Robots Solving Serial Means-Means-End Problems

Johannes Kulick   Stefan Otte   Marc Toussaint

Email: prename.surname@ipvs.uni-stuttgart.de

Abstract—Recently, Auersperg et al. [1] demonstrated that Goffin's cockatoos (*Cacatua goffini*) are able to solve complex means-means-end problems. This is an impressive cognitive ability and it is desirable to build models to understand such abilities. In this paper we describe a project that models such behavior and recreates the experiment on a robotic platform. First preliminary results suggest that the serial structure of the problem does not require much planning.

I. INTRODUCTION

Serial problem solving is a key capability of cognitive systems. In the recent research Auersperg et al. [1] showed that Goffin's cockatoos (*Cacatua goffini*) are able to solve means-means-end problems which require up to five dependent actions. More specifically, the cockatoos had to unlock a complex mechanism with different locks to reach their reward, a cashew nut.

Recent work of Kulick et al. [6] showed that learning the dependency structure of complex mechanisms from interaction is possible. In their work a robot uncovered the dependency structure of simpler mechanisms (e.g., keys locked a drawer) compared the cockatoo experiment. We believe that such a model is a good starting point to achieve behavior similar to the cockatoo's.

In this paper we will lay the groundwork towards recreating the cockatoo experiment with a robotic platform. Our plan is to confront our PR2 robot with a mechanism similar to the cockatoos'. We will show that the joint dependency structure model of [6] is well suited to capture many important aspects of the cockatoo experiment. In [6] no goal driven actions and no external reward were used to uncover the dependency structures. To capture these aspects we will integrate state-of-the-art planning mechanisms with the existing dependency model. Learning success as an intrinsic motivation and reward for reaching the goal will be used.

Given that this paper is mostly a project description, our results up to the point are still preliminary. However, first results in simulation look promising. They suggest that the serial nature of the problem encourages actions that lead to the correct solution without the need of planning.

II. RELATED WORK

The research is mainly inspired by the experiments on Goffin's cockatoos by Auersperg et al. [1]. They confronted several cockatoos with a box, locked by a complex mechanism as shown in Fig. 1a). To reach the reward, the animals had to unlock all five locks in a serial way. One cockatoo was able to

solve the whole apparatus without any pre-training on similar mechanism, whereas other subjects needed pre-training on the individual locks and others needed "coaching" by a cockatoo that was already able to open the box. Our goal is to realize the cockatoo experiment with a robot instead of the animals.

An important aspect of the experiment is uncovering the dependency structure of serial locks. We will use the approach described by Kulick et al. [6]. They model the dependency structure of the mechanism as probability distribution. Using actively gathered observations (the locking state, the position of the joints, and force-torque feedback) the posterior is calculated. The posterior then allows to make statement such as: "if the screw is between 3 and 15 mm out, it locks the bolt".

To integrate rewards (learning success and external reward) we use the *Monte-Carlo tree search (MCTS)* algorithm for planning. MCTS is the state-of-the-art planning method. For more details about MCTS have a look at Browne et al. [2]. We use planning in belief space as we plan in a partially observable domain (see Kaelbling et al. [4]).

As intrinsic motivation reward, we use measures from Bayesian experimental design [3] to reward action that lead to novel situations. In particular we use the expected cross entropy measure from Kulick et al. [5] to reward actions that challenge the current belief.

III. MODELING AND PLANNING

For modeling he dependency structure of the mechanisms we follow Kulick et al. [6] strictly. We thus have a discrete distribution for each mechanism, capturing which other joint locks and unlocks it.

In order to use intrinsic motivation as one reward, we need a notion of goodness of a state/configuration of the mechanism. We use a novelty measure for that. A good measure as shown by Kulick et al. [5] is the *cross entropy* between the current belief and the expected belief after an action is taken. We use this measure as reward in the planning process beside typical goal reward.

For planning we use MCTS [2] in belief space. It builds a tree of belief state nodes by simulating actions. As moving a mechanism to a position is a continuous action (since the desired position of the joint is continuous) the branching factor would be infinite. Even for grid discretization the space would be immense. Thus, we discretize the actions as follows. For each joint we can compute the action which most likely locks (and unlocks) the joint. We can also compute the action that maximizes the expected cross entropy over the belief before and after the action taken. This yields 3 actions for each joint
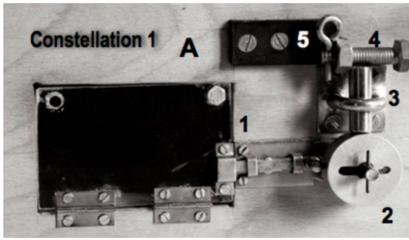
(a) The original mechanism the cockatoos had to open. (Picture from [1]).
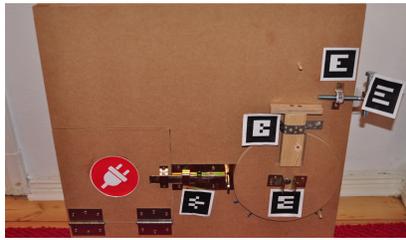
(b) The bigger version of the mechanism to fit the needs of a PR2 robot.
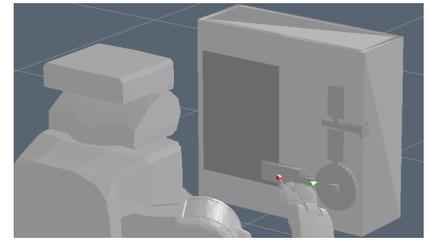
(c) The lockbox with the PR2 robot in simulation.

and $3 \cdot k$ total actions where k is the number of joints of the mechanism.

IV. EXPERIMENTAL SETUP

For this paper we implemented the lock box in simulation (see Fig. 1c). For the final experiment on the real robot this will be an enlarged version of the lock box to enable the PR2 to grasp all parts. A prototype of the real world mechanism is shown in Fig. 1b. Each mechanism must be slid or turned from its initial configuration to its other limit configuration to unlock the next mechanism in the chain. We assume that the robot has the motor skills to move each mechanism into the desired configuration. The robot can observe both the position of each mechanism as well as each locking state directly. The goal is to unlock the last mechanism to open the box.

V. PRELIMINARY RESULTS

We run the experiment in simulation with the following scenarios: (a) no reward at all (random actions), (b) only intrinsic motivation, (c) and intrinsic motivation and goal reward. With random actions the robot is not able unlock the last door within 30 action. With both only intrinsic motivation as well as the additional goal reward the robot was able to reach the goal state within 5 actions, i.e., it directly unlocks each mechanism in a row.

Note that these are early results and further investigation is needed.

VI. DISCUSSION

It is not surprising that random actions do not lead to the desired goal. However, this baseline strategy shows that the problem is non-trivial.

It is surprising that the robot reaches the goal quickly without any goal reward. This is due to the serial nature of the experiment. Intuitively it makes sense that moving a joint as far as possible has the highest chance of changing a state and thus is considered most interesting. In our setup of the experiment moving the joints to the opposite end always unlocks the next joint. In turn moving the next joint is more interesting as it has not been moved yet. Now the same argument as before applies and the joint is again moved as far as possible. With this sequence the mechanism is unlocked with only 5 actions.

This argument is also backed by the finding that the cockatoos in the original experiment had problems with the only mechanism (the wheel) that did not unlock the next

joint at the opposite end. The wheel had to be turned to a particular position to unlock the next joint. It was among the two mechanisms that needed the most time in Auersperg et al.s experiment.

VII. CONCLUSION

In this paper we laid the groundwork to repeat the cockatoo experiment by Auersperg et al. [1] with a robot instead of animals. First preliminary results show that the serial structure of the problem helps to find its solution. It would be interesting to confront the animals with mechanisms that do not have this serial structure (e.g., several mechanisms have to be in a certain configuration in parallel until the next mechanism is unlocked), or confront them with wheel-like mechanisms that are not unlocked by simply moving them to the opposite position.

The results are promising from a robotic perspective as they show how useful intrinsic motivation might be to solve planing problems. The comparability of the cockatoo and the robot is of course somewhat limited as we did not address the perception and motor skill problems. For the cockatoos one important aspect of the opening process was to learn, e.g., how to unscrew the screw. Adding motor skill learning and perception as additional problems would make the experiments more comparable to the original experiments and is an interesting direction for future work.

REFERENCES

[1] Alice MI Auersperg, Alex Kacelnik, and Auguste MP von Bayern. Explorative Learning and Functional Inferences on a Five-Step Means-Means-End Problem in Goffins Cockatoos (Cacatua goffini). *PloS one*, 8(7):e68979, 2013.

[2] Cameron Browne, Edward Powley, Daniel Whitehouse, Simon Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakus, and Simon Colton. A Survey of Monte Carlo Tree Search Methods. *Transactions on Computational Intelligence and AI in Games*, 4(1):1–49, 2012.

[3] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[4] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1):99–134, 1998.

[5] Johannes Kulick, Robert Lieck, and Marc Toussaint. Active learning of hyperparameters: An expected cross entropy criterion for active model selection. *arXiv preprint arXiv:1409.7552*, 2014.

[6] Johannes Kulick, Stefan Otte, and Marc Toussaint. Active Exploration of Joint Dependency Structures. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2015.