAT to establish high-level features early in evolution and later elaborate on them. Further, NEAT's ability to evolve increasingly complex ANNs fits well with this paper's motivation of evolving cognitive behaviors, which require potentially complex evolved structure.

Note that as originally described, NEAT speciates the population to encourage genotypic diversity. Because some experiments in this paper explore the effects of diversity maintenance on search, NEAT is extended to run without speciation as a baseline, and speciation is also replaced in some setups by the age-layered population structure (ALPS; [6]), a diversity maintenance technique encouraging diversity among the age of genomes instead of among ANN topologies.

Additionally, in some experiments in this paper NEAT is also extended with a mechanism for lifetime adaptation. In particular, in the learning task described later, ANNs are augmented with neuromodulated plasticity, a biologically-plausible [3] method for behavioral plasticity applied in similar previous learning experiments [17, 21]. In this model, an unsupervised learning rule modifies connection weights based on postsynaptic activity. In addition, special neuromodulatory neurons can be added to the ANN by random mutations; their role is to modulate the magnitude of weight changes for the neurons to which they connect. In particular, for each non-modulatory neuron, the sum of incoming activation from modulatory neurons (which do not contribute to the non-modulatory neuron's traditional activation level) determines the magnitude of weight changes for the non-modulatory node's incoming connections. In this way, neuromodulatory neurons can allow an ANN itself to decide not only *how* to adapt connection weights but also *when*. Following Risi et al. [17], the particular plasticity rule is:

$$\Delta w_{ji} = \tanh(m_i/2) * 35.95 a_i, \quad (1)$$

where w_{ji} is the weight of the connection from neuron j to i, m_i is the sum of incoming modulatory activation, and a_i is the activation of neuron i.

2.3 Novelty Search

In contrast to most EAs, which tend to converge the population, novelty search is a *divergent* evolutionary technique. It is inspired by natural evolution's drive towards novelty. In novelty search, novel behavior is rewarded directly *instead* of progress towards a fixed objective [7, 8]. The idea is that novelty can act as a proxy for many creative forces in natural evolution. In this way, rewarding only novelty allows investigating the impact of such creative forces independent of adaptive pressure to better fit a particular niche. This paper applies novelty search both as a means of illustrating deception when compared with objective-driven search, and as a potential practical alternative to such objective-driven search when evolving cognitive behaviors.

Tracking novelty requires little change to an evolutionary algorithm aside from replacing a fitness function with a *novelty metric*. Such a metric measures how different an individual is from other individuals, thereby creating a constant pressure to do something new. The key idea is that instead of rewarding performance on an objective, divergence from prior behaviors is rewarded.

In order to implement novelty search, novelty needs to be measured. A novelty metric characterizes how far away the new individual is from the rest of the population and its predecessors in *behavior space*, i.e. the space of unique behaviors. A good metric should thus compute the *sparseness* at any point in the behavior space. Areas with denser clusters of visited points are less novel and therefore rewarded less.

A simple measure of sparseness at a point is the average distance to the k-nearest neighbors of that point. Intuitively, if the average distance to a given point's nearest neighbors is large then it is in a sparse area; it is in a dense region if the average distance is small. The sparseness ρ at point x is given by

$$\rho(x) = \frac{1}{k}\sum_{i=0}^{k} \text{dist}(x, \mu_i), \tag{2}$$

where μ_i is the ith-nearest neighbor of x with respect to the distance metric dist, which is a domain-dependent measure of behavioral difference between two individuals in the search space. Candidates from more sparse regions of the behavior space then receive higher novelty scores.

If novelty is sufficiently high at the location of a new individual, i.e. above some minimal threshold ρ_{min}, then the individual is entered into the permanent archive that characterizes the distribution of prior solutions in behavior space. The current generation plus the archive give a comprehensive sample of where the search has been and where it currently is; that way, by attempting to maximize the novelty metric, the gradient of search is simply towards what is *new*, with no other explicit objective.

However, even without an explicit objective, novelty search is still driven by meaningful information. Behaving in a novel way often requires exploiting the structure of the domain. For example, if the novelty of a robot's behavior is measured across many independent trials, learning from experience in any way (i.e. not necessarily such that it improves objective performance) may help a robot to differentiate itself from previous behaviors that behave *uniformly* when exposed to repeated trials in the same environment. Importantly, such learning behavior, first evolved without objective benefit, may later facilitate evolving the task's solution.

Once objective-based fitness is replaced with novelty, the underlying evolutionary algorithm operates as normal, selecting the most novel individuals to reproduce. Over generations, the population spreads out across the space of possible behaviors.

3. THE T-MAZE DOMAIN

The approach in this paper is to compare the effectiveness of different reward schemes in variants of an ER domain designed to require evolving different cognitive abilities.

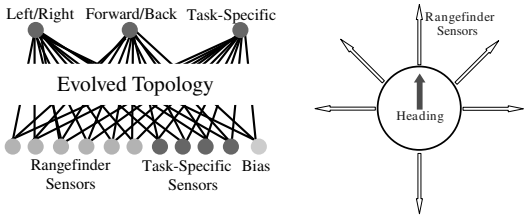
The common experimental setup consists of a simulated mobile robot embedded in a T-Maze domain typical of ER experiments [13, 17]. In the T-Maze, the robot begins in a corridor and travels to the corridor's end, where the path splits into two branches. At the end of one branch there is a high reward, while the other branch contains a low reward. The imposed time limit prevents a robot from traversing both branches, and the robot cannot differentiate rewards until one is collected. Thus a successful robot can only maximize collecting the high reward by intelligently deciding which branch to traverse.

An advantage of the T-Maze is that it offers a simple mechanism for isolating cognitive properties of robot behaviors through repeated trials of binary decision making. If the ANN controlling the robot is flushed between trials (i.e. each neuron's activation is reset), information from a robot's previous trials can affect successive trials only through explicit allowances in experimental design. For example, in experiments with communication, the goal is to evolve networks that extract information from a given trial and communicate it to the robot attempting the same task in the following trial. A task-specific output of the evolved ANN is recorded at a trial's end and is interpreted as a communication signal, which is given as an additional input to the robot at the beginning of the next trial. The task-specific output and input together constitute a single continuous-valued communication channel spanning trials; no other information can leak between trials. In this way, any systematic changes in behavior between such trials can be attributed to utilization of the communication channel. Thus if trials are sequenced such that success depends upon intelligently modulating behavior, a successful outcome ensures that the desired cognitive behavior has actually been evolved.

In contrast, when information is not isolated in such a way, evolution often discovers mechanisms simpler than intended. For example, it is often possible to exploit the environment as a form of external memory or to leverage neural activation between trials to differentiate behavior instead of through synaptic learning or explicit communication.

The robot and its ANN controller are shown in figure 1, and the two T-Maze maps such robots must explore in this paper's experiments are shown in figure 2. The Standard T-Maze map (figure 2a) is similar to the classic T-Maze imported from experiments in animal cognition [14], and has been previously explored in previous ER studies [13, 17, 21]. Note that in the Standard T-Maze map, to navigate from the starting point to one of the branches requires a relatively simple behavior: the robot must be able to go forward and decide which way to turn at the junction.

In contrast, in the Extended T-Maze map (figure 2b) the required behavior is more complex: The robot must success-



(a) Neural Network (b) Sensors

Figure 1: **The T-Maze Robot.** In the T-Maze domain studied in this paper, the ANN in (a) controls the robot. The environmental sensors of the ANN are augmented with task-specific inputs and outputs. In the Memory task, such inputs enable the robot to perceive external stimuli. In both the Communication and Learning tasks, such inputs allow the robot to perceive which reward it collects. Additionally, in the Communication task such neurons provide a channel through which to communicate and receive information between trials. The layout of the robot's sensors is shown in (b). Each arrow outside of the robot's body in (b) is a rangefinder sensor that indicates the distance to the closest obstacle in that direction. The solid arrow indicates the robot's heading.

(a) Standard T-Maze (b) Extended T-Maze

Figure 2: **T-Maze Maps.** In both maps, an unfilled circle represents the starting position of the robot and the R symbols represent the two reward locations. In each T-Maze instance, a high and a low reward will appear at the end of the maze's branches. The goal of a robot is to collect the high reward as often as possible over all of its trials. The Standard T-Maze is shown in (a), while (b) extends the T-Maze slightly to require an additional turn, interfering with the optimal turn decision later on in the maze. Note that while in both maps the high reward is depicted on the left, the rewards' locations will swap multiple times across the repeated trials of a robot's evaluation.

fully move forward, turn right, move forward to the junction, and only then decide whether to turn left or right. Thus while the extension to the maze may appear minor, the required behavior is much more complex. In this way, the idea is to explore the scalability of approaches and to determine whether deception is exacerbated as problem difficulty increases.

4. COGNITIVE TASKS

The three separate cognitive tasks tested in the underlying T-Maze domain are designed to require evolution of communication, memory, and learning (figure 3). In all three tasks, a robot is placed in either the Standard or Extended T-Maze for many independent trials (figure 4), with the goal of collecting the high reward as often as possible. However, the high reward switches branches systematically such that consistently collecting it requires evolving a particular cognitive behavior. Note that the robot's ANN is flushed between trials to control information flow between trials, and that to prevent overfitting, in each trial the robot's initial location and heading are slightly perturbed.

The Communication and Learning tasks are similar, but differ in the mechanism by which information from one trial can affect a robot's behavior in successive trials. In both tasks, robots are augmented with two sensors indicating which reward the robot reaches at the end of an evaluation. This reward information, combined with a mechanism for information transfer, makes it possible to modulate behavior intelligently in the future.

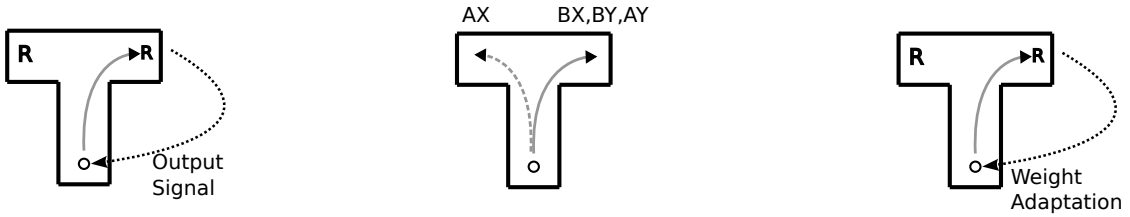
In the Communication task, such information transfer is provided by augmenting the robot's ANN with an additional output and input. Through such added neurons, the robot can communicate information to itself in the following trial (i.e. the added output is queried at the end of each trial, and is supplied to the robot as an additional input in the following trial). The communication input is set to 0 for the first trial because in that trial there is no previous communication to relay. Note that because the robot sends the signal to an identical ANN, the experiments do not investigate the evolution of language. Instead, the task tests the ability to distill information into a communication signal and to effectively modulate behavior through interpreting the signal. Thus a successful communicating agent will observe when it collects the low reward and communicate the need to visit the opposite branch to itself in the following trial.

In contrast, in the Learning task, networks are augmented with the ability to change network weights through neuromodulated plasticity as in several previous ER learning experiments [17, 21]. That is, NEAT in this particular task is extended with a model allowing the ANN to change its weights through modulated learning rules. While network activation is cleared between trials, connection weight modifications remain throughout an evaluation. Thus, a successful learning agent, after observing which reward it has collected at the end of a trial, can change its synaptic weights if necessary to guide it to the high reward in the following trial.

For the Communication and Learning tasks, how the reward location varies over the trials that constitute an individual's evaluation is shown in figure 4a. The idea is to keep the reward's location fixed long enough for adaptation to be beneficial; that is, if reward location is constant for at least three trials, an agent can consistently outperform a policy always taking the same branch by adapting to take the opposite branch whenever it collects the low reward.

The Memory task, unlike the Communication and Learning tasks, does not require learning *between* trials. As a result, no information from one trial of the robot is allowed to influence the next trial. The reason is that the tested behavior is to store information over the course of a single trial, as in [13]; in particular, the task is inspired by the AX-CPT working memory test [2] where two stimuli are presented in succession. Thus, to behave differentially over combinations of presented stimuli the ANN must remember the first one; such memory is possible in NEAT's ANNs through recurrent connections that maintain signals over time through feedback loops.

In the beginning of each trial in the Memory task, the ANN first receives as input either the A or the B stimulus. The ANN is then activated (without the robot being allowed to move) for 25 timesteps before the first stimulus is cleared. The ANN's inputs are cleared and it is then activated for 25



(a) Communication Task (b) Memory Task (c) Learning Task

Figure 3: **Separate Cognitive Tasks within the T-Maze Domain.** In all three tasks a robot's goal is to collect the high reward as often as possible. (a) In the Communication task, the robot must communicate between trials to consistently collect the high reward. An output added to the ANN is queried at the end of an evaluation, and communicated as an input in the following trial. (b) In the Memory task, a robot is first presented with time-delayed stimuli before it navigates the map. To consistently collect the high reward, the ANN must develop short-term memory through recurrent connections. (c) In the Learning task, the ANN model is extended such that a robot can modify its ANN's connection weights to maximize collecting the high reward before it switches locations again.

(a) Communication and Learning Tasks

(b) Memory Task

Figure 4: **Schedule of Reward Locations for Different Cognitive Tasks.** The locations of the high and low reward vary over the trials composing a robot's evaluation. Figure (a) shows the locations for the 50 successive trials of the Communication and Learning tasks and figure (b) the locations for the 24 successive trials of the Memory task. Filled rectangles indicate that the high reward is on the left branch of the T-Maze, while unfilled rectangles indicate that it is on the right branch. Note that the schedule is the same for the Communication and Learning tasks, and the number of trials before the reward switches is varied to prevent overfitting. In the Memory task, the presented stimuli cycle every four trials between AX, AY, BX and BY. Because the test is to distinguish AX from the other stimuli, in the first trial and every fourth succeeding trial the high reward is on the left branch, and for all others it is on the right branch.

additional timesteps without any stimulus. Then the ANN receives as input either the X or Y stimulus for 25 timesteps. Following the presentation of the stimuli, the robot is then allowed to navigate through the T-Maze. If the stimuli presented in a particular trial are A and X, then the high reward is placed on the left branch of the T-Maze; in the other three situations (i.e. AY, BX, and BY) the high reward is placed on the right branch. Thus when presented with either the X or Y stimulus, only by remembering whether the A or B stimulus had been presented previously can the robot consistently choose the correct branch. The schedule of reward locations for each trial in an individual's evaluation is shown in figure 4b.

5. METHODS COMPARED

For all experiments in this paper the underlying NEAT neuroevolution algorithm [23] is applied. However, the evolutionary incentive and type of diversity maintenance are varied to explore the degree of deception in cognitive domains. In particular, several objective-driven methods are compared to a non-objective search method.

For the objective-driven methods, the main evolutionary incentive is provided by an objective-based fitness function typical of such domains. In each trial a robot's fitness is 0 if it fails to collect any reward, 250 if it collects the low reward, and 500 if it collects the high reward. A robot's overall fitness is the sum of its fitness score over all of its trials. The intuitive motivation for this fitness function is that it reflects the desirability of outcomes and is similar to those applied in previous ER T-Maze domains [13, 17, 21].

The motivation for the objective-based methods is to apply a range of representative techniques. In total, four objective-based methods are evaluated:

- The **Baseline** method is a neuroevolution algorithm without any diversity maintenance. That is, for the baseline runs NEAT's standard diversity maintenance method, speciation, is disabled. The motivation is to determine whether a more sophisticated technique is necessary to solve the tasks, and to provide a baseline against which to compare more sophisticated methods.
- In the **High Mutation** method, mutation rates are increased to make adding neurons and connections to the ANN more likely; such increased structural mutation rates may encourage evolving more complex ANNs. This method follows a common practical principle of simply increasing mutation rates as remedy for premature convergence.
- The **Speciation** method is identical to the Baseline setup except genotypic speciation is enabled, i.e. this method is the standard NEAT algorithm as described in Stanley and Miikkulainen [23]. The purpose of genotypic speciation is to encourage exploring ANNs with diverse topologies. In this way, speciation may aid evolving cognitive behaviors because realizing them may require evolving more complex ANNs.
- Finally, the **ALPS** method augments NEAT with an implementation of ALPS [6], which is a popular ER diversity maintenance technique. The basic idea in ALPS is to protect younger genotypes from competing with older genotypes by segregating genomes based on their age. This method is complemented by regular infusion of randomly-generated genotypes. In this way, search never completely converges to a single local optimum. For a more complete description of ALPS, see Hornby [6]. Note that speciation is also disabled in this method so that there is only one diversity-maintenance mechanism being tested.

With the aim of illustrating deception, a non-objective-driven setup is also evaluated. In particular, the **Novelty**

method rewards only behavioral novelty (through novelty search), instead of progress towards the desired objective. Note that NEAT's speciation is disabled in this method because preliminary experimentation showed that it decreased performance.

Applying novelty search requires defining a space of behaviors through which it can search. The approach taken in this paper is to define such a space from high-level summary properties of a robot's behavior across the trials of a cognitive task. The idea is to create a tractable space of behaviors that is similar in spirit to how a human might summarize a robot's performance. In this way the resulting measure of novelty might grossly reflect a human's measure of behavioral distance. For example, in the Memory task, one might summarize a robot's behavior according to which branches the robot most often visits and how often it collects the high and low rewards. Thus novelty search's behavioral characterization in this task includes four values: what fraction of trials the robot visits the left branch, visits the right branch, collects the high reward, and collects the low reward.

In the Communication and Learning tasks a human observer would likely pay special attention to a robot's behavior in the trials when the location of the high reward switches branches, and in the trial immediately following such a switch. Such trials are informative because they encompass when the robot is expected to make a mistake and then adapt its behavior. Thus in both these tasks, beyond the four fractions measured in the Memory task, four additional fractions are added to the behavioral characterization: the fraction of trials in which the robot reaches the high and low rewards, for both trials in which the rewards switch locations and the trials that immediately follow such a switch.

Note that when novelty search explores such a space, it is rewarded for finding combinations of such fractions different from those the search has encountered in the past, with no fixed objective. In other words, if neither behavior has yet been encountered, a behavior maximizing the robot's ability to crash without collecting any rewards is as viable as a behavior that consistently collects the high reward. A more granular behavioral characterization was explored in initial experimentation consisting of concatenating together the outcomes of each trial. However, it performed worse than the coarse behavioral characterization described in this section, which was then adopted for the experiments.

6. EXPERIMENTS

All experiments were run for $500,000$ evaluations. For all methods except ALPS, the steady-state rtNEAT algorithm was applied with a population size of 500. For ALPS a generational algorithm based on Hornby [6] was implemented, with five age-layers of 100 individuals each (for a total population size of 500) with a polynomial aging scheme and an age gap of 20. For all methods except High Mutation, NEAT's add node mutation probability was 0.03 and its add link probability was 0.15. In High Mutation, these probabilities were raised to 0.1 and 0.3, respectively. For all methods, the weight mutation power was 2.5 and the survival threshold was set to 0.4.

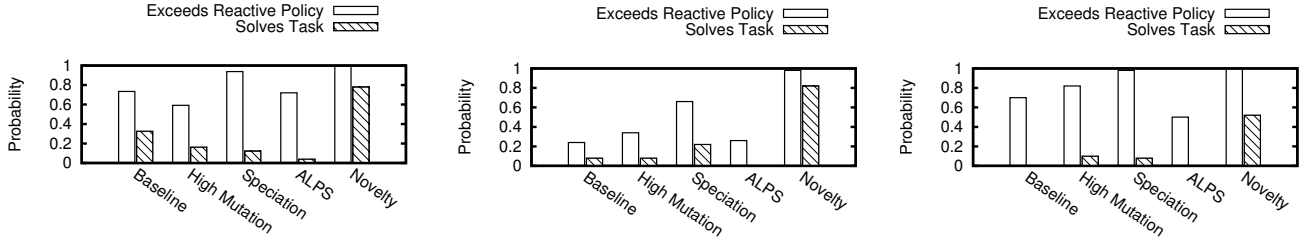
Figure 5 compares the performance of the objective-based search methods with that of novelty search across the three cognitive tasks in the Standard T-Maze map. Indicative of deception, in these tasks novelty search evolves solutions in more than half of all runs, while the objective-based methods taken together evolve solutions in less than 20% of all runs. In particular, in every pair-wise comparison between novelty search and the objective-based variants in each domain, novelty search solves the task significantly more often (Fisher's exact test; $p < 0.01$). Additionally, of the 150 runs of novelty search in the Standard T-Maze map, only one failed to evolve a behavior that exceeded the performance of the simplest purely-reactive controller (i.e. one that always turns the same direction at the end of the T-Maze). In contrast, objective-based search more often failed to do so, particularly in the Memory task.

Figure 6 similarly compares performance in the more difficult Extended T-Maze. In every pair-wise comparison between novelty search and the objective-based variants, novelty search both solves the task more often and is more likely to evolve a policy outperforming purely reactive behavior (Fisher's exact test; $p < 0.01$). Overall, performance of all methods declines in this more difficult version of the T-Maze. However, novelty search still consistently evolves behaviors outperforming the basic reactive policy, and evolves solutions in a significant percentage of runs. In contrast, the only combination of objective-based method and cognitive task to be solved in more than two out of 50 runs was High Mutation evaluated on the Communication task. Performance degradation was particularly acute in the Memory task, where the objective-based methods never solved the task and novelty search solved it in only 16% of runs.

Based on a pair-wise comparison over all tasks and domains, Speciation was most effective out of the objective-based methods in exceeding the performance of a reactive agent (Fisher's exact test; $p < 0.01$). However, Speciation did not evolve *solutions* to tasks more consistently. Thus while directly encouraging accumulating ANN structure may sometimes help narrowly to increase performance it may not always be enough to overcome deception.

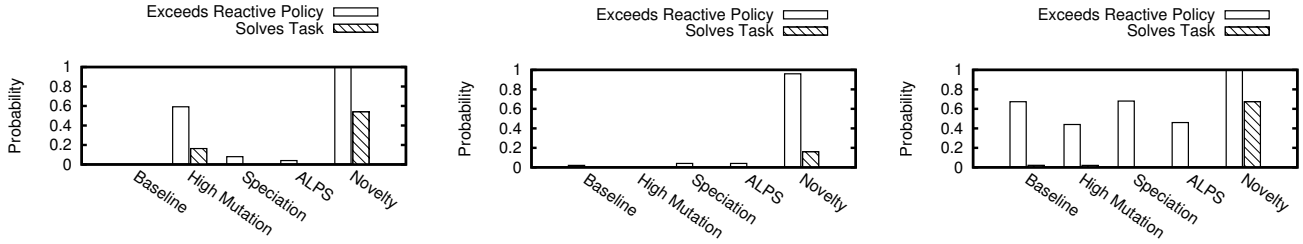
7. DISCUSSION AND FUTURE WORK

The results in this paper support the hypothesis that deception is a central factor underlying the difficulty to evolve cognitive behaviors. Such a result is important because it indicts objective-focused selection pressure rather than the encoding or evolutionary algorithm itself, which are more often the focus of refinement. The tendency of objective-based search to converge to reactive behaviors highlights that the necessary stepping stones leading to solving a cognitive task often may not lie conveniently on the gradient of increasing behavioral similarity to a task's solution. For example, Risi et al. [17] illustrated that the stepping stones towards evolving learning behavior often perform worse according to the task objective than the population so far. Similarly, figure 7 plots the objective fitness scores for the lineage of a solution for the Communication task in this paper: The innovations facilitating the desired cognitive behavior are not recognized by the objective-based fitness function. While deception has previously been shown to become increasingly problematic as problems become more complex [9, 10], this study is the first to show that deception may be systemic when evolving cognitive behaviors in particular. To overcome such deception likely requires refining selection. However, the results here show that diversity maintenance techniques alone are not always sufficient. The reason is that large digressions from the gradient of objective-based fitness are needed to establish the neural structure that enables cognitive behav-



(a) Communication Task (b) Memory Task (c) Learning Task

Figure 5: **Performance in the Standard T-Maze across Cognitive Tasks.** The ability for different methods to exceed the performance of a reactive agent that traverses the same branch every trial, and to successfully solve the task, is shown for (a) the Communication task, (b) the Memory task, and (c) the Learning task. Indicative of deception, novelty search solves each of the tasks more consistently, and the objective-based methods fail to consistently evolve policies outperforming a purely-reactive one.



(a) Communication Domain (b) Memory Domain (c) Learning Domain

Figure 6: **Performance in the Extended T-Maze across Cognitive Tasks.** The ability for different methods to exceed the performance of a reactive agent that traverses the same branch every trial, and to successfully solve the task, is shown for (a) the Communication task, (b) the Memory task, and (c) the Learning task. Illustrating that deception is worse in this more complex T-Maze variation, all of the objective-based methods struggle to consistently evolve solutions outperforming reactive policies. Novelty search's performance also degrades, although it still solves each task in a significant percentage of runs.

ior. In particular, cognitive behavior often requires integrating multiple independent adaptations that taken alone do not increase performance in benchmark tasks like those studied in this paper. For example, for communication to improve performance in the T-Maze the information necessary for correctly adapting behavior in the following trial must be first distilled, such information must then be propagated to the communication output of the ANN, and the information from the communication input must be used to actually modify behavior in the following trial. In this way, developing any of these adaptations in isolation will not improve performance, which may undermine their evolution by objective-driven search.

In contrast, novelty search rewards fledgling instances of cognitive ability that enable novel behavior in the T-Maze even if it serves no objective purpose. The performance advantage of novelty search in this paper both highlights deception and points to the importance of open-ended search processes in general when tackling problems requiring higher-complexity behaviors. Beyond novelty search, other examples of more open-ended searches, such as multi-objectivization [12, 13, 25], artificial life virtual worlds [16], and methods driven by curiosity [15, 20] are thus also promising approaches for generating complex cognitive behaviors. In particular, combining pressure towards behavioral diversity with pressure towards an objective is often an effective approach in practice [10, 13, 25]; however, searching only for novelty (i.e. without any objective-based optimization [8]) is used in this paper to isolate the role of deception.

In addition to open-ended search processes, more open-ended *domains* may also be necessary to scale to increasingly complex cognitive tasks. The problem is that domains imported from animal cognitive tests, like the T-Maze, narrowly constrain possible agent behavior to a single highly-structured task. In doing so, such domains similarly constrain evolution's creativity, which may often proceed through exapting structures evolved for one purpose

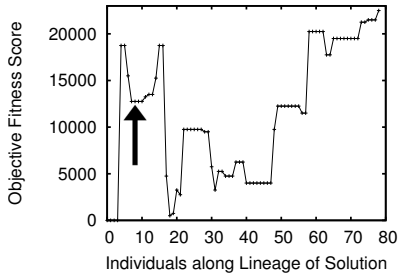


Figure 7: **Deceptive Stepping Stones in the Communication Task.** The objective fitness scores are shown for the ancestors of the eventual solution evolved in a representative run of novelty search in the Standard T-maze for the Communication task (crossover was disabled for this particular run to enable a single line of descent). Key innovations that facilitate the eventual solution receive lower fitness scores than their predecessors, and therefore would be unlikely to be preserved by objective-based search – yet novelty search can build on them because they are deemed more novel than their predecessors. For example, the ancestor highlighted by the arrow receives a fitness score of $12,750$, which is a significantly worse than the one received by the simple reactive policy of always going to the same branch of the T-maze ($18,750$), discovered earlier in evolution (by the fifth individual in the lineage). Yet because the lower scoring policy reacts when the reward location changes, it proves to be a better stepping stone to an effective solution policy.

to suit another [4]; if every behavior must relate to navigating through a T-Maze, there is little to exapt. In other words, both objective-based search and constrained domains may serve to focus evolution, when what may sometimes be needed is *less* directed focus.

An interesting direction for future work is to apply novelty search to a single task that requires integrating multiple cognitive behaviors. For example, evolved agents able to communicate and learn could potentially enable effective teams for Robocup soccer or other complex group-based tasks.

8. CONCLUSION

This paper explored the hypothesis that deception complicates evolving ANNs capable of cognitive behavior. Evidence for such a hypothesis was provided through experiments contrasting objective-driven search with novelty search in tasks requiring communication, memory, and learning. Overall, novelty search more effectively evolved solutions across the tested tasks, although the performance of all tested methods declined in the harder domain variant. The conclusion is that evolving cognitive behaviors may require both increasingly open-ended search processes and domains.

9. ACKNOWLEDGEMENTS

This research was supported in part by NSF grants DBI-0939454 and IIS-0915038, by NIH grant R01-GM105042, and by the FRI program of the College of Natural Sciences at the University of Texas at Austin.

References

[1] T. Aaltonen et al. Measurement of the top quark mass with dilepton events selected using neuroevolution at CDF. *Physical Review Letters*, 2009.

[2] Todd S. Braver, Jonathan D. Cohen, and David Servan-Schreiber. A computational model of prefrontal cortex function. *Advances in Neural Information Processing Systems: 7*, 7:141, 1995.

[3] Thomas J. Carew, Edgar T. Walters, and Eric R. Kandel. Classical conditioning in a simple withdrawal reflex in aplysia californica. *The Journal of Neuroscience*, 1(12):1426–1437, 1981.

[4] Stephen Jay Gould and Elisabeth S. Vrba. Exaptation-a missing term in the science of form. *Paleobiology*, pages 4–15, 1982.

[5] Inman Harvey. *The Artificial Evolution of Adaptive Behavior*. PhD thesis, School of Cognitive and Computing Sciences, University of Sussex, Sussex, 1993.

[6] Gregory S. Hornby. ALPS: the age-layered population structure for reducing the problem of premature convergence. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 815–822, New York, NY, USA, 2006. ACM.

[7] Joel Lehman and Kenneth O. Stanley. Exploiting open-endedness to solve problems through the search for novelty. In *Proc. of the Eleventh Intl. Conf. on Artificial Life (ALIFE XI)*, Cambridge, MA, 2008. MIT Press.

[8] Joel Lehman and Kenneth O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evol. Comp.*, 19(2):189–223, 2011.

[9] Joel Lehman and Kenneth O. Stanley. Novelty seach and the problem with objectives. In *Genetic Programming in Theory and Practice IX (GPTP 2011)*, chapter 3, pages 37–56. Springer, 2011.

[10] Joel Lehman, Kenneth O. Stanley, and Risto Miikkulainen. Effective diversity maintenance in deceptive domains. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2013)*. ACM, 2013.

[11] Andrew P. Martin. Increasing genomic complexity by gene duplication and the origin of vertebrates. *The American Naturalist*, 154(2):111–128, 1999.

[12] Charles Ollion, Tony Pinville, and Stéphane Doncieux. Emergence of memory in neuroevolution: impact of selection pressures. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion*, pages 369–372. ACM, 2012.

[13] Charles Ollion, Tony Pinville, and Doncieux Stéphane. With a little help from selection pressures: evolution of memory in robot controllers. In *Artificial Life*, volume 13, pages 407–414, 2012.

[14] David S. Olton. Mazes, maps, and memory. *American Psychologist*, 34(7):583, 1979.

[15] Pierre-Yves Oudeyer. Intelligent adaptive curiosity: a source of self-development. In *Proceedingsof the Fourth International Workshop on Epigenetic Robotics*. Lund University Cognitive Studies, 2004.

[16] Thomas S. Ray. Evolution, complexity, entropy and artificial reality. *Physica D: Nonlinear Phenomena*, 75 (1):239–263, 1994.

[17] Sebastian Risi, Charles E Hughes, and Kenneth O Stanley. Evolving plastic neural networks with novelty search. *Adaptive Behavior*, 18(6):470–491, 2010.

[18] Sebastian Risi and Kenneth O Stanley. A unified approach to evolving plasticity and neural geometry. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8. IEEE, 2012.

[19] Gregory M. Saunders and Jordan B. Pollack. The evolution of communication schemes over continuous channels. *From Animals to Animats*, 4:580–589, 1996.

[20] J. Schmidhuber. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2):173–187, 2006.

[21] Andrea Soltoggio and Ben Jones. Novelty of behaviour as a basis for the neuro-evolution of operant reward learning. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 169–176. ACM, 2009.

[22] Kenneth O. Stanley, Bobby D. Bryant, and Risto Miikkulainen. Real-time neuroevolution in the NERO video game. *IEEE Transactions on Evolutionary Computation Special Issue on Evolutionary Computation and Games*, 9(6):653–668, 2005.

[23] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10:99–127, 2002.

[24] Kenneth O. Stanley and Risto Miikkulainen. Competitive coevolution through evolutionary complexification. *Journal of Artificial Intelligence Research*, 21:63–100, 2004.

[25] Paul Tonelli and Jean-Baptiste Mouret. On the relationships between synaptic plasticity and generative systems. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 1531–1538. ACM, 2011.

[26] Gregory M. Werner and Micheal G. Dyer. Evolution of communication in artificial organisms. In *Proceedings of the Second International Conference of Artificial Life*, pages 659–687, 1991.

[27] Brian M. Yamauchi and Randall D. Beer. Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, 2(3):219–246, 1994.

[28] Tom Ziemke and Mikael Thieme. Neuromodulation of reactive sensorimotor mappings as a short-term memory mechanism in delayed response tasks. *Adaptive Behavior*, 10(3-4):185–199, 2002.