CHAPTER 7

Rock and roll: exploring and exploiting an object affordance

> *[The waterfall] was the second highest anywhere on the Disc and had been discovered in the Year of the Revolving Crab by the noted explorer Guy de Yoyo. Of course, lots of dwarfs, trolls, native people, trappers, hunters, and the merely badly lost had discovered it on an almost daily basis for thousands of years. But they weren't explorers and didn't count.* (Pratchett, 1991b)

In the end what matters in life and robotics is action, and perception should reflect that priority. Perception can be seen as "basically an implicit preparation to respond" (Sperry, 1952). This chapter introduces an approach to perception that is *explicitly* about preparation to respond. The perceptual system is assigned the task of continually preparing a set of actions that are possible in the robot's current situation, and which simply need to be selected and activated to take effect. This approach has similarities with Gibson's notion of affordances (Gibson, 1977), which is reviewed.

7.1 What are affordances?

Affordances are possibilities for action. If a creature can perform some action on an object, then the object is said to afford that action, and that action is an affordance of the object. For example, a cup affords grasping and drinking. The idea of affordance is actor-specific; a leaf might afford support to an ant but not an elephant. The existence of an affordance depends only on whether the creature can perform the appropriate actions, and does not depend on the ability of the creature to perceive it. Other authors have used the term in different ways. Those concerned with interface design, such as the Human-Computer Interface community, often use both perception and action as the defining characteristics of the term – see McGrenere and Ho (2000) for a review. We will take "perceived affordances" to refer to the actions a creature believes are possible on an object, which are potentially distinct from the "true affordances" that are physically realizable.

In Gibson's work, there is an implication that affordances can be perceived "directly" and are in some sense "picked up" from the environment – as opposed to being inferred. This is not a particularly helpful notion for robotics, and although there is a significant literature on the notion of direct perception, it will not be reviewed here (see Hurley (2001) for a discussion). Gibson did make good points about vision being easier when done dynamically from a moving perspective, ideas that cropped up later as active/animate vision. Gibson pointed out the power of optic

flow information, which this thesis has benefited from (and in other collaborative work even more (Fitzpatrick and Metta, 2002)).

7.2 Why think about affordances?

In robotics, possibilities for action are captured concisely in a robot's configuration space. A configuration space contains all the parameters (e.g. joint angles) necessary to uniquely specify the robot's physical state. Actions correspond to trajectories in the configuration space. So why do we need another way to think about the space of possible actions?

Configuration space is very useful for planning the details of motor control, such as getting from one point to another without trying to move through any impossible joint angles. If there are complicated constraints on action then this tactical level of analysis is unavoidable. However, at the strategic level, it isn't as helpful. When the robot is deciding what it should do now, joint angle trajectories are the wrong level of abstraction. One choice would be to switch into a space with a representation of goals and possible operators, and do planning, and then later translate operators back into motor actions using plan execution. This involves a significant architectural commitment. It is useful to consider if there are alternatives that don't involve such a dramatic "phase change" between motor control and perception. Affordances offer such an alternative. If there is a predominantly bottom-up system assessing what is possible in the robot's current situation, then it can prepare the appropriate control parameters for the available actions, and describe the actions in a very low-bandwidth way relative to this – with all the awkward details suppressed. Is this different from planning, picking an action, and then assessing what parameters are appropriate to carry it out? Not in principle, but in practice it could significantly simplify the decisions that need to be made.

Configuration space ideas do have the benefit of being formalized and clear, unlike affordances. We could define an "affordance space" as the set of control parameters output by perception so that initiated actions are channelled appropriately, and then a set of action flags specifying which actions seem possible. For example, an affordance-based perceptual system for a mobile robot might chose to signal "turn" as a possible action to its behavior system, while setting up appropriate control parameters to achieve a good turn or to continue on straight. If all the robot does is navigate then there is not much benefit to this; but if the robot has a wide vocabulary of actions then this may be a useful simplification. The danger of using a weaker perception system that makes minimal judgements itself is that it will add delay, and potentially leave decisions in the hands of a less well-informed module. Of course, not everything is evident from perception, so integration is still important.

7.3 Exploring an affordance

The examples of affordances that are most commonly discussed include different kinds of grasping, twisting, or chewing. All of these require quite sophisticated motor control. As a starting point, it seemed more sensible to choose actions that have all the properties of an affordance, but have a lower cost of entry in terms of dexterity. The author identified object rolling as an excellent candidate. Only certain objects roll well, and to make them roll requires matching the robot's action to the object's pose in an intelligent manner. For example, Figure 7-1 shows four objects that have quite distinct properties in terms of a "rolling affordance". The rolling affordance is perfectly within reach of the robot, given the capabilities already developed. It can poke an object from different directions, and it can locate familiar objects and recognize their identity. Active segmentation played two roles

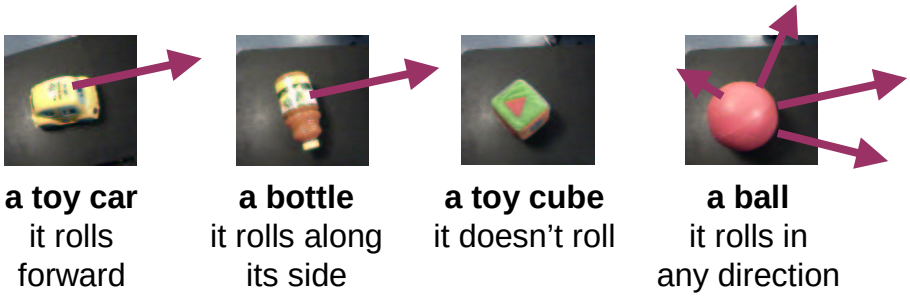

Figure 7-1: Different objects roll in different ways. A toy car rolls forward, a bottle rolls on its side, a ball rolls in any direction, and a cube doesn't roll easily at all.

in this experiment: collecting data for later object recognition and localization, and providing a good segmentation for tracking the motion of the object after contact. Chronologically, this experiment was performed before the techniques for tracking, recognition, and localization described elsewhere in this thesis were fully developed, so simpler methods were used (color histogram back-projection for localization and recognition, optic flow based tracking over a small number of frames). This system was developed in collaboration with Giorgio Metta.

We designed two experiments that use poking and the visual segmentation described in Chapter 3 to move an object on the basis of the rolling affordance. In the first experiment the robot poked the set of objects shown in Figure 7-1 (an orange juice bottle, a toy car, a cube, and a colored ball) using one of four possible actions (the motor repertoire). Actions are labelled for convenience as back-slap, side-tap, pull-in, and push-away. These actions correspond to different patterns of poking. In a side-tap, the robot sweeps its arm in across the field of view. In a back-slap, the robot first raises its arm and draws it in to its torso, then sweeps outwards. Normally these actions are used interchangeably and at random during poking, as the segmentation algorithm is agnostic about the source of object motion (see for example Figure 3-11). The toy car and the bottle tend to roll along a definite direction with respect to their principal axis. The car rolls along its principal axis, and the bottle rolls orthogonal to it. The cube doesn't really roll because of its shape. The ball rolls, but in any direction. Shape information can be extracted from the segmentation produced by poking, so these relationships could in principle be learned – and that is the goal of this experiment.

The robot poked the set of objects shown in Figure 7-1 many times (approximately 100 each), along with other distractors. The segmented views were clustered based on their color histogram. For each poking episode, shape statistics were gathered at the point of impact, and the overall translational motion of the object was tracked for a dozen frames after impact. Over all poking events (470 in all) the gross translation caused by poking was computed as a function of the type of poking applied (back-slip, side-tap, pull-in, push-away), as shown in Figure 7-2. This is necessary since the effect that the poking fixed action pattern has is not known to the robot's perceptual system; this procedure recovers the effect (and also reveals that the motor control on Cog's arm is very erratic). Using this procedure, the robot automatically learns that poking from the left causes the object to slide/roll to the right, as a general rule. A similar consideration applies to the other actions. Next, object-specific models are built up relating the effect of object orientation on the translation that occurs during poking. Figure 7-3 shows the estimated probability of observing each of the objects rolling along a particular direction with respect to its principal axis. Here the peculiar properties of the car and bottle reveal themselves as a "preferred direction" of rolling. The ball and cube do not have such a preference. At the end of the learning procedure the robot has built a representation of each object in terms of:

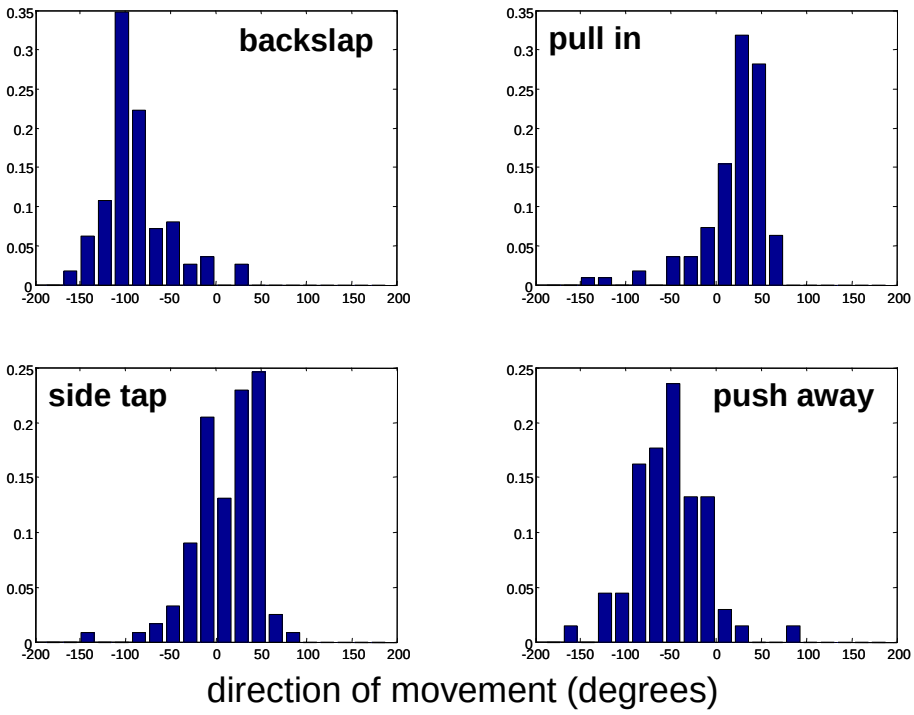

Figure 7-2: Histogram of the direction of movement of object for each possible poking action. For each of the four plots the abscissa is the direction of motion of the object where the 0° direction is parallel to the x axis, and -90° to the y axis. The ordinate is the empirical probability distribution of the direction of motion of the objects.

- ▷ Pictorial information in the form of color histograms, following (Swain and Ballard, 1991).
- ▷ Shape information in the form of a measure of the average size of the object, an index of the elongation of the object with respect to its principal axis, and a set of Hu moments (Hu, 1962).
- ▷ Detailed histograms of the displacement of the object with respect to its initial orientation given that a particular motor primitive was used.
- ▷ The summary histograms shown in Figure 7-3 which capture the overall response of each object to poking.

After the training stage, if one of the known objects is presented to Cog, the object is recognized, localized and its orientation estimated (from its principal axis). Recognition and localization are based on the same color histogram procedure used during training (Swain and Ballard, 1991). Cog then uses its understanding of the affordance of the object (Figure 7-3) and of the geometry of poking to make the object roll. The whole localization procedure has an error between 10° and 25° which which is perfectly acceptable given the coarseness of the motor control We performed a simple qualitative test of the overall performance of the robot. Out of 100 trials the robot made 15 mistakes. A trial was classified as "mistaken" if the robot failed to poke the object it was presented with in the direction that would make it roll. The judgements of the appropriate direction, and whether the robot succeeded in actually achieving it, were made by external observation of the behavior of the robot. Twelve of the mistakes were due to imprecise control – for example the manipulator sometimes moved excessively quickly and shoved the object outside the field of view. The three

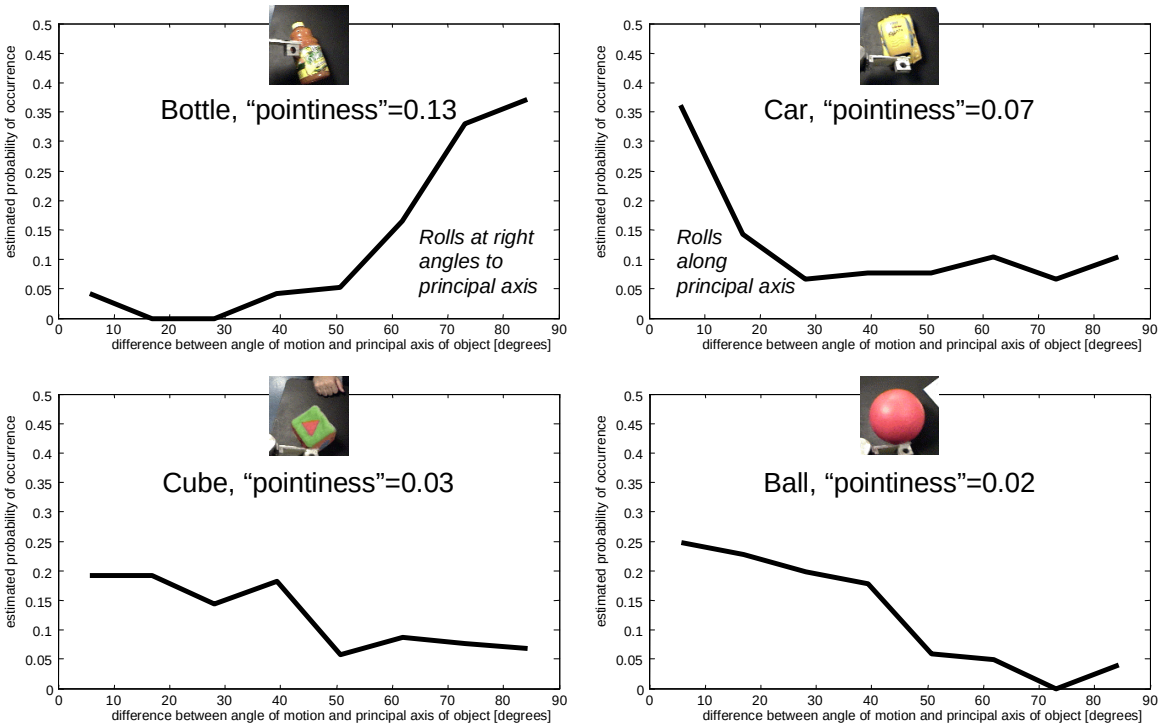

Figure 7-3: Probability of observing a roll along a particular direction for the set of four objects used in Cog’s experiments. Abscissae represent the difference between the principal axis of the object and the observed direction of movement. Ordinates are the estimated probability. The principal axis is computed using the second Hu moment of the object’s silhouette (Hu, 1962). The “pointiness” or anisotropy of the silhouette is also measured from a higher order moment; this is low when the object has no well-defined principal axis, as is the case for the cube and the ball. The car and bottle have clear directions in which they tend to roll. In contrast, the cube slides, and the ball rolls, in any direction. These histograms represent the accumulation of many trials, and average over the complicated dynamics of the objects and the robot’s arm to capture an overall trend that is simple enough for the robot to actually exploit.

remaining errors were genuine mistakes due to misinterpretation of the object position/orientation. Another potential mistake that could occur is if the robot misidentifies an object – and, for example, believes it sees a bottle when it in fact sees a car. Then the robot will poke the object the wrong way even if it correctly determines the object’s position and orientation.

7.4 Mimicry application

With the knowledge about objects collected in the previous experiment we can then set up a second experiment where the robot observes a human performing an action on the same set of objects, and then mimics it. In fact, the same visual processing used for analyzing a robot-generated action can be used in this situation also, to detect contact and segment the object from the human arm, as described in Chapter 6. The robot identifies the action observed with respect to its own motor vocabulary. This is done by comparing the displacement of the object with the four possible actions, as characterized in Figure 7-2, and choosing the action whose effects are closer to the observed

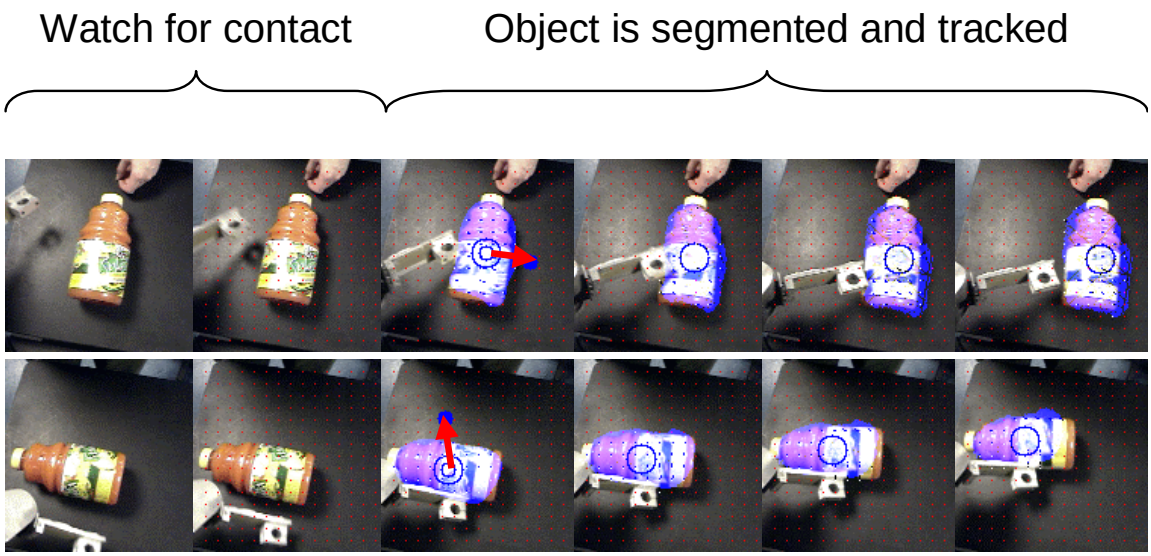

Figure 7-4: Frames around the moment of contact are shown. The object, after segmentation, is tracked for 12 frames using a combination of template matching and optic flow. The big circles represent the tracked position of the bottle in successive frames. The arrow displayed on the frame of contact (3^{rd} from the left) projects from the position at the time of contact and at the 12^{th} frame respectively. In the first sequence, the bottle is presented to the robot at an orientation that makes a side-tap appropriate for rolling, and that is what the robot does. In the second sequence, the car is presented at a different angle. The appropriate action to exploit the affordance and make the bottle roll is now a back-slap.

displacement. This procedure is orders of magnitude simpler than trying to completely characterize the action in terms of the observed kinematics of the movement.

The robot can then mimic the observed behavior of the human if it sees the same object again. The angle between the preferred direction of motion of the object (as characterized in Figure 7-3) and the observed displacement is measured. During mimicry the object is localized as in the previous experiment and the robot picks the motor action which is most likely to produce the same observed angle relative to the object. If, for example, the car was poked at right angle with respect to its principal axis Cog would mimic the action by poking the car at right angle, despite the fact that the car's preferred behavior is to move along its principal axis. Examples of observation of poking and generation of mimicry actions are shown in Figures 7-5.

7.5 Conclusions

Describing a problem the right way is an important step to solving it. Affordances provide a means to implement bottom-up influence on the terms in which the current situation is described. For example, in the work described here, if the perceptual system detects that the robot is looking at an object that can roll, then the motor options automatically change. Poking will now automatically push the object in the right direction to make the object roll. The option to strike the object awkwardly is now available – which the robot is pretty good at anyway, but now it can do it deliberately, to mimic human action for example. Control in terms of side-taps and back-slaps is still possible of course, but that level of detail is no longer necessary.

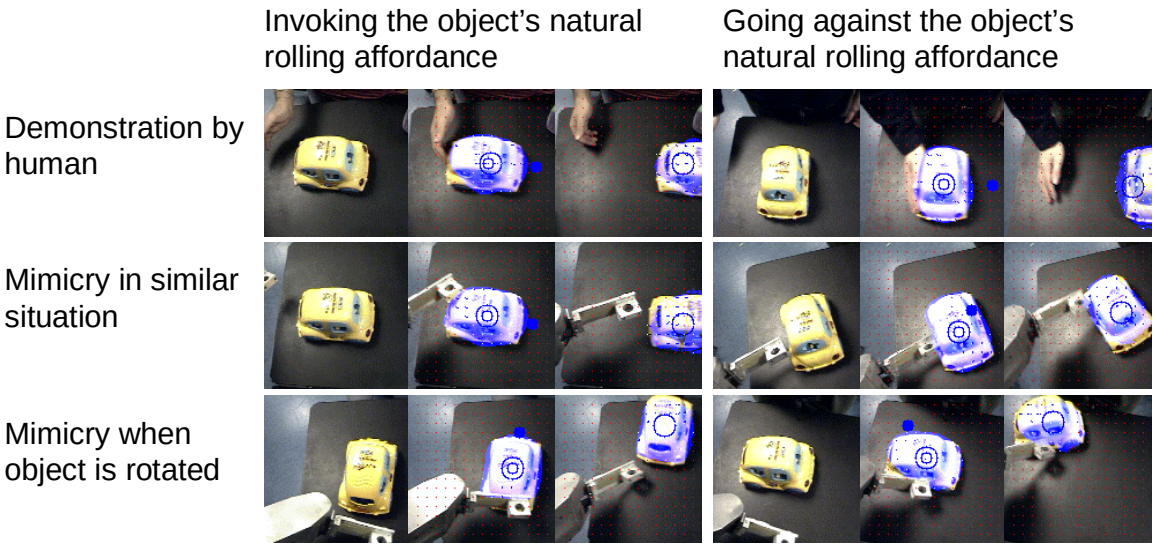

Figure 7-5: A mimicry example using the toy car. The first row shows human demonstration of poking operations, which the robot then mimics. The sequences on the left show the robot mimicking a human exploiting the car's rolling affordance. The sequences on the right show what happens when the human hits the car in a contrary fashion, going against its preferred direction of motion. The robot mimics this "unnatural" action, suppressing its usual behavior of trying to evoke rolling. Mimicry is shown to be independent of the orientation at which the car is presented.

Figure 7-6: You don't have to have high dexterity to explore some properties of objects.