Testing for Machine Consciousness Using Insight Learning

Catherine Marcarelli and Jeffrey L. McKinstry

Point Loma Nazarene University
3900 Lomaland Drive San Diego, CA 92106
CMarcare@pointloma.edu JeffMcKinstry@pointloma.edu

Abstract

We explore the idea that conscious thought is the ability to mentally simulate the world in order to optimize behavior. A computer simulation of an autonomous agent was created in which the agent had to learn to explore its world and learn (using Bayesian Networks) that pushing a box over a square would lead to a reward. Afterward, the agent was placed in a novel situation, and had to plan ahead via "mental" simulation to solve the new problem. Only after learning the environmental contingencies was the agent able to solve the novel problem. In the animal learning literature this type of behavior is called insight learning, and provides possibly the best indirect evidence of consciousness in the absence of language. This work has implications for testing for consciousness in machines and animals.

Introduction

One possible definition of consciousness is the ability to simulate behavior mentally (Picton and Stuss 1994; Hesslow 2002). If consciousness involves mental simulation, then machine consciousness should be achievable, since it seems possible to build a conscious artifact which can learn from experience and use that experience to simulate its own actions and their sensory consequences in order to optimize behavior.

Although many believe that a conscious artifact (or machine consciousness) is possible, there is no consensus as to how one could prove that a machine was conscious (Edelman 1989; Franklin 2003; Aleksander 2003). Three possible tests for consciousness are through conscious report using language, neurobiological signals, and behavior (Edelman et al. 2005). The gold standard for detecting consciousness is conscious report through language (Edelman 1989); however, if the evolution of machine consciousness is similar to that of animal consciousness, then language is likely to come after the first conscious artifact. The second method for detecting consciousness by looking for neurobiological signals (e.g. gamma-band oscillations) is useful for machine consciousness only to the extent that the model is biologically inspired. Testing for consciousness using behavior, therefore, is of considerable interest to researchers seeking to build a conscious artifact.

One behavioral test that could be taken as evidence for consciousness is "insight learning". Insight learning is the ability to evaluate and combine prior experiences to solve new problems, for example Kohler's famous monkeys and bananas problem in which bananas suspended from a ceiling are reachable only by moving a box under the bananas and then climbing on top of the box. How to move boxes and how to climb on boxes were known from prior experience, but putting the two actions together in a sequence to obtain hanging bananas was novel. Insight learning has been reported in primates (Kohler 1959) and ravens (Heinrich 1995), and could be accomplished by something akin to mental simulation of actions and their sensory consequences, a process equated with conscious thought (Hesslow 2002). The fact that animals sometimes take a significant amount of time to find the solution in the absence of overt trial and error suggests the possibility that the animals are using mental trial and error. If consciousness is related to mental simulation, then insight learning is a reasonable test, since 1) insight learning tasks seem to require mental simulation in order solve, 2) introspection indicates that insight learning tasks can be solved by mental simulation in humans, and 3) insight learning in animals looks from the outside as if it involves mental simulation as well. However, the argument is not fully convincing, since it is difficult to rule out other ways in which insight learning behavior may be accomplished without mental simulation. In addition, one cannot guarantee that the machine has conscious experience, even if it does use a model of "mental simulation".

In order explore the possibility of testing for machine consciousness using insight learning, we created a computer simulation of an autonomous agent which learned to successfully solve an insight learning task using a technique which models "mental" simulation at a high level. Thus the model satisfies one definition of consciousness, and performs well on a task that seems to require flexible mental simulation in order to solve. Such complex, flexible behavior is thought to provide indirect evidence of consciousness in animals which cannot give a conscious report (Edelman, Baars, and Seth 2005). However, we will argue that the model is unconscious and discuss the implications for future studies of animal and machine consciousness.

Methods

Agent environment and task

The interaction between the simulated agent and its environment can be described as follows. The agent is placed in a 6x6 arena surrounded by black walls. The initial configuration of the environment and agent is illustrated in figure 1. The environment contains two blue boxes which the agent can push around. The floor of each grid location is cyan, except for three orange goal locations in the environment, represented by unfilled squares in figure 1. The agent is rewarded for pushing any box into any of the goal locations. A trial is terminated immediately upon goal attainment or if the maximum number of moves is reached. At the end of the trial the environment is reset to the initial configuration for the start of the next trial, however the agent remains in its current location. (If the agent occupies the initial location of one of the two boxes, then both boxes are placed in the alternative locations adjacent to the goals shown in figure 1.)

This task is meant to be a simplified version of the monkey and bananas problem that still satisfies the definition of insight learning. The agent learns about pushing boxes around by getting behind them and moving. The agent also learns that pushing a box into a goal area gets the reward. In the novel situation, the box can be up to 4 moves away from the goal, and the goal is in a new location (the goal in the center of arena in figure 1). (It is impossible for the agent to have ever pushed the "boxes" away from the walls during training, since it can't get behind them.) Thus, the agent must apply prior learned knowledge of the effects of its movements and the effects of the box moving to the goal in order to obtain a reward in a new situation; we argue that this is an insight learning task sufficient to demonstrate flexible "mental" simulation. Although the agent has already pushed a "box" one step in order to obtain the "bananas" before, it will still have to discover the exact sequence of moves to receive the reward at a new goal location, and determine that the sequence will work using only "mental simulation".

Motor commands

The agent can select one of three actions: move forward, turn left, and turn right. If the agent moves into the square containing a box, then the box will move to the adjacent square in the direction of the agent's movement. (If the agent would push the box into a wall, the movement is not allowed). If the agent attempts to move into a wall, the agent stays in the same location.

Sensory input

The agent has several simple "transducers" for sensing its environment at each simulation cycle. Sensors include a "camera", an internal "value" sensor, and "touch" sensor. The camera can detect the color at the current location, and the adjacent location in the direction the agent faces. Given the unique colors of the walls, boxes, goal locations, and the non-goal locations, the agent can use this sensory input to distinguish between them. The internal value sensor can distinguish between pain (due to walking into a wall or pushing a box into a wall), hard work (pushing a box), light work (movement without pushing a box), and reward (received from the external environment when the agent receives a reward). Finally, the touch sensor can distinguish between no force, the force caused by pushing a box, and the forced caused by hitting a wall or pushing a box into a wall.

Exploration during training

The insight learning task has two phases. It begins with a learning phase where the agent initially has no knowledge of the environment and cannot predict the consequences of its environment. It learns about its surroundings as it moves around and collects sensory data from the environment. By the end of the learning phase, the agent has an accurate internal model of the environment in the form of an egocentric map and a Bayesian network that predicts the sensory consequences of the agent's motor actions. In the testing phase, the agent the agent is able to utilize its learned model to simulate its environment without acting by looking ahead a certain number of moves decide upon an action sequence that will lead to a positive reward.

During exploration, the agent issues random actions and wanders around its environment enough to learn the layout of the environment and the relationships between variables in the environment. As the agent explores, the sensory data returned from the environment is stored in the agent in the form of an egocentric map. The map is initially empty. The map turns and shifts appropriately as the agent turns and moves in its environment like a GPS controlled map in an automobile. The map expands as necessary as the agent explores new areas. The map stores the data from the camera during exploitation. Using this internal map the agent will be able to obtain the following additional values from its memory after sufficient exploration: the color 2 squares ahead, the presence of an object 1 square ahead, and the color underneath a box. (Whenever the agent has discovered the floor color underneath a box, the map contains the color of the floor rather than the color of the box). A special value is used to indicate that the color is unknown. Object-1-ahead indicates whether an object is directly in front of the agent.

For simplicity, the DBN is trained in batch mode; once the agent has executed a given number of actions, the sensory-motor data collected during exploration is used to create a Dynamic Bayesian Network (DBN). Kevin Murphy's open source Bayes Net Toolbox (http://bnt/sourceforge.net/) for Matlab was used to dis-

cover both the DBN structure and the network parameters given the data. The following 7 variables at time t, and $t+1$ were used as input for a total of 14 variables: Current Reward, Current Force, Current Color, Color-1-ahead, Color-2-ahead, Selected action, and Object-1-ahead. Structure learning determines the causal relationships between these variables in time step t and $t+1$. The technique used by the toolbox maximizes the mutual information between variables at time t and variables at time $t+1$ (see the toolbox documentation). Given these causal relationships, the probability distribution of each variable given its parents was calculated using the toolbox function which finds the maximum likelihood estimate given complete data.

Exploitation during testing

Once the agent has learned a model of its environment, it can then use the model for exploitation. This is accomplished by "mental" simulation without receiving actual sensory input from the environment, but only using the predicted sensory input obtained from its internal model, via inference in the learned DBN and recall from the map, in response to a series of proposed actions. If a sequence of proposed actions leads to a reward, the actions are carried out to obtain that reward.

"Mental" simulation

In order to model the agent's mental simulation during exploitation, a depth-first tree search of possible moves was made starting with the current environment inputs. The number of moves that the agent could explore in advance was an experimental parameter. As each move was proposed, the new sensory inputs were predicted using inference in a DBN using the Matlab Bayesian Toolbox. In these predictions, the object-1-ahead and current, 1 ahead, and 2 ahead colors are all predicted using the agent's egocentric maps. The values for reward and force, however, are predicted through inference with the DBN using the J-Tree inference algorithm. Finally, the tree search selects the action sequence of a given length that produces the maximum cumulative reward as a result of this mental simulation.

Intelligent Agent Acting in the Environment

For the last two experiments the agent behaves as follows during its testing, or exploitation phase. Right before the agent is about to issue an action it mentally looks ahead 3 steps (as described above). If it predicts getting a positive reward within three actions, it will issue the first move in the action sequence that it predicts will lead to reward. If the agent does not predict that a positive reward can be obtained within three actions, it will issue a random action. For this part of the experiment the agent is given the same number of trials and moves per trial as it was during random exploration. Just like during exploration, the environment is reset if the maximum number of moves per

trial is reached or the positive goal is reached. Ultimately, acting intelligently in the environment should cause the agent to obtain an increased number of rewards since it knows how to get the positive reward and seeks out to do just that. It should be noted, however, that, due to the resetting of the boxes in the environment, the agent's perception of where the boxes are in the environment might be incorrect, initially. These faulty perceptions of the environment might cause the agent to think certain actions will lead to reward when they, in fact, do not due to the boxes being moved. Nevertheless, as the agent explores, it will choose to act on any sequence of length 3 or less that it things will lead to reward, increasing greatly the likelihood of obtaining rewards.

Simulation parameters:

The results below were achieved using the following parameters unless otherwise noted:

> Maximum moves per trial: 50
> Training trials: 600
> Testing trials: 600

Results

Dynamic Bayesian Network Structure

We adjusted the number of trials for exploration in order to adequately train the DBN. As the maximum number of movements (number of trials × number of moves per trial) during exploration was increased, the DBN structure became more and more consistent. When the maximum number of movements during exploration was set to 30,000 (600 trials with 50 actions per trial), the same DBN structure was consistently produced over multiple runs. The structure shown in Figure 2 closely reflects causal relationships in the environment.

Direct test of the utility of mental simulation

After training the agent, specific tests were run in which it was possible for the agent to obtain a reward in from 1 to 4 moves by pushing a box into a novel goal location. These test cases were set up by putting the boxes and the agent in their desired locations and forcing actions to get the agent facing the desired initial direction. At this point the agent knew where it had been moved as well as where the boxes are in the environment. The goal location for these tests was never used for obtaining a positive reward during the exploration phase (the goal in the center of the arena in figure 1) since the boxes were against the wall during training and therefore could not be pushed away from the wall. Three different starting scenarios each were created for the experiments one, two, and three moves away from goal attainment. A single start configuration was used for the 4 move test.

To give the agent enough data to correctly predict the variables in the environment 30,000 maximum moves were used (600 trial with 50 moves per trial – the same number used to create the consistent DBN structure).

To illustrate that the agent has learned to predict, act, and receive reward in the environment and that this ability was not hardwired into the agent, the test cases were run on the set of 5 trained agents and a set of 5 naive agents. The naive agents were allowed to explore for only 40 trials instead of 600, enough to learn the basic layout of the environment but not enough to fully learn the causal relationships between variables, whereas the trained agents were fully trained as described above. Agents were placed initially in test configurations and only allowed to make the minimum number of moves necessary to receive a reward. For example, if the configuration required two moves to obtain a reward, then the agent was only allowed to make two moves. A total of 50 tests were performed for each of the possible number of moves (one through four). The average percentage of possible rewards attained for each condition is plotted in Figure 3. The results confirm that the trained agents receive more rewards than the naive agents. While the trained agents all received 100% regardless of the number of steps necessary to obtain the goal, the naive agents received significantly fewer rewards in all cases ($p<0.01$ in all 4 comparisons, Wilcoxon Rank Sum), receiving an average of slightly below 40% for one step look ahead; performance approached zero percent as the number of moves from the goal increased. This confirms the necessity for exploration prior to "mental" simulation of the environment. The experiment also confirms that the agents are using prior experience to solve the new problem rather than innate abilities, since the naïve agents lacking in prior experience perform worse on the task; the task satisfies the requirement that insight learning applies prior experience to a novel situation.

The Utility of "Mental" simulation
Several additional tests of the utility of mental simulation in solving this insight learning problem were performed. When the trained agent uses mental simulation (three steps ahead) as it exploits its environment during testing, the amount of rewards it obtains significantly increases compared to acting randomly for the same number of steps. The median number of rewards received by the trained agents acting in the environment using mental simulation received more than twice as many rewards than control agents exploring randomly given the same amount of maximum moves (See Figure 4). The data was significantly different, as illustrated by the box-and-whisker plots in Figure 4.

Not only did the trained, "mental simulating" agents receive more rewards, but they also received the rewards in a shorter period of time. Figure 5 shows the median number of moves during random exploration and during exploitation using continuous mental simulation. Although both the random agents and the trained agents had the same maximum number of moves, a significantly lower number of total moves were made by the mental simulation agent, than agent that performs random actions. This is possible since a trial terminates when the agent receives a reward.

These two experiments illustrate the utility of a learned internal model of the environment coupled with "mental simulation" by showing that such an experienced agent can optimize its behavior and receive more rewards in less time than controls.
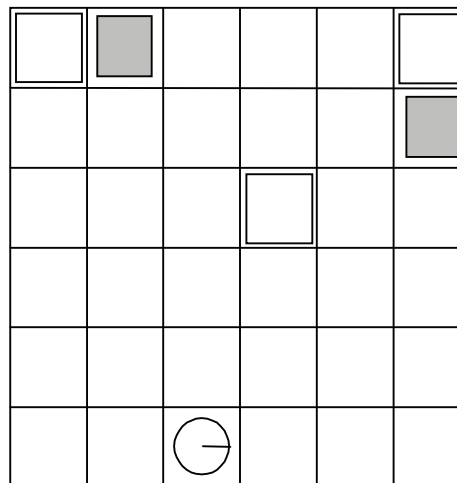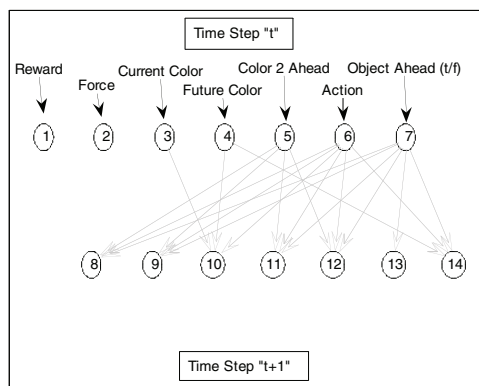

Figure 1. The initial environment setup.

Figure 2. The DBN learned after sufficient exploration reflects the causal structure of the environment.

Figure 3. The agent with experience significantly outperforms the naïve agent in the insight learning task (P<0.01 for all four pairs, Wilcoxon Rank Sum).

Figure 4. The trained agents using "mental simulation" receive significantly more rewards than random agents.

Figure 5. The trained agents using "mental simulation" receive significantly more rewards than random agents.

Conclusions

An insight learning task was used to test a model agent which could carry out simulations of its environment using an internal "model" that was learned through experience with the environment. After exploring its environment to learn the consequences of its actions, the agent successfully solved a novel insight learning task by simulating behavior through the learned model instead of acting. The agent successfully optimized its behavior by continuously searching for a sequence of actions that would obtain a reward before acting.

If consciousness is mental simulation, or if insight learning and flexible behavior is evidence for consciousness as some have argued, then this simulated agent could have been conscious to a limited degree. First, mental simulation has been conjectured to be the same as conscious thought. Hesslow (2002) argued that conscious thought can be carried out by 1) mental simulation of actions internally which activate the same motor systems as during overt action, 2) mental simulation of perception whereby brain areas that are activated by the senses are instead activated by the brain, and 3) anticipation whereby associations are formed between actions and their sensory consequences such that mental simulation of actions will result in the appropriate sensory consequences. Our simulated agent does all three in a highly reduced model "nervous system", with a Bayesian network learning to anticipate the sensory consequences of the agent's actions in conjunction with an ego-centric map of the environment. Secondly, flexible behavior such as insight learning has been argued as indirect evidence of animal consciousness and can be accomplished via mental simulation. Our model can solve an insight learning task by explicit, internal simulation of the consequences of its actions in its environment. Thus, by both criteria we could argue that our model has a limited degree of consciousness.

In addition, Aleksander (2003) has proposed a list of capabilities that a minimally conscious artifact should have: sense of place, imagination, directed attention, planning, decision making and emotion. All of these can be attributed to our model to a limited degree.

However, it seems quite unlikely that this simple simulation was conscious (although we cannot prove it). If not, why?

One possibility is that the mental simulation was too highly constrained. To really be convincing, one would have to have a conscious artifact that could learn and simulate a wide variety of tasks. Edelman (1989) suggests that one might be convinced that an animal is conscious if it performs sufficiently difficult tasks under a wide variety of situations, and suggests a task whereby
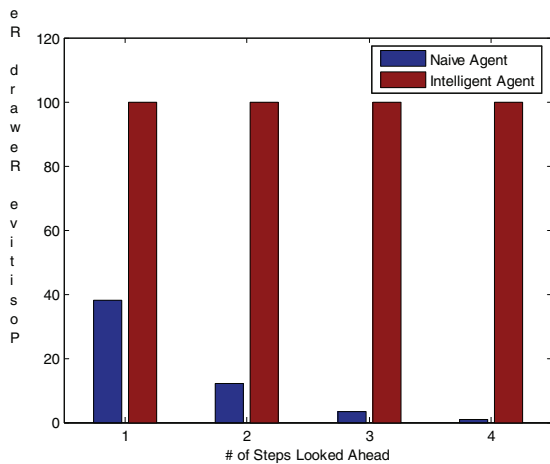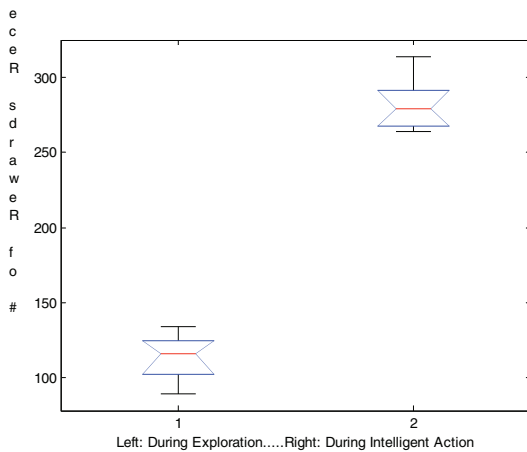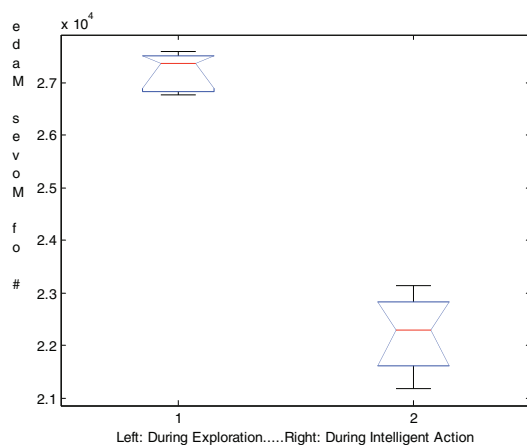
one monkey might learn to do whatever another monkey does, for example. Although our simulation learned its environment via exploration, the mental simulation used in our agent had a very limited repertoire of sensations and actions. **We propose therefore, that a conscious artifact should be capable of flexible mental simulation demonstrable by learning to solve a wide range of tasks with mental simulation.** Otherwise it is assumed that the system was carefully crafted to solve only the demonstrated problems; if consciousness has utility, then it is for flexible behavior (Koch 2004), therefore such flexibility should be demonstrated in a conscious artifact.

Another shortcoming of the model agent was the algorithmic nature of the mental simulation. Although tree search has been considered as part of a model for animal intelligence (Daw, Niv, and Dayan 2005), the uninterruptible, deterministic tree search used in this work does not seem to model human problem solving, which has been described by cognitive scientists as means-ends analysis, working both forward from the current situation and backward from the goal (Newell and Simon 1972). **We propose that a conscious artifact must be capable of monitoring its own problem solving progress under the control of its value system in order to be able to flexibly choose proposed actions for mental simulation** (See Edelman (1989) for more on the importance of the value system)**.** Our model lacked such self-monitoring capability, but slavishly followed a depth-first-search algorithm during mental simulation.

This work has implications for testing for animal consciousness as well. If insight learning alone did not prove that our agent was conscious, then it follows that it alone does not prove that an animal is conscious. Perhaps animals, like our automaton agent, are *unconsciously* following a search algorithm.

One still wonders whether such a machine that passes our proposed tests would be conscious. Edelman, Baars, and Seth (2005) have argued that the necessary conditions for primary consciousness in non-mammalian animal species are 1) brain structures equivalent to the thalamo-cortical structures in mammals, 2) neural activity similar to that in mammals during conscious states, and 3) rich discriminatory behavior demonstrating links between sensory perception and memory. If behavior alone is not enough to prove consciousness in clever, living, animals like birds and cephalopods, then the same is certainly true for a conscious artifact. Thus **the most convincing conscious artifact will be modeled after the mammalian brain.** (See for example, Holland's CRONOS project (Holland and Goodman, 2003)).

A machine that passes the three tests above may still be unconscious. Without a conscious report from our artifact, pushing towards more flexible behavior may be the best we can do to indirectly demonstrate consciousness, especially if the model is not a direct attempt to model the brain. Assuming that evolution did the same, perhaps we will converge on the same solution; if there are no zombies as some have argued (Edelman 1989; Koch 2004), then a conscious artifact that behaves as a conscious animal must necessarily be conscious.

References

Aleksander, I., and Dunmall, B. 2003. Axioms and tests for the presence of minimal consciousness in agents, J Consciousness studies, 10:7-18.

Daw, N. D., Niv, Y., Dayan, P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 8:1704-11.

Edelman G. 1989 *The remembered present: a biological theory of consciousness*. New York: Basic Books.

Edelman, D. B., Baars, B.J., Seth, A. K. 2005. Identifying hallmarks of consciousness in non-mammalian species. Conscious Cogn. 14:169-87.

Franklin, S. 2003. IDA: A conscious artifact? Journal of Consciousness Studies 10:47-66.

Heinrich, B. 1995. An experimental investigation of insight into Common Ravens, Corvus corax. Auk 112:994-1003.

Hesslow, G. 2002. Conscious thought as simulation of perception and behavior. Trends in Cognitive Science, 6:242-247.

Holland, O., Goodman, R. 2003. Robots with internal models. In O. Holland (Ed.), *Machine consciousness*. Exeter: Imprint Academic.

Koch, C. 2004. *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Co.

Kohler, W. 1959. *The Mentality of Apes* (2nd ed.), (E. Winter, translator). Vintage Books: New York.

Newell, A. and Simon, H.A. 1972. *Human Problem Solving*, Englewood Cliffs, NJ.: Prentice Hall.

Picton, T.W. and Stuss, D.T. 1994. Neurobiology of conscious experience. Curr Opin Neurobiol. 4:256-65.