12

Intelligent Object Exploration

Robert Gaschler[1], Dov Katz[2], Martin Grund[1],
Peter A. Frensch[1] and Oliver Brock[2]
[1]Humboldt-Universität zu Berlin,
[2]Technische Universität Berlin
Germany

1. Introduction

Tool use is considered to be the hallmark of higher cognitive abilities (compare e.g. Blaisdell, 2008). It is therefore the target of an extensive body of work in psychology. The mechanisms that enable the discovery of affordances in humans and animals are still not fully understood. Tool use has been observed predominantly in primates but also in other animals such as crows. Weir and Kacelnik (2007), for instance, report on a New Caledonian Crow modifying aluminum strips in order to retrieve food. The crow correctly chooses between bending and unbending aluminum strips, depending on the specific type of jar it is presented with. Studies on tool use suggest that the potential application of objects to achieve manipulation objectives can be discovered through exploration. When an affordance of an object is discovered, it becomes a tool.

Recently, tool use has begun to gain the attention of a different research field: robotics. The goal of research in robotics is to produce artificial agents capable of accomplishing manipulation tasks. Many of these manipulation tasks require the usage of tools. Thus, significant progress in robotics will be achieved by developing the necessary mechanisms for tool use.

We believe that much can be gained by integrating the methods of psychology and robotics. The mutual interest in tool use creates an opportunity for fruitful collaboration. On the one hand, roboticists can leverage insights gained by decades of research on tool use in humans and animals. On the other hand, psychologists can benefit from quantifiable and easy-to-reproduce experiments conducted on artificial agents. A theory about tool use, for example, could be tested on a robotic system, allowing for precisely controlled experimental conditions, without the complexity involved in testing human subjects.

In this chapter, we present a new collaborative effort between researchers from the areas of psychology and robotics. We focus on exploration of kinematic structures as a first step towards advancing our understanding of human and robotic tool use. Through the study of kinematic structures, we hope to gain insights into the principles that govern tool use at a more general level. The work presented here is still in its early stages. Nevertheless, our preliminary results are encouraging.

To assess whether a new object can be a useful tool for a certain task, an agent must be able to explore the object's properties, such as its shape or the possible relative motions its constituent parts are able to perform. An understanding of object exploration is therefore a prerequisite for explaining tool use in humans and in robots. As in many environments exploration is costly, learning of efficient exploration strategies and transfer to novel objects is a major concern of the current chapter. To understand the principles underlying object exploration and the acquisition of exploration strategies, we turn our attention to a specific class of objects: rigid articulated objects. Articulated objects are objects composed of rigid parts that are connected to each other via degrees of freedom (joints). Exploration provides an agent with an understanding of how an object's degrees of freedom can be used and thus with knowledge about the potential use or the function of the object. Consequently, the ability to explore the kinematic structure of an object is a prerequisite for using it as a tool.

We describe a novel simulation environment, which will allow for robots and humans to interact with the same objects, in order to determine their kinematic structure. This simulated environment enables us to study object exploration in humans and robots. We hope that by studying exploration of articulated objects we can advance the current level of understanding of human object exploration and tool use, and that this improved understanding will play an important role in advancing robotic tool use. We also hope that this collaborative research will encourage a greater exchange of ideas and techniques between robotics and psychology.

In the following paragraphs we will discuss what psychology and robotics can contribute to intelligent object exploration in terms of representational formats, the dilemma of exploration vs. exploitation, and the interplay of passive knowledge accumulation and active testing. In each case we will briefly sketch how the robotics approach to object exploration can benefit from incorporating approaches from psychology and how psychological studies and applications concerning object exploration can be fostered by borrowing from and providing interfaces with robot object exploration. After this, we will describe in more detail our current approach and results concerning robot object exploration. Finally, we report findings on human exploration that render as highly promising future attempts to provide robots with human exploration strategies and to construct interfaces that allow for humans and robots to manipulate the same objects.

1.1 Symbolic and relational representations

The form of representation of object structure is highly relevant for intelligent object exploration. In order to benefit from past exploration episodes, experience needs to be stored in a way that allows application to novel cases and abstraction from irrelevant features. Both of the latter criteria pose a serious challenge to instance-based models that have been proposed in psychology and robotics for improving performance by storing processing episodes. We argue that symbolic and relational representations will be crucial to capture human and robot object exploration.

In cognitive psychology, instance-based approaches have been developed that successfully model human skill acquisition with narrow sets of stimuli (e.g., Logan, 1988). In these models, processing episodes (e.g. an arithmetic problem and its solution) are stored in memory. When the situation reoccurs, the solution can be retrieved from memory which is often faster and more efficient than the original way of computing the solution. While successful in explaining some aspects of human skill acquisition, these models are less helpful for capturing exploration. However, research on information reduction (e.g., Haider & Frensch, 2002; Gaschler & Frensch, 2007, 2009) suggests that humans spontaneously explore object structure and parse objects into relevant and irrelevant parts – even in repetitive tasks with a narrow set of material that would be solvable by an instance-based approach. They generate knowledge about abstract structural properties of the task material which allows them to process novel and unfamiliar objects just as efficiently as highly practiced ones.

Roboticists have also been using instance-based approaches to link specific states of the environment to actions. While machines are good at storing tables of states and actions, the problem of applying the knowledge to novel situations remains. Instance-based representations usually contain too much information due to the storage of irrelevant features. Unless repeated exposure to various exemplars of objects allows for pruning of the irrelevant features, they will hinder application of the knowledge to novel situations (e.g., Sun, Merrill, & Peterson, 2001 for an interesting approach). This is because novel situations that might in principle be suitable for application of the instance knowledge might have low similarity on irrelevant features. Other research in artificial intelligence and robotics has tried to tackle the problem of matching knowledge to states of the environment by considering more abstract representation formats. Katz and Brock (2008) proposed to capture information about robot object exploration episodes by symbolic relational representations. This work focuses on the domain of articulated objects. It leverages a representation of link properties that only includes properties that are relevant for manipulation. Application of the knowledge of the structural properties of the object therefore does not depend on irrelevant features such as the configuration of an object in every exploration episode. This type of approach avoids problems of combinatorial explosion when planning and learning from specific physical interactions that instantiate exploration episodes. Also, a robot system that has a clear-cut representation of an exploration event allows for an interface with a human. More specifically, a robot may learn from human exploratory behavior through observation.

In psychology and robotics representation can be regarded as an important component of intelligent object exploration and tool use. Representation format is crucial to solve the problem of the application to novel cases and abstraction from irrelevant features as well as the problem to provide an interface between human and robot exploration behavior. A high-dimensional world can be represented such that human and robot object exploration can rely on helpful restrictions of what has to be considered to solve a task. Through experiments, psychologists can reveal insights about the representations used by humans for solving exploration and tool use tasks. Roboticists can then test such representations in real-world or simulated experiments and try to design mechanisms that can choose or generate the most appropriate representation format for a new tool use task. Psychologists can gain from strict validation of their experimental results on a robotic platform. Human theories of exploration and tool use can be cast in the representation used for robot exploration and then tested in a simulation environment that allows for a large array of systematic variations of object structure.

1.2 Exploration vs. exploitation

Object exploration can lay the ground for tool use by delivering the knowledge about structural properties that is necessary to infer functional properties. If exploration is a means rather than an end, efficiency considerations are relevant. Humans or robots can either invest more time and energy in acquiring more knowledge about the structure of an object or rather capitalize on the (potentially incomplete) knowledge acquired so far and exploit it to boost performance.

Based on human and animal research, learning and decision theories have proposed models that balance exploration and exploitation by proposing that the probability to choose an option different from the one that is currently associated with the highest reward is inversely related to the strength of the evidence for the option deemed best (e.g., Luce, 1959). Through updating of the estimates choice preferences will change from trial to trial. Crucial for adaptation to dynamic reward structures, reinforcement learning will provide estimates that grant opportunities for further learning as eventually the option deemed less profitable is chosen occasionally. Indeed, humans often show probability matching. If human participants are instructed to choose between options (e.g. card decks) in reward discrimination experiments, probability matching has been frequently observed (e.g., Gaissmaier & Schooler, 2008). Their participants did not constantly choose the deck with the highest reward probability (i.e., optimizing) but rather chose all decks – with frequencies paralleling reward probabilities. Probability matching ensures quick adaptation to changes in the reward schedule as information on all options is constantly gathered. If, for instance, the previously less rewarding option would surpass the previously most rewarding option, a participant relying on optimization would miss the chance to profit from the new highest reward, while one engaging in probability matching would not. Notably, the authors furthermore showed that probability matching was not the result of probabilistic action selection in all cases. Rather, some of the participants constantly tested hypotheses about potential deterministic regular structures in the reward schedule of the task (even when the task was still probabilistic). Exploring various hypotheses subsequently led to choice patterns that were consistent with probability matching. Once a regular rather than a probabilistic structure was present in the material, many of the participants discovered it and switched to exploitation of the discovered regularity. Thus, human behavior can, at least on the aggregate level, be captured well by probabilistic approaches linking the strength of evidence to the balance between exploration and exploitation. This however does not preclude the development and usage of rule knowledge.

In a similar way, reinforcement learning approaches used in robotics inherently balance exploration and exploitation and are flexible enough to adapt to probabilistic as well as deterministic regularities in object structure. For instance, Katz and Brock (2008) have proposed a reinforcement learning approach to make robot object exploration more efficient. Robot actions that lead to the discovery of properties of the structure of an object are assigned a reward. The higher the evidence that a specific action will lead to a substantial increase in the amount of knowledge about the structure of an articulated object, the more likely it is to be executed. By this mechanism regular object structures will lead to strong manipulation knowledge which is exploited in order to boost exploration performance.

While reinforcement learning had originally been developed in psychology, the extension to temporal difference learning that is nowadays of high impact in model-based neuroscience and behavioral research has been sparked by work conducted from a machine learning perspective (e.g., Sutton & Barto, 1998). Investigating how humans balance exploration and exploitation will enable us to develop robots that are capable of exploring their environment in an efficient way. For instance, the conditions under which humans discard (seemingly) irrelevant object features from processing or, alternatively, start to rely on a feature first deemed irrelevant that then turned out to be highly correlated with outcomes and easy to assess might be of special interest for efficient object exploration and tool use in robots. Hypotheses about potentially efficient exploration strategies can be tested on simulated or physical artificial agents. Furthermore, heuristics for efficient exploration discovered or validated in robotic systems can be provided to humans. For instance, the visual saliency of object parts in a virtual environment can be dynamically adapted in order to guide attention of a human novel to a task based on comparing online eye tracking data and exploration knowledge of the robot.

1.3 Watching vs. doing

To explore a new object, we typically begin by poking it, rotating it, etc. This indicates that for humans action and perception are closely linked. Nevertheless, action and perception have often been studied separately. Similarly, robots are usually modeled as input-computing-output devices. Indeed, the most dominant paradigm in robotics is called sense-plan-act. This paradigm has led to a separation in the study or robot perception and robot motion.

In psychology, integrative approaches have been flourishing on different levels. On the one hand, this relates to the issue of action-enabled perception. On the other hand, this relates to the issue of how learning can lay the ground for developing agency and specifically for the ability to choose actions according to goals. Concerning the issue of action-enabled perception, animal research has early on pointed out that active movements are necessary in order to develop a functioning visual cortex (e.g., Held & Hein, 1963). While kitten actively moving in a controlled visual environment developed normal vision, yoked-control kittens being exposed to the very same movements passively did not. Recent findings by Craighero, Leo, Umiltà, and Simion (2011) suggest that based on a bias in perception, action and perception are linked in humans even before own action starts. They reported that two-day-old human newborns preferentially attend movements directed toward an object. The authors systematically manipulated (a) presence of an object, (b) direction of the arm movement, and (c) hand shaping, and found that the newborns oriented more frequently and looked longer at a hand shapes that were consistent with a movement goal. Apart from this, research and everyday experience suggests that moved or moving objects are detected faster as compared to items stationary to the background.

One of the first examples of robots interacting with the environment to simplify perception was proposed by Katz and Brock (2008; see also Katz, Pyuro, & Brock, 2008, and Katz, Orthey, & Brock, 2010). This work proposes that by pushing and pulling on an object, a robot can distinguish the object from the background and track which parts of the object are connected to each other by degrees of freedom. Consequently, when faced with a novel object, the robot can discover its kinematic structure by interacting with it.

The link between perception and action leads to an interesting perspective when considering how learning enables action. Perception and learning of co-variations between states in the environment and own movements can enable robots and humans to later exploit their knowledge about useful interactions to explore new objects. One line of research on human action control, often referred to as ideo-motor theory, proposes that we develop the capability to intentionally influence our environment by inverting observations of accidental moves and effects in the environment. For instance, Elsner and Hommel (2001) suggested that humans become intentional agents in a two-step procedure. Due to a lack of knowledge relating own movements to changes in the environment, a baby might at first have little basis for intentional action in the form of selecting motor programs that lead to desired goals. At first, a baby may in fact randomly execute motor programs and observe changes in objects. The co-occurrence of movements (i.e., shaking the leg) and changes in object (i.e., the mobile starts to move) are stored. Once established, associations between motor programs and changes in objects can be applied in reverse direction in the next step. For instance, a representation of the moving mobile might evoke the motor program that made it move accidentally in the first place. The baby becomes an active agent as it can select the motor programs leading to desired goals, because pleasurable states are linked to motor programs that anteceded them. By this it can test and establish causal knowledge linking motor programs to changes in objects. While later in life, we surely possess action-effect knowledge with respect to many objects and domains, we might use similar ways to obtain structured knowledge about novel objects. We develop the capability to intentionally use novel objects as tools by storing co-occurrences of movements and effects and by active exploration. To this end we can employ principled testing, but also learn from effects that were not brought about by active testing.

A central motivation of our work on combining the study of object exploration in humans and in robots is our belief that much can be gained by better integrating learning from watching and learning from doing. Humans and robots can explore objects successfully by capitalizing on our embodiment, as well as by taking advantage of opportunities for passive learning. While active exploration and systematic experimentation is the key strategy to test hypotheses about causal relations between own movements and changes in the object, the hypotheses to be tested might in part be derived from a rather passive subsystem that observes movements in the actor and the object. Research on implicit learning in humans (e.g., Frensch & Rünger, 2003) suggests that co-occurrence statistics about a multitude of features in the actor and the object might accumulate and the strongest of the co-variations may fuel active testing that can lead to causal knowledge and a symbolic level of representation. We thus propose that it may be beneficial to follow the human example, and develop robotic systems that combine active exploration with passive learning. Likely humans can, if novel to an object exploration or tool use task, profit from a robot model as a starting point for generating and testing hypotheses. Furthermore, human object exploration might benefit from robot systems that dynamically adjust training schedules in order to foster efficient hypothesis testing. For instance, the robot might track the hypothesis space explored by the human and enforce overlooked tests by blocking access to object parts, visual highlighting of object parts or online composition of transfer objects that pinpoint hypotheses about object structure.

1.4 From puzzle boxes to virtual environments

Since Thorndike's research on cats learning to escape from puzzle boxes and gaining access to food (Thorndike, 1911), various researchers have employed similar devices in order to study exploration. Approaches have to balance between (a) the goal of providing a rich environment to explore and (b) the consideration that data logging and quantification of behavior are necessary for most research purposes. If participants are provided with a rich environment and can explore it in a multitude of different ways, researchers will be faced with the task of categorizing and summarizing many instances of rather unique behavior. For instance, they will have to determine which motor patterns are functionally equivalent.

We are currently developing a virtual environment in which humans can explore and use kinematic structures through a haptic interface. This system allows to log the forces applied to the object and object movements. As discussed in the previous sections, it is often far from obvious whether or not a behavior allows to conclude that a system is behaving in a goal directed manner and through exploration is building up a knowledge base capturing relevant parts of the structure of the object – unless one has designed the system and has access to the process parameters. Braitenberg (1984) provided a vivid demonstration of how a few simple mechanical building blocks can, when combined, produce complex behavior in purely reactive creatures that nevertheless readily leads to attribution of agency by humans. Braitenberg suggested combining the analysis of biological systems with a synthetic approach in order to (a) guard against the pitfalls of attributing agency where there is none and to (b) generate fruitful research hypotheses from one approach to the other and vice versa. This is exactly the research agenda we are currently following to understand exploration of kinematic structures of objects. Robots and humans are faced with very similar tasks. In a virtual environment, humans can interact with kinematic objects over haptic devices transmitting force back and forth between hand and object. Simulated robots can explore the same objects in the same environment. Apart from offering the possibility to test robots and humans on the same task, the virtual environment has several other advantages. Data logging is precise and easy to automate. Different variants of fully specified and exactly reproducible objects can be created. Experimental manipulations of perceptual capabilities, regularities in the structure of the object, and manipulation capabilities are possible. Most importantly, depending on research goals, objects and manipulation capabilities can be designed in such a way that prior knowledge is either of key relevance or has little impact.

Our current approach to study how humans and robots explore the kinematic structure of objects and how exploration strategies change with experience is twofold. We are using high and low constrained environments in order to combine the strengths of both approaches. In each case we are using kinematic chains as objects that can be explored. The chains are equipped with different types of joints. The agent has to determine what type of joint is located at which part of the object in order to capture the structure and functionality of the chain. In the environment with few constraints the agent can apply continuous amounts of force to different parts of the object. By pushing or pulling, the agent creates the opportunity to track movements of the different parts of the chain. Based on observing how the parts of the chain move in relation to one another, the agent can (a) distinguish object from background, (b) infer which parts of the moving object(s) are linked rather than independent objects, and (c) infer which type of joint is located at which position.

2. Robot learning to explore objects by manipulating them through grounded relational reinforcement learning

Undirected object exploration can be time consuming. In many environments, an agent may thus be required to explore new objects efficiently. The ability to plan an exploration sequence by considering past experience with similar objects is therefore essential.

To decide how to explore an articulated object, a robot must determine a sequence of interactions with it. These interactions would result in relative motion between the parts of the object, enabling the robot to acquire knowledge of the object's shape and kinematic structure. In this section, we describe a simulation environment within which an agent can interact with novel objects. We will demonstrate that it is possible to gather and generalize manipulation expertise that will enable the robot to efficiently direct its future interactions within the environment. Our robot interacts with an object by pushing or pulling it, while observing the object's motion. As these interactions create a change in the configuration of the object, the robot incrementally discovers the object's intrinsic and extrinsic degrees of freedom (intrinsic = between the parts of a single object, extrinsic = between different objects). The robot learns to select interactions that are most likely to reveal the maximum information about the kinematic structure. The acquired manipulation knowledge substantially reduces the number of interactions required to obtain an accurate kinematic model. Furthermore, manipulation knowledge acquired by modeling one object transfers to other objects even if they have different kinematic structures.

In the approach introduced by Katz and Brock (2008), manipulation expertise is learned based on a relational state representation. This representation is essential, as it renders learning tractable by collapsing large regions of the state space onto a single, task-relevant, relational state. The symbolic representation is carefully grounded in the perceptual and interaction skills of the robot. This grounding ensures that relationally learned knowledge remains applicable in the physical world. We begin our discussion by introducing the relational representations of kinematic structures that forms the basis of our learning-based approach to manipulation. Next, we describe how this representation can be grounded using the perception and manipulation capabilities of the robot. We proceed to discuss the relational learning framework and how it can be grounded with respect to the relational representation. Finally, we demonstrate the effectiveness of our approach in manipulation experiments with articulated objects.

2.1 Relational representation of kinematic structure

To describe the state space associated with manipulating rigid articulated objects using a propositional representation, we would have to include a proposition for every object encountered by the agent. We would also have to include a proposition for every action applicable to this object. Gathering and generalizing manipulation expertise becomes impossible with this representation due to the combinatorial explosion of actions and states. A relational representation allows us to describe an infinite number of states and actions using a finite set of relations. It is thus critical to the success of our learning-based approach to manipulation. Our relational representation leverages the following insight: an agent may encounter, for example, many types of scissors. These scissors may vary in color, shape, and size. All scissors, however, have the same kinematic structure. This kinematic structure can be captured by a single relational formula.

What object properties should our relational representation include? To represent the kinematic structure of an object, we must consider joint types (revolute, prismatic, or disconnected), link properties (e.g. color and size), and the kinematic relationships between links. Therefore, our relational representation uses the following predicates:

1. Revolute Joint: R(·,·,…)
2. Prismatic Joint: P(·,·,…)
3. Disconnected: D(·,·,…)

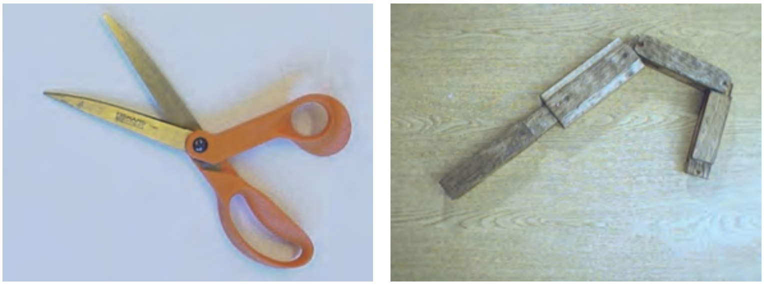


Fig. 1. Two examples of kinematic structures: scissors with a single revolute joint and a wooden toy with a prismatic joint and two revolute joints.

Fig. 1 shows two examples of kinematic structures. The scissors have a single revolute degree of freedom and the wooden toy is a serial kinematic chain with a prismatic joint (on the left of the figure) and two revolute joints. Our relational representation enables us to describe the kinematic structure of the scissors as: $D(L_B, R(L_1, L_2))$, where L_1 and L_2 represent the two links of the scissors and L_B is a disconnected background link. Similarly, the kinematic structure of the wooden toy can be represented as $D(L_B, R(L_4, R(L_3, P(L_1, L_2))))$. The notation is constructed based on a table that indicates the kinematic relationship between every pair of rigid bodies.

This representation is not unique. The wooden toy could also be represented as: $D(P(L_4, R(R(L_1, L_2), L_3)), L_B)$. The specific representation used by the agent depends on the order of discovery of the joints. The most deeply nested relation is the one discovered first. The representation of links can also be extended to an m-ary relation: L(·,·,…) where m>0 (m can be any positive integer implying that any finite number of properties can be captured by the representation). This representation supports a variety of link properties such as size, color, and composition. In the work we describe here, we limit ourselves to a single link property: size. The extension to a larger number of link properties, however, is straightforward. Using the extended link representation, the wooden toy can be described by: $D(L_B, R(L(S,F_4), R(L(S,F_3), P(L(S,F_1), L(S,F_2)))))$ where S stands for the property size=small and F_i spatially identifies link i in the physical world. To complete our relational representation, we must also provide a representation for the actions performed by the agent. The actions that we allow are limited to pushing or pulling a link. Each action can be applied either along the major axes of the link or at a forty-five degree angle to the major axes. An action is represented as A(L(·,·,…),α), where L(·,·,…) represents a link and α is an atom describing one of the six possible actions. The relational representation of links, joints, and actions allows us to reason and learn about objects based on their kinematic structure. The experience that an agent may acquire by manipulating scissors can be applied to all other scissors. The properties of an object that affect its manipulation behavior may not be limited to its kinematic structure. The relational representation of a link can be extended to include other relevant properties. With additional link properties, our agent will be able to distinguish between identical kinematic structures. The advantage of this approach is that it ignores information about the physical manifestation of objects (i.e. position, orientation, and configuration), as well as other properties irrelevant for generic description and control of manipulation. As a result, we achieve a significant reduction in the dimensionality of the state space, rendering the learning problem tractable.

2.2 Grounding the relational representation

The relational representation described in the previous section can only support the learning of manipulation knowledge if it is grounded in the physical capabilities of the robot. Grounding bridges between the symbols of our representation and the physical, continuous world (Harnad, 1990). It ensures that we can symbolically interpret the observations made by the robot with regards to its interactions with the world. At the same time, grounding ensures that the resulting symbolic manipulation knowledge maintains its relevance and predictive power for the robot's real-world interactions.

To ground our relational representation, we bind the relations R(·,·,…), P(·,·,…), and D(·,·,…) as well as the links' properties to real-world perceptual capabilities of the robot. These perceptual capabilities enable a robot to model rigid articulated objects (Katz & Brock, 2008). The robot's perceptual capabilities provide adequate grounding for our relational representation of links and their kinematic relationship.

2.3 Acquiring manipulation expertise

With the grounded relational representation of states (links and joints) and actions, we can now cast the problem of incremental acquisition of manipulation knowledge as a relational reinforcement learning problem (Džeroski, de Raedt, & Driessens, 2001; Tadepalli, Givan, & Driessens, 2004; van Otterlo, 2005). In reinforcement learning, an agent learns an optimal policy for solving a task. This policy tells the agent which action to perform in a particular state (e.g., where to affect the object and whether to push or to pull). The process of acquiring the policy is incremental; the agent learns the policy through a sequence of interactions with the environment. With every action, the robot may or may not discover new information. To formulate this process as a reinforcement learning problem, in our experiments, we simply assign a reward for every degree of freedom and every link property discovered by the robot. We expect the robot to incorporate new experiences into its policy, improving it over time. If learning succeeds, our robot will have acquired an effective policy for modeling the kinematic structure of novel rigid articulated objects. For more information about the implementation, we refer the reader to Katz, Pyuro, and Brock (2008).

Given unlimited time for exploration, an agent can gather enough manipulation experiences to learn an optimal policy. To comply with the time constraints imposed by manipulation in unstructured environments, our agent must be able to discover an optimal (or nearly optimal) policy quickly. To that end, it must balance between exploration and exploitation. Exploration refers to the execution of an action to improve the robot's estimate of the associated reward. In other words, when an agent explores, it either chooses an action it has never tried before, or the action the outcome of which the agent is most uncertain about. Exploitation, in contrast, refers to action selection based on maximizing reward. The balance between exploration and exploitation is important. If the agent explores too much, it will miss to employ knowledge. If it exploits too early, it will perform poorly because it has not gathered enough experience.

To decide if a new action should be executed, we compute the fraction of actions for which the robot already has gathered experience. It then selects one of the actions associated with its current knowledge about the object unless the random number generator indicates that exploration should be executed instead. If the robot is to retrieve an action based on its experience, we use Interval Estimation (IE) (Kaelbling, 1993), which picks the action that has the highest potential to perform well. Thus, IE also balances between exploration and exploitation.

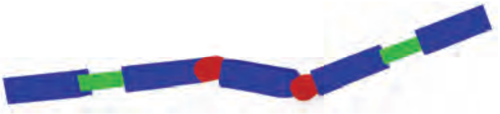


Fig. 2. Example of an articulated object. Links (rigid bodies) are shown in blue. Revolute joints are represented by red cylinders, and prismatic joints are illustrated as green boxes. Joint types are only marked for illustrative purposes but not in the experiments.

2.4 Experiments with planar objects

To evaluate the effectiveness of our learning-based approach to manipulation of articulated objects, we perform two types of experiments, previously published in Katz and Brock (2008). First, we show that manipulation knowledge can be gathered from experience. And second, we show that the acquired experience transfers to previously unseen objects. We perform experiments in a simulated environment. This environment is based on the Open Dynamics Engine (ODE), a popular dynamics simulator. ODE is an open source, high-performance library for simulating the dynamics of rigid bodies. It features various joint types and integrated collision detection. It simulates gravity, various sources of friction, and allows for some non-determinacy. In our experiments, a robot interacts with an articulated planar (two-dimensional) object to extract its kinematic structure. An example object is shown in Fig. 2. Links (rigid bodies) are shown in blue. Revolute joints are represented by red cylinders, and prismatic joints are illustrated as green boxes.

An experiment consists of a sequence of trials. A trial is composed of a number of steps. In each step, the robot applies a pushing or pulling action to the articulated object. The trial ends when an external observer (independent thread of the simulation) signals that the robot has obtained the correct kinematic structure of the object. In each step of every trial the robot accumulates manipulation experiences that improve its future performance. The number of steps per trial measures the number of interactions necessary to discover the correct kinematic structure of the articulated object. It therefore measures the efficiency with which the robot accomplishes the task. Each step of a trial can be divided into three phases:

1. The robot selects an action and a link for interaction. The action is instantiated using the current state and the experience stored in the agent's memory.
2. The selected action is applied to the link, and the resulting object motion is simulated. The observed motion is reported to the agent. If the agent pushes or pulls the object at a suitable location, the resulting object motion might deliver information concerning multiple links at the same time.

3. The agent analyzes the observed motion and determines the kinematic properties of the rigid bodies observed so far. These properties are then incorporated into the robot's current state representation.

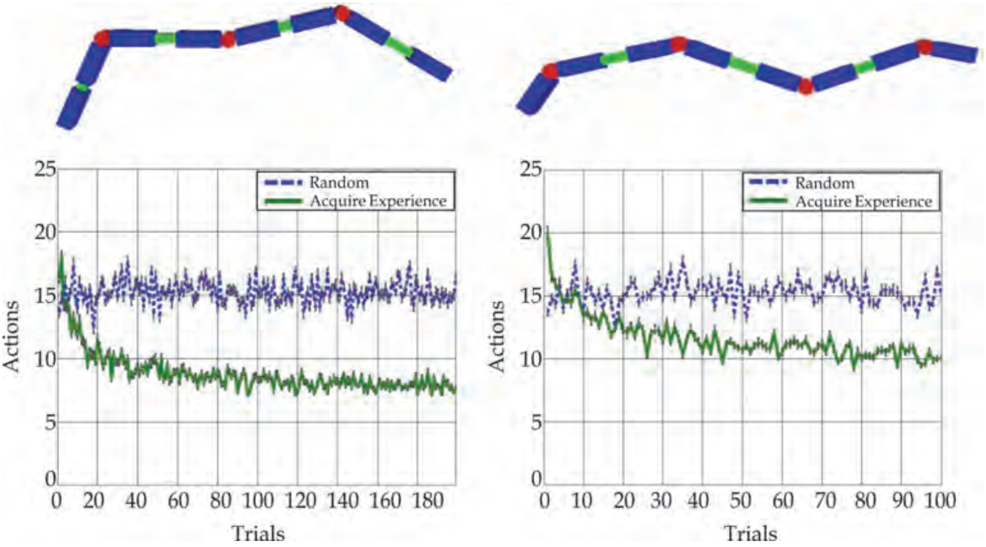


Fig. 3. left panel. Experiments with a planar kinematic structure (PRPRPRP). The object possesses seven degrees of freedom (R = revolute, P = prismatic). The right panel shows the experiment with the structure (RPRPRPR; seven degrees of freedom). Error bars in all figures reflect the standard error of the mean.

2.5 Gathering manipulation knowledge

Our first type of experiments shows that manipulation knowledge can be gathered from experience. To demonstrate the effectiveness of learning, we observe the practice-related decrease in the number of actions required to discover a kinematic structure. We compare the performance of the proposed grounded relational reinforcement learning approach to a random action selection strategy. Fig. 3 and 4 show the objects presented to the robot, as well as the results (learning curves) of four experiments. For each trial, we report the average number of interactions used to discover the correct kinematic structure. This average is computed over 10 independent replications.

In the first experiment, we presented the robot with an object with seven degrees of freedom and eight links. The resulting learning curve is shown in Fig. 3 (left panel). Action selection based on the proposed relational reinforcement learning approach results in a substantial reduction of the number of actions required to correctly identify the kinematic structure. As to be expected there is a stable and high number of actions required in the baseline, random action selection. This improvement already becomes apparent after about 10 trials. Using the learning-based strategy, an average of eight pushing actions is required to extract the correct kinematic model of the object. Compared to the approximately 16 pushing actions required with random action selection, learning achieves an improvement of about 50%. In the second experiment, we presented the robot with another object with seven degrees of freedom and eight links. The resulting learning curve is shown in the right panel of Fig. 3. The improvement achieved by our learning approach becomes apparent after about 20 trials. Using the learning-based strategy, an average of 10 pushing actions is required to extract the correct kinematic model of the object. Compared to the approximately 15 pushing actions required with random action selection, learning achieves an improvement of about 30%.

In the third experiment, we presented the robot with an object with eight degrees of freedom and nine links. The resulting learning curve is shown in Fig. 4 (left panel). The learning-based strategy requires an average of eight pushes at asymptote, whereas the random strategy uses approximately 20 pushing actions. Learning achieves an improvement of about 60%. In the fourth experiment, we present the robot with an object with nine degrees of freedom and ten links. The resulting learning curve is shown in the right panel of Fig. 4. The learning-based strategy requires an average of 10 pushes, whereas the random strategy uses approximately 22 pushing actions. Learning achieves an improvement of about 60%. These four experiments demonstrate that our approach to manipulation enables robots to gather manipulation knowledge and to apply this knowledge to improve manipulation performance.

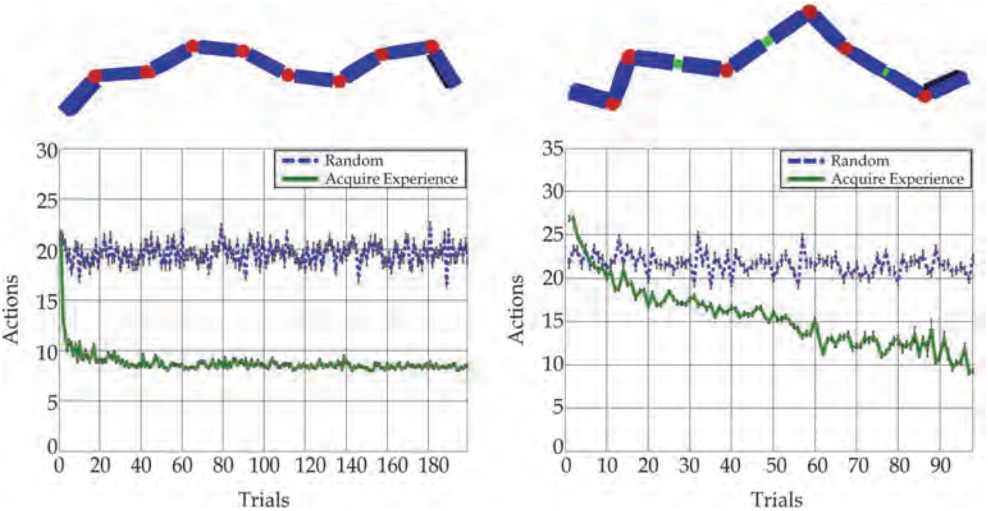


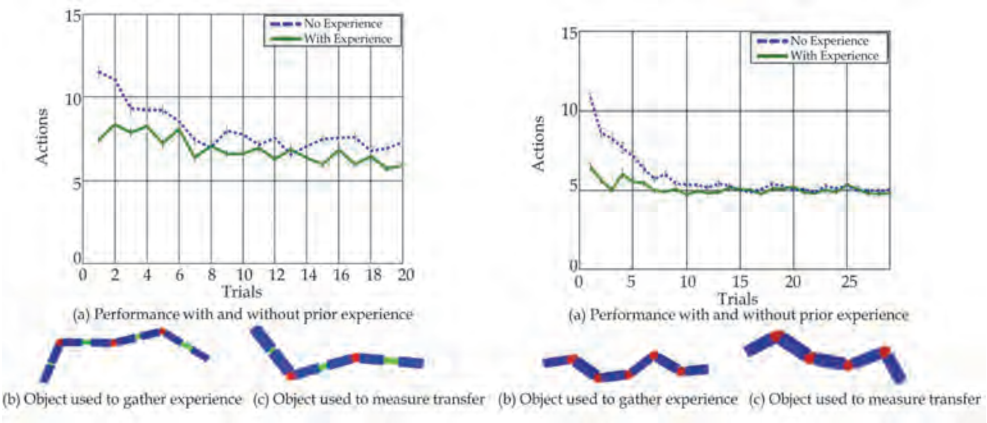
Fig. 4. Structure RRRRRRRR (eight degrees of freedom) plus learning curve and baseline on the left panel as well as structure RRPRPRRPR (nine degrees of freedom) on the right panel.

2.6 Transferring manipulation knowledge

Our second type of experiment shows that manipulation experience acquired with one object transfers to other objects. To demonstrate the effectiveness of knowledge transfer, we again observe the number of actions required to discover a kinematic structure. We compare the performance of the proposed grounded relational reinforcement learning approach with and without prior experience. Fig. 5 and 6 show the objects presented to the robot, as well as the results (learning curves) of four experiments.

In the first transfer experiment, the robot gathers experience with an articulated object with seven degrees of freedom (see Fig. 5, left panel). After 50 trials, the robot is given a simpler object with only five degrees of freedom. The simpler structure is a substructure of the more complex one. We compare the robot's performance with that of a robot without prior experience. The robot with prior experience consistently outperforms the robot without experience. In the first trial, which is the most important for real-world manipulation, the experienced robot requires only 40% as many pushes. Over the following five trials, the performance improvement is approximately 20%. In trials 5 to 20, the performance improvement is much smaller.

In the second transfer experiment, the robot learns to manipulate a complex articulated object with five revolute joints (see Fig. 5, right panel). After 50 trials, the robot is given a slightly simpler structure that only possesses four revolute joints. Again, the simpler structure is a substructure of the more complex one. We compared the robot's performance after these initial 50 trials to the performance of a robot without prior experience. The experienced robot achieves convergence almost immediately. This corresponds to a performance improvement of about 50% in the first trial, relative to the robot without experience. After about 15 trials, both robots converge to approximately the same performance. This is to be expected for simple structures, exclusively consisting of revolute joints. The third transfer experiment complements the second experiment. Here, the robot learns to manipulate an articulated object with four revolute degrees of freedom (see Fig. 6, left panel). After 50 trials, the robot is given a structure with an additional revolute joint (five altogether). We compare the robot's performance after these initial 50 trials to another robot's performance without prior experience. Again, experience results in an improved performance in the first few trials (about 30%). After about eight trials, both robots converge towards the same number of interactions.

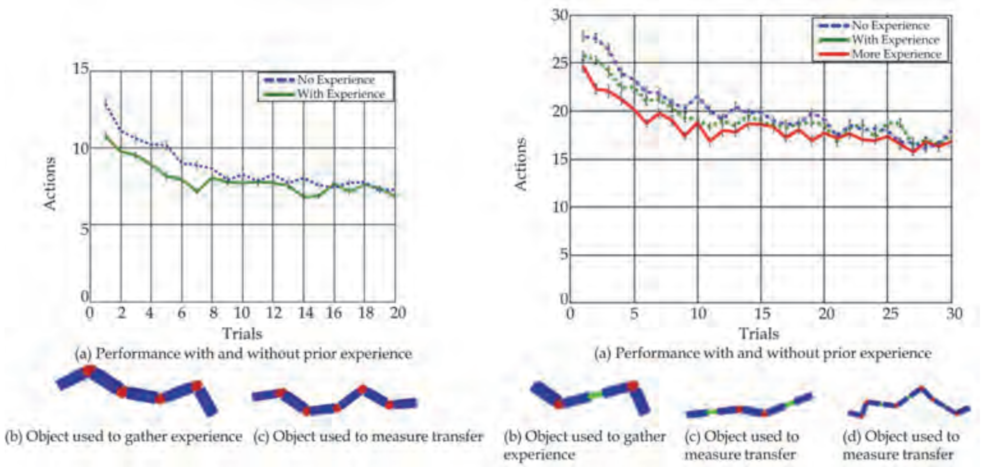


(a) Performance with and without prior experience

(a) Performance with and without prior experience

(b) Object used to gather experience (c) Object used to measure transfer (b) Object used to gather experience (c) Object used to measure transfer

Fig. 5. left panel. Experiment on transfer of knowledge acquired with PRPRPRP to PRPRP. The right panel shows the experiment on transfer from RRRRR to RRRR.



(a) Performance with and without prior experience

(b) Object used to gather experience (c) Object used to measure transfer

(a) Performance with and without prior experience

(b) Object used to gather experience (c) Object used to measure transfer (d) Object used to measure transfer

Fig. 6. Left panel. Experiment on transfer of knowledge acquired with RRRR to RRRRR. The right panel shows the transfer (RPR to PRRP and RRPRPRRPR).

The fourth experiment (see Fig. 6, right panel) takes a step towards long term learning of manipulation expertise. Here, we compare the performance of three robots: The first has no prior knowledge, the second's prior knowledge is based on interactions with one object, and the third's prior knowledge relies on interactions with two objects. The results show the advantage that the more experienced robots have in the first few trials. More importantly, this experiment suggests that the more experience a robot gathers, the more it can transfer to new situations.

To summarize, our experimental results provide strong evidence that learning from past experience can significantly boost the robot's manipulation performance. Learning enables a robot to autonomously acquire manipulation expertise by interacting with the environment. Our results show that this expertise transfers across different instances of the manipulation task and substantially improves manipulation performance. Learning and generalization of manipulation knowledge become possible due to our relational representation of states and actions. This representation collapses the otherwise intractable state space and renders reinforcement learning feasible. We believe that the effectiveness of our approach is due to the proper, task-specific grounding of our relational representation in the robot's perceptual and interactive capabilities.

2.7 Experiments with 3-D objects

We are currently working on the development of a new simulation environment for three-dimensional objects. This work is still in its early stages. Our primary objective is to replicate the success of learning for planar objects in the more general case of 3-D articulated objects. An example of the type of three-dimensional objects we plan to explore in the new simulation environment is shown in Fig. 7 (left panel). In this simulation environment, we also intend to explore the relevance of a variety of object properties, such as size, color, texture, the existence of parallel lines, or sharp changes in contrast.

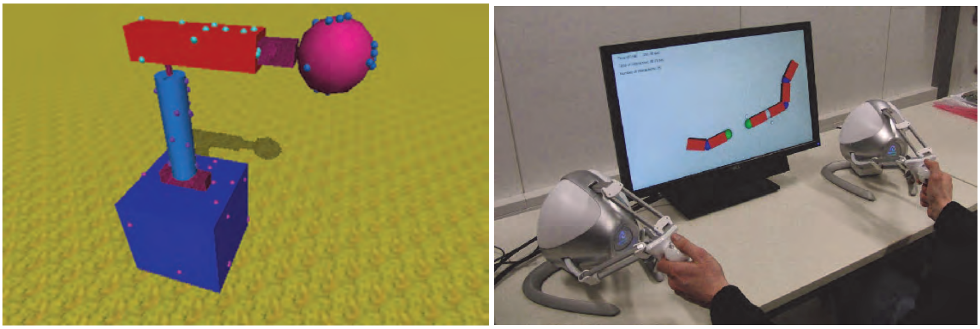


Fig. 7. Left panel. Simulated three-dimensional rigid articulated object. On the right side: two haptic devices operated by a human subject to interact with an object.

The new simulator is designed to facilitate research in the intersection between human and robotic object exploration. The simulator features a haptic interface (see Fig. 7, right panel), enabling human subjects and the simulated robots to interact with the same kinematic structures. Encouraged by the success of our learning approach for the domain of planar objects, we intend to use the new simulation environment to further develop our robot's skills in exploring new objects. We hope that by studying how human subjects approach object exploration, balance exploration and exploitation we will be able to extract knowledge that will advance the state of the art in autonomous manipulation.

3. Exploration of simple structures in humans

If, in the long run, robot exploration is to take advantage of adaptive strategies from human exploration behavior, one has to demonstrate in the first place that there is in fact an adaptive processing in humans while performing an exploration task that is suitable for robots. This is the goal of this section. Ideally, one could substantiate the notion that participants make use of clever exploration strategies, show systematic exploration and generate rules about the characteristics of the material. This could motivate research that employs these behavioral patterns for robot object exploration. To this end, we study here human exploration of simple structures with a simplified interface. We aim first to demonstrate principled exploration in a narrow environment before expanding to more complicated object structures and elaborate interfaces in the future. For instance, in the virtual 3-D environment described above, humans will use haptic interfaces to manipulate objects. Robots can (a) observe and try to learn from human moves and (b) manipulate the same objects. Notably, on the human side, psychomotor abilities and cognitive aspects of exploration will jointly determine performance. Failure can either be attributed to lack of knowledge about where and how to physically affect the object in order to learn the most about its structure, or can be attributed to difficulties in skillfully executing and completing manipulation plans. Likewise, for a researcher or a robot trying to extract valuable patterns of exploration behavior from human manipulation there is the problem of parsing behavior into discrete attempts to affect the object by applying force to a specific part of the object.

In order to provide a firm basis for our future attempts to tackle these problems, we first tried a divide and conquer strategy, setting apart the more cognitive aspects of exploration from the more psychomotor aspects. As detailed below, we started our work on human exploration of articulated objects with a highly simplified exploration environment excluding the need for skillful application of force and limiting the space of possible tests and strategies. With this, we wanted to provide evidence for systematic and adaptive exploration strategies in a variant in which the parsing of the exploration by a machine would be trivial. This should lay the ground for tackling more complicated object structures and less constrained continuous exploration behavior while making use of a haptic interface. As described in detail in the next section, in the high constrained environment participants were confronted with a short chain with space for three joints on each trial. Participants were asked to conduct discrete tests for each of the different potential types of joints in each of the locations of the chain in order to discover the structure of the chain.

3.1 Setup of the task

We designed an experiment to test whether and how systematic exploration of highly constrained structures occurs in humans. On each trial, participants were provided with a chain on the screen and were asked to test the different joints (compare Fig. 8).

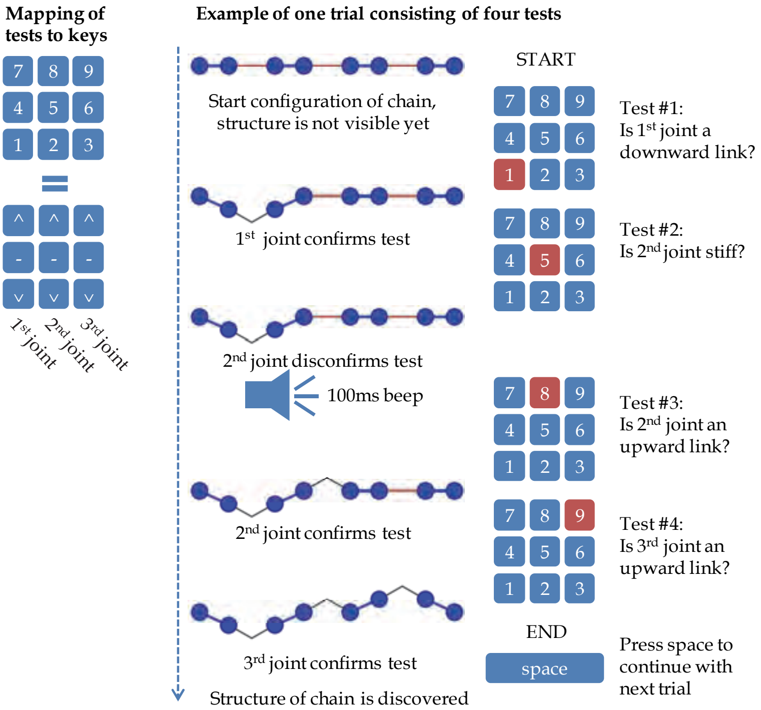

Mapping of tests to keys
7 8 9
4 5 6
1 2 3
=
^ ^ ^
- - -
v v v
1st joint
2nd joint
3rd joint
Example of one trial consisting of four tests
START
Start configuration of chain, structure is not visible yet
Test #1: Is 1st joint a downward link?
1st joint confirms test
Test #2: Is 2nd joint stiff?
2nd joint disconfirms test
100ms beep
Test #3: Is 2nd joint an upward link?
2nd joint confirms test
Test #4: Is 3rd joint an upward link?
3rd joint confirms test
END
space
Press space to continue with next trial
Structure of chain is discovered

Fig. 8. Setup and example trial of the highly constraint exploration task for humans.

Each chain had three joints. There were three different kinds of joints: bending upward, bending downward, and stiff connections. We used the number pad of a regular keyboard. The leftmost column of the 3 by 3 matrix of the number pad was assigned to testing whether the leftmost joint was bending upward (upper key), was stiff (middle key) or bending downward (lower key). The same arrangement was in place in the second and third column of the number pad with respect to the second and third joint (counted from the left). Tests were executed in a discrete manner. If the participant wanted to test whether the leftmost joint was able to bend upward, the participant pressed the upper left key on the 3 by 3 matrix of the number pad. Then, the display on the screen indicated if the joint indeed bended upward (if this was the characteristic of this joint) or if it was instead either a stiff joint or one able to be bent downwards. Then a tone sounded as feedback on the discrete test while the visual display remained constant.

3.2 Selection of training material

In order to test for systematic and adaptive exploration in humans we used material that normatively favored some strategies over others. We judged the systematism of human exploration behavior based on whether participants adapted to the structure of the material. On a finer level, we inquired whether humans either developed rule-like knowledge about the structure of the chains occurring in the training material or rather learned which exemplars of chains existed in the practice set. While we first describe the different regularities we built into the material for different groups of participants, we then discuss how it is possible to distinguish between an adaptation to these regularities that is based on rule knowledge vs. one based on exemplar knowledge.

We distinguished four conditions with different training materials. As we constructed chains with three joints each selected from three different types of joints, there were 27 different chains in principle. The training material was selected from the pool of 27 possible chains according to one of three different rules. Each of the rules led to the selection of twelve chains and allowed for clear predictions on how learning should change exploration. The first rule was tested on two different groups of participants. They explored chains in which *joints 2 and 3 were never stiff* (they were either bending upward or bending downward). If participants learned about the structure of the chains, they should stop testing whether the joints 2 and 3 are stiff. As detailed below, we varied the frequency of specific chains during training in 13 participants so that four of the twelve chains were repeated four times per learning block and others just once. For nine participants the same chains were presented with balanced frequencies – each twice per block of 24 trials. The third group of participants (N=9) was provided with chains in which *no neighboring joints were identical*. If participants adapted to the structure of the material, they should often switch to a different type of test when switching to test another joint. The fourth group of participants (N=9) explored the structure of chains in which *two neighboring joints were identical*, either the joints 1 and 2 or 2 and 3. We hypothesized that participants would adapt to the material by often executing identical tests on neighboring joints, especially once one joint had been correctly classified.

3.3 Frequency variation to test for rule knowledge

In the following we consider the condition in which *joints 2 and 3 were never stiff* in some more detail for two reasons. First, this training material allows for a strategy change towards faster exploration by discarding irrelevant aspects of the exploration task (refraining from stiffness tests on joints 2 and 3). Research on information reduction (i.e., Gaschler & Frensch, 2007, 2009) has argued that the discarding of irrelevant aspects of tasks from processing is a major basis of skill acquisition and of expertise acquisition. One can argue that by learning which aspects are relevant and which can be ignored, experts learn to use their time and cognitive resources very efficiently (in their domain of expertise). Similarly, learning to avoid less useful tests on kinematic objects helps to focus on hypotheses concerning their structure, to save time and energy, and to reduce risks that might be involved in executing tests in adverse environments. Second, in the research on information reduction we have proposed means to test whether simplification of task processing is based on rule-like knowledge and voluntary strategy change. We therefore wanted to apply such a test first on the data of the group with the setup most similar to the one used in research on information reduction so far. A test of rule-based performance is very useful for our goal to demonstrate systematic, principled exploration behavior in humans.

In research on information reduction we have been arguing that observations of people simplifying task processing, for instance by ignoring irrelevant aspects of stimuli, are widespread in various domains of applied psychology. However, special manipulations are necessary in order to test exactly how the simplification of task processing takes place and what kind of knowledge about regularities in the task material is acquired. For instance, the widespread observation *that* after some practice, participants ignore aspects of stimuli that are less relevant does not suffice to judge whether rule-like knowledge has developed or whether participants have instead adapted to the specific training exemplars. We successfully applied manipulations of exemplar frequency to specify the type of knowledge being acquired during practice and the mode of exploitation that this knowledge leads to. In particular, we varied the frequency with which specific training exemplars were processed during practice. If knowledge about the structure of the material would be bound to the specific instances encountered during training, then one would expect that learning should occur early in training for the frequently encountered exemplars, but much later for the examples presented only infrequently. Already early in training, participants could accumulate substantial experience with frequently presented exemplars and, for instance, start to ignore irrelevant parts in these exemplars, while still fully processing the infrequent exemplars until a similar amount of experience with these has been gathered. If, however, participants generate rule-like knowledge, then practice should modify the processing of the frequently and less frequently encountered instances at the same time and to the same extent. The latter is what we observed in the studies on information reduction. Participants learned to ignore the irrelevant parts of infrequently presented exemplars at the same point in time during practice and managed to ignore these to the same extent as the frequently encountered exemplars. It was not the case that participants dared to ignore the irrelevant parts of well-known items while still fully processing infrequently presented and novel exemplars. Rather, there was an all-or-none strategy change.

Here we employed a similar approach in order to judge whether or not participants developed rule knowledge when confronted with material in which the joints 2 and 3 were never stiff. Counterbalanced across participants, either the four chains with the first joint bending upward or the four chains with the first joint bending downward were repeated four times rather than once per block. The frequency manipulation allowed to distinguish between gains in exploration efficiency based on rules knowledge vs. on representations of specific exemplars of chains. If, on the one hand, participants rely on knowledge about specific exemplars, then the rate of testing whether joint 2 and 3 are stiff should decrease much more quickly per block of practice for the four frequent in comparison with the infrequently presented chains. People would e.g. learn that for the four frequently presented chains starting with an upward bending joint 1 there is no need to test whether joints 2 and 3 are stiff, but learn little about the other eight chains confirming to the rule that were presented less frequently. If, on the other hand, participants acquire knowledge that can be described as a rule, then the frequency of training exemplars should be irrelevant. The rate of testing whether joints 2 and 3 are stiff should decrease at the same rate per block and to the same level for both frequently and infrequently presented chains.

3.4 Procedure

Participants were instructed that their task was to explore chains by determining in each trial the types of the joints. Participants were provided with the mapping of keys and tests for the three different types of joints. They then performed four blocks of 24 trials on the training material selected as detailed above. In block 5, participants from all conditions were exposed to all 27 possible chains. We randomly sorted the chains in each of the five blocks for each participant . From the perspective of the participants, there was no signal for the beginning or end of a block.

3.5 Results

3.5.1 Overall learning

As an initial learning check, we analyzed whether practice led to a decrease in the number of tests required to determine the structure of a chain. This analysis confirmed that participants learned to explore chains more efficiently from block to block (compare Fig. 9, left panel). A mixed analysis of variance with training block as factor varied within participants and composition of training material varied between participants confirmed the general training effect as there was a significant main effect of training block, *F*(2.24, 80.52) = 20.1, *MSE* = 253, *p* < .001, η_p^2 = .358. We applied Greenhouse-Geisser correction here and whenever warranted in the analyses of variance (ANOVAs). The average amount of tests per trial was similar in each of the four groups and decreased at the same rate over blocks. The ANOVA neither showed a main effect of the composition of training material (*F* = 1.07) nor an interaction of training material and training block (*F* < 1).

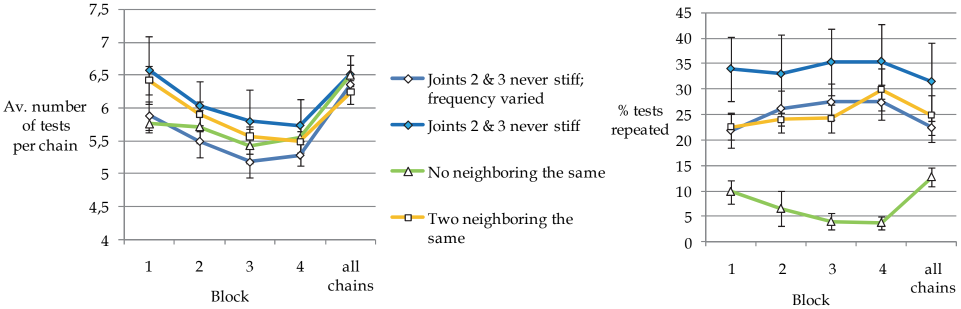


Fig. 9. Practice-related changes in human exploration performance. The left panel shows the decrease of the average number of tests executed per trial to determine the structure of the chain. The right panel depicts the average % of test repetitions that were observed when participants changed from testing one joint to testing a different one.

The decrease in the number of tests per chain in blocks 1 to 4 could either be the result of learning about the structure of the chains or of other practice effects (e.g., learning to operate the keyboard to execute the tests or learning to avoid test repetitions). Therefore, transfer to a situation in which the pool of chains changed while the exploration task stayed constant was essential. Comparison of the last block of training with the subset of the material and the final block with all possible 27 chains suggests that participants indeed learned of the structure of the chains presented in blocks 1 to 4. There was a sharp increase in the average number of tests executed per chain between block 4 and the final block, bringing performance back to the starting level. This rules out that the decrease in the number of tests executed per chain was due to general training effects unrelated to the chains presented. An ANOVA of the last two blocks confirmed the visual impression. There was a main effect of block, $F(1, 36) = 89.46$, $MSE = .125$, $p < .001$, $\eta_p^2 = .713$. Again the specific rule applied to select the training set neither influenced the average amount of tests per chain in main effect nor in interaction with block (*F*s < 1).

3.5.2 Practice related changes in tests on neighboring joints

While the above analyses suggest that participants learned about the chains they encountered during training, it does not specify what exactly was learned. In the next two sections we therefore analyzed the groups of participants separately according to whether and how they adapted to the specific regularity present in their training material.

First we analyzed the average rate of trials in which one type of test was repeated on subsequent tests on different joints of the same chain. Participants confronted with chains selected from the pool of 27 possible chains under the constraint that no neighboring joints were identical should refrain from repeatedly executing the same test. After having tested, for instance, whether joint 1 bends upward, they should not execute the same test on joint 2 but rather check for the ability of joint 2 to bend downwards. The right panel of Fig. 9 suggests that this was indeed the case. Participants adapted to the regularity during the first few trials. The proportion of subsequent identical tests on different joints was already very low in this condition as compared to the other conditions in the first block and decreased further over the three training blocks. The reverse should hold true for participants trained on chains selected so that 2 neighboring joints were identical. They should execute the same tests on neighboring joints. Unexpectedly however, no marked boost of the rate of executing identical tests subsequently on different joints was evident. Differences in overall rate of repeating tests subsequently on different joints amongst the conditions as well as differences in the dynamics were confirmed by an ANOVA on the data of training blocks 1 to 4. There was a main effect of the composition of the training material, $F(3, 36) = 18.98$, $MSE = 774.15$, $p < .001$, $\eta_p^2 = .613$, as well as an interaction of composition condition and training block, $F(7.49, 89.86) = 2.75$, $MSE = 103.17$, $p = .011$, $\eta_p^2 = .187$.

3.5.3 Practice related changes in tests as joints 2 and 3 were never stiff

Testing for systematic exploration in humans, we next analyzed the data of the participants trained on chains in which joints 2 and 3 were never stiff. We focused on how the average number of tests for whether joints 2 or 3 were stiff decreased with practice. As detailed above, in and of itself a practice-related decrease in the average number of tests is compatible with many views on what exactly is being learned. For testing whether exploration leads to rule-like knowledge and systematic exploration, the variation of the frequency with which specific chains were presented per block has to be taken into account.

Consistent with the view that exploration is systematic and related to rule knowledge, we found that participants learned as quickly to perform efficient exploration on *infrequently presented* chains as they did on frequently presented chains. General knowledge about the characteristics of the chains rather than knowledge about specific chains that were frequently presented was driving performance. Fig. 10 shows average frequencies of testing joints 2 and 3 in the group with the frequency variation (lines named high frequency vs. low frequency) and in the group of participants in which all chains were presented twice per block of practice (equal frequency line). In order to investigate the impact of presentation frequency on the number of tests for stiff joints 2 and 3, we charted the same data in two different ways. On the left panel we averaged the data per block (which we consider first), while on the right panel we averaged the data based on counting the occurrence of the specific exemplar of a chain during training. In the blockwise analysis we can, for instance, determine whether the rate of tests for stiff joints 2 and 3 had decreased to the same level by training block 4 in the infrequent (fourth presentation) and the frequent chains (14th-16th presentation). Indeed, there was no difference in the rate of tests for stiff joints 2 and 3 for the latter chains. More generally, the performance on high and low frequency chains was highly similar in all blocks of training. The uniform increase in exploration efficiency is in line with an account proposing that participants are acquiring rule knowledge that is applied to frequently encountered and infrequently encountered chains alike.

Interestingly, in tendency more of the stiffness tests on joints 2 and 3 were observed in the first training block of the group of participants exploring each chain with equal frequency as compared to the number of tests in the group of participants with frequency variation. This might suggest that learning of the regularity in the structure of the material was faster or was exploited faster for efficient exploration in participants with frequency variation. It is conceivable that knowledge of regularities in the material is generated relatively quickly based on the chains presented four times per block and then immediately transferred to the chains presented less frequently (compare Gaschler & Frensch, 2007). However further experimentation would be necessary to determine in detail whether knowledge develops in the frequent chains and transfers to the infrequent ones or vice versa. This would, first of all, include a replication of the data pattern as the ANOVA was not fully decisive with regard to the question of whether the equal frequency group deviated from the course of practice observed in the group of participants with frequency variation. There was a main effect of block, $F(2.19, 43.7) = 23.98$, $MSE = .016$, $p < .001$, $\eta_p^2 = .545$. The interaction of block and training group was marginal, $F(2.19, 43.7) = 2.59$, $MSE = .016$, $p = .082$, $\eta_p^2 = .115$. There was no main effect of group of participants ($F < 1$).

The blockwise analysis suggests equal increases in exploration efficiency for the high and the low frequency chains. While this null effect is consistent with the interpretation that participants were acquiring and employing rule knowledge to increase exploration efficiency, one could wonder whether the setup is actually suitable to demonstrate any influence of the rate of presentation of specific chains on performance. We therefore also charted the same data based on counting the occurrence of the specific example of a chain during training. As the high frequency chains were presented four times in each of four blocks, we have 16 data points. The four presentations of the low frequency chains over the course of practice lead to four data points, and sorting the presentation of the specific instances in the equal frequency group led to eight data points. The graph suggests that the reduction in the rate of tests for stiffness in joints 2 and 3 was much faster for low frequency compared to high frequency chains when plotted based on the instance counter. On average, the very first encounter with an infrequent chain led to a much lower rate of testing joints 2 or 3 for stiffness as compared to the first encounter with a high frequency chain. Notably, the first encounter with a specific low frequency chain usually occurred at a point in time during training in block one, when several frequently presented chains had already been processed. Apparently, the knowledge acquired during the processing of the latter was immediately transferred to the former. As practice on high frequency chains affected performance on low frequency chains from their first presentation onwards, the knowledge acquired cannot be specific to the high frequency chains. Rather, it seems to be rule-like. The observation that learning changed the performance on low frequency chains faster than on high frequency chains (if charted per presentation of the specific chain) was substantiated with a within-subjects ANOVA on the data of the group of participants with the frequency variation. This analysis was restricted to the first four encounters with each specific high frequency chain and included all four encounters with low frequency chains. As the rate of stiffness tests on joints 2 and 3 was overall lower in the low frequency chains, the ANOVA showed a main effect of frequency, $F(1, 12) = 6.24$, $MSE = .063$, $p = .028$, $\eta_p^2 = .342$. The overall decrease in the rate of testing stiffness in the joints 2 and 3 was reflected in a main effect of instance counter, $F(2.18, 26.19) = 15.4$, $MSE = .03$, $p < .001$, $\eta_p^2 = .562$. The steeper slope of learning on the high frequency as compared to the low frequency chains over the first four encounters led to an interaction of frequency and instance counter, $F(1.98, 23.75) = 3.72$, $MSE = .04$, $p = .04$, $\eta_p^2 = .237$.

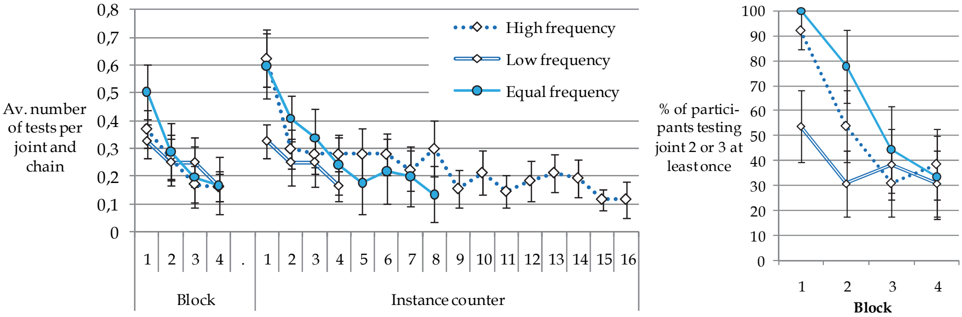


Fig. 10. For participants exposed to a selection of chains in which the joints 2 and 3 were never stiff, the average number of tests for stiffness per joint (2 and 3 averaged) is displayed over the course of practice, either by aggregating per block of practice or by aggregating per encounter with the specific chain. On the right side, we display the practice related decrease in the percentage of participants still testing stiffness in joints 2 or 3.

In summary, we can conclude that participants adapted to the regularity in the material. When confronted with material in which joints 2 and 3 were never stiff (but rather bending upward or downward) participants showed a marked reduction in the average number of tests for stiffness per chain on these joints. As exploration efficiency increased at the same time in practice and to the same extent for high and low frequency chains, we suggest that participants employed systematic exploration and developed rule knowledge on the structure of the chains. The data are consistent with the view that humans develop relational representations that capture high-level features and regularities of the objects. For the future, this encourages us to provide robots with human behavior in this and similar exploration situations in order to grant them with a set of adaptive exploration sequences they can expand upon. Comparison with robot object exploration will in turn allow us to judge what aspects of human exploration behavior come close to optimal exploration sequences and which aspects might be improved. This also counts for approaches to the exploration-exploitation dilemma. For instance, our analyses suggest that most of the participants eventually ceased exploration and started exploitation of the acquired knowledge. As shown in the right panel of Fig. 10 from block 3 onwards, the majority of participants did not check stiffness in joint 2 or 3 at all. They switched from exploration to exploitation mode. For instance, they would not have noted whether multiple characteristics had been ascribed to single joints. While all participants tested stiffness in the high frequency chains in block 1, results of these tests were apparently transferred to low frequency chains, by some participants already within the first block.

4. Summary

In this chapter we have described first steps for studying tool use in humans and robots in a common framework by focusing on how humans and robots explore the kinematic structure of objects. We have gathered initial evidence that representational formats, which make the problem of discovering and representing kinematic structure tractable for robots, may have similar counterparts in humans. Exploration experience could be used to render exploration more efficient both by robots and by humans. As we observed that humans adapt their exploration strategies to the constraints present in the pool of objects, future work can target the possibility that robots use the observation of human exploration behavior as a starting point for acquiring efficient strategies. So far we have used tasks in which exploration was applied to completely unravel the kinematic structure of an object. Expanding upon our results on humans exploiting redundancies in the structure of the objects, future research can address how exploration can be terminated once sufficient structural properties are discovered for tool use according to the current goal.

5. References

Blaisdell, A. P. (2008). Cognitive Dimension of Operant Learning, In: *Learning and Memory: A Comprehensive Reference*, Vol.1, R. Menzel, & J. Byrne, (Eds.), pp. 173-195, Elsevier, ISBN 0-12-370504-5, Oxford, Great Britain

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*, MIT Press, ISBN 0-262-02208-7, Cambridge, MA, USA

Craighero L, Leo I, Umiltà C, Simion F. (2011). Newborns' Preference for Goal-Directed Actions. *Cognition*. [Epub ahead of print] PubMed PMID: 21388616, ISSN: 0010-0277

Džeroski, S., de Raedt,L., & Driessens, K. (2001). Relational Reinforcement Learning. *Machine Learning*, Vol. 43, No. 1-2 , pp. 7-52, ISSN 0885-6125

Elsner, B., & Hommel, B. (2001). Effect Anticipation and Action Control. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 27, No. 1, pp. 229-240, ISSN 0096-1523

Frensch, P. A., & Rünger, D. (2003). Implicit Learning. *Current Directions in Psychological Science*, Vol. 12, No. 1, pp. 13-18, ISSN 0963-7214

Gaissmaier, W. & Schooler, L. J. (2008). The Smart Potential Behind Probability Matching. *Cognition*, Vol. 109, pp. 416-422, ISSN: 0010-0277

Gaschler, R., & Frensch, P. A. (2007). Is Information Reduction an Item-Specific or an Item-General Process?. *International Journal of Psychology*, Vol. 42, No. 4, pp. 218-228, ISSN 0020-7594

Gaschler, R., & Frensch, P. A. (2009). When Vaccinating Against Information Reduction Works and When It Does Not Work. *Psychological Studies*, Vol. 54, No. 1, pp. 43-53, ISSN 0033-2968

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, Vol. 42, No. 1-3, pp. 335-346, ISSN 0167-2789

Haider, H., & Frensch, P. A. (2002). Why Aggregated Learning Follows the Power Law of Practice When Individual Learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 28, pp. 392-406, ISSN: 0278-7393

Held, R., & Hein, A. (1963). Movement-Produced Stimulation in the Development of Visually Guided Behavior. *Journal of Comparative and Physiological Psychology*, Vol. 56, pp. 872-876, ISSN: 0021-9940

Kaelbling, L. (1993). *Learning in Embedded Systems*, MIT Press, ISBN 9780262512787, Cambridge, MA, USA

Katz, D., & Brock, O. (2008). Manipulating Articulated Objects with Interactive Perception, *Proceedings of the IEEE International Conference on Robotics and Automation 2008,* ISBN 978-1-4244-1646-2, Pasadena, CA, USA, May 2008

Katz, D., Orthey, A., & Brock, O. (2010). Interactive Perception of Articulated Objects. In: *The 12th International Symposium of Experimental Robotics (ISER) 2010*

Katz, D., Pyuro, Y., & Brock, O. (2008). Learning to Manipulate Articulated Objects in Unstructured Environments Using a Grounded Relational Representation, *Proceedings of Robotics: Science and Systems IV*, ISBN 978-0262513098, Zurich, Switzerland, June 2008

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review*, Vol. 95, pp. 492-527, ISSN: 0033-295X

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley. ISBN 0-486-44136-9.

Sun, R., Merrill, E., & Peterson, T. (2001). From Implicit Skills to Explicit Knowledge: A Bottom-up Model of Skill Learning. *Cognitive Science*, Vol. 25, pp. 203-244.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*, ISBN 978-0262193986, Cambridge, MA: MIT Press.

Tadepalli, P., Givan, R., & Driessens, K. (2004). Relational Reinforcement Learning: An Overview. *Proceedings of the Workshop on Relational Reinforcement Learning at ICML '04*, Banff, Canada, July 8, 2004

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*, Macmillan, New York, NY, USA

van Otterlo, M. (2005). A Survey of Reinforcement Learning in Relational Domains. *CTIT Technical Report series TR-CTIT-05-31*, Centre for Telematics and Information Technology University of Twente, Enschede, ISSN 1381-3625 [for more: http://eprints.eemcs.utwente.nl/1879/]

Weir, A. A. S., & Kacelnik, A. (2006). A New Caledonian Crow (Corvus Moneduloides) Creatively Re-designs Tools by Bending or Unbending Aluminium Strips. *Animal Cognition*, Vol. 9, No.4, pp. 317-334, ISSN 1435-9448