Chapter 2: The Basic Framework

2.1 Focus of Attention Methods

2.1.1 Introduction

To ease its learning task, an agent can employ a range of methods for focus of attention. It can be selective in terms of what sensor data it attends to as well as what internal structures it considers when acting and learning. These forms of focus of attention are termed *perceptual* and *cognitive selectivity* respectively. They are illustrated by the left and right braces respectively in Figure 1 below, and discussed in more detail in the following sections.

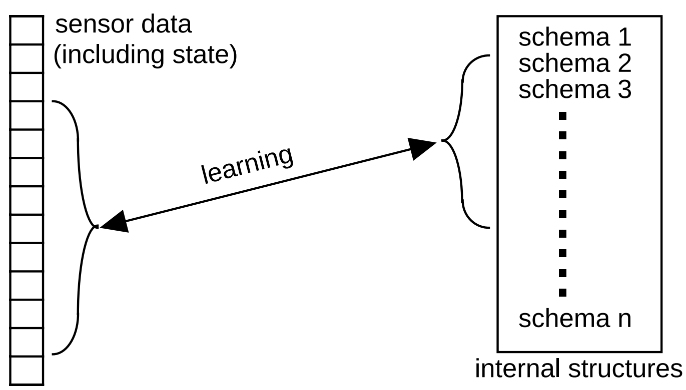

Figure 1: Sensory (left) and cognitive (right) pruning

Along another dimension, there is a distinction between *domain-dependent* and *domain-independent methods* for focus of attention. Domain-independent methods represent general heuristics for focus of attention that can be employed in any domain. For example, one can attempt only to correlate events that happened close to one another in time. Domain-dependent heuristics, on the other hand, are specific to the domain at hand. They typically have been preprogrammed (by natural or artificial selection or by a programmer). For example, experiments have shown that when a rat becomes sick to its stomach, it will assume that whatever it ate recently is causally related to the sickness. That is, it is very hard for a rat to learn that a light flash or the sound of a bell is correlated to the stomach problem because it will focus on recently eaten food as the cause of the problem [Garcia 72]. This demonstrates that animals have evolved to pay attention to particular events when learning about certain effects.

Finally, the focus mechanism can be *goal-driven* and/or *world-driven*. Focus of attention in animals is both strongly world- and goal-driven. The structure of the world and the sensory system determines which sensory or memory bits may be "usually" ignored (e.g. those not local in time and space), while the task determines those which are relevant some of the time and not at other times. For example, when hungry, any form of food is a very important stimulus to attend to; learning how to get to some would presumably take on greater important in this case.

The results reported in Chapter 3 concern world-dependent, domain-independent cognitive and sensory selectivity. Such pruning depends on invariant properties of the environment[1] and common tasks, and does not take into account what the current goal of the agent is. The methods can be applied to virtually any domain. While it is true that, in complex worlds, goal-driven and domain-dependent pruning is quite important, it is surprising how much of an advantage even goal-independent pruning can convey. Using Chapter 3's work as a base, Chapter 4 then investigates the added leverage of adding goals to the learning

1. In a physical, terrestrial environment, such properties might include causality (actions must precede their effects), locality (most effects are near their causes), space, time, gravity, etc.

system. That is, the actions taken and cognitive and perceptual pruning that occurs are controlled by the short-term goals of the agent.

2.1.2 Perceptual selectivity

Perceptual selectivity limits what stimuli might possibly be attended to at any one time, which puts limits on what might be learnable at that time. For example, a real creature would not pay attention to every square centimeter of its skin and try to correlate every nerve ending therein to every possible retinal cell in its eyes at every moment. Consequently, it might never learn some peculiar correlation between a particular patch of skin and a flash of light on some part of its retina, but presumably such correlations are not important to it in its natural environment.

Obvious physical dimensions along which to be perceptually selective include *spatial* and *temporal* selectivity.[2] The universe tends to display spatial locality: many causes are generally located nearby to their effects (for example, pushing an object requires one to be in contact with it). Further, many causes lead to an observable effect within a short time (letting go of an object in a gravity field causes it to start falling immediately, rather than a week later). Real creatures use these sorts of spatial and temporal locality all the time, often by using eyes that only have high resolution in a small part of their visual field, and only noticing correlations between events that take place reasonably close together in time. While it is certainly *possible* to conceive of an agent that tracks every single visual event in the sphere around it, all at the same time, and which can remember pairs of events separated by arbitrary amounts of time without knowing a priori that the events might be related, the computational burden in doing so is essentially unbounded.[3] The algorithm dis-

2. Another dimension of selectivity concerns the amount of preprocessing done to the input. For example, Drescher points out that, in a realistic world, correlating unprocessed retinal input is not very useful, because it does not map well onto aspects of the world that are good building blocks for inductive generalization [Drescher 94]. Such changes of level are not addressed in this research.

cussed in this research implements temporal selectivity as well as spatial selectivity to reduce the number of sensor data that the agent has to correlate with its internal structures (see Figure 1, on page 23). Note that the perceptual selectivity implemented is of a passive nature: the agent prunes its “bag of sensory bits,” rather than changing the mapping of that bag of bits to the physical world by performing an action that changes the sensory data (such as changing its point of view). The latter would constitute active perceptual attention (e.g., [Aloimonos 93] among others).

2.1.3 Cognitive selectivity

Cognitive selectivity limits what internal structures are attended to at any given moment.

Notice that for any agent that learns many facts,[4] being cognitively selective is likely to be even more important than being perceptually selective, in the long run. The reasons for this are straightforward. First, consider the ratio of sensory to memory items. While the total number of possible sensory bits is limited, the number of internal structures may grow without bound.[5] This means that, were we to use a strategy which prunes all sensory information and all cognitive information each to some *constant fraction* of their original, unpruned case, we would cut the total computation required by some constant factor—but this factor would be much larger in the cognitive case, because the number of facts stored would likely far outnumber the number of sensory bits available.

3. Many algorithms for learning from experience employ an extreme form of temporal selectivity: the agent can only correlate events that are “one timestep” apart.
4. In this case, since we are discussing a causal model builder, these facts are correlations of actions and their results.
5. The assumption here is that every fact learned requires some internal structure to represent it. If the learning algorithm in use must examine prior facts to decide whether to invalidate the fact, create a new one, etc, then the computational effort of the algorithm will tend to increase as more facts are learned.

Second, consider a strategy in which a *constant number* of sensory bits or a *constant number* of remembered facts are attended to at any given time. This is analogous to the situation in which a real organism has hard performance limits along both perceptual and cognitive axes; no matter how many facts it knows, it can only keep a fixed number of them in working memory. In this case, as the internal structures grow, the organism can do its sensory-to-memory correlations in essentially constant time, rather than the aforementioned $O(n^2)$ time, though at a cost: as its knowledge grows, it is ignoring at any given time an increasingly large percentage of all the knowledge it has.

Compromise strategies which keep growth in the work required to perform these correlations within bounds (e.g., less than $O(n^2)$), yet not give in completely to utilizing ever-smaller fractions of current knowledge (e.g., more than $O(1)$) are possible.

For example, one can use properties of the world or characteristics of the sensor data to restrict the number of structures looked at (as is the case in the algorithms described in this thesis). Not all internal structures are equally relevant at any given instant. In particular, internal structures that refer neither to current nor expected future perceptual inputs are less likely to be useful than internal structures which do. This is the particular domain-independent, goal-independent, world-driven heuristic for cognitive selectivity which is employed in Chapter 3. One might argue that, in real creatures, evolution optimizes them to ignore those aspects of the environment which do not change; for example, there is little reason to perceive nor reason about the existence of air unless one is in an environment in which it is not ubiquitous.

Another way to compromise is to use the current *goal* to help select what facts are relevant; such *goal-driven pruning* is discussed in Chapter 4. Since generally only a small number of goals are likely to be relevant or applicable at any one time (often only one), this can help to keep the amount of correlation work in bounds.[6] Again, in real creatures domain-dependent and goal-driven cognitive selectivity play a large role too. For example, the subset of internal structures that are considered at some instant is not only determined by what the agent senses and what it expects to sense next, but also by what it is “aiming” to sense or not sense (i.e., the desired goals).

2.2 The testbed microworlds

In order to evaluate the effects of adding focus of attention to a system that can learn a world model, we need something which learns (the *agent*) and some world model for it to learn in (the *domain* or *environment* of the agent). The two agents chosen here were both software agents, operating in a a *microworld* consisting of a simulated environment. The combination of a particular agent and its associated environment is called a *scenario*.

The sections immediately following describe the two scenarios used here. Later, in Section 3.2.1, on page 48, we describe what the learning system does with the sensory information it receives from either environment.

In both scenarios, the world may be unpredictable; actions are allowed to have no perceivable effect for any of several reasons, including incomplete sensor information (in neither scenario does the agent have an all-encompassing sensor view of the world), and external motions or actors in the world that cannot be controlled by the agent.

In general, the learning system is connected “at arm’s length” to either of the two microworlds, and can issue only one of a small number of commands at each timestep of the simulation. It gets back a collection of sensory bits describing what the simulated sensors are perceiving, and does not have any other access to the internal state of the micro-

6. Another way to compromise might be to investigate much more of memory when other demands on the agent’s time are minimal, essentially doing as much extra work as possible when not otherwise occupied with immediate concerns or goals which must be completed under tight deadlines.

world. Actions are taken at random in the case of goal-independent learning, but are informed from the goal system when performing goal-dependent learning.

Appendix A describes the overall system architecture used in this research, and shows a schematic of how the learning system, the action selection system, and the goal system interact. Figure 20, on page 141, describes what pieces of the system communicate with which other pieces, and Figure 21, on page 143, shows details of when the learning system and the microworld are allowed to exchange information. Appendix A also provides some implementation details specific to this particular system, which might be of use in reproducing it.

2.2.1 The infant/eyehand scenario

The most extensively studied scenario, both in prior work with the particular learning system employed [Drescher 91], and in this research, concerns a simulated infant in a simple, mostly (but not completely) static microworld.

Drescher's original system is concerned with Piagetian modeling, so his microworld is oriented towards the world as perceived by a very young infant (younger than eight months old, e.g., before early fourth Piagetian stage). The simplified microworld, shown in Figure 2, on page 30, consists of a simulated, two-dimensional "universe" of 49 grid squares (7 by 7). Each square can be either empty or contain some object. Superimposed upon this universe is a crudely simulated eye which can see a square region 5 grid squares on a side, and which can be moved around within the limits of the simulated universe. This eye has a fovea of a few squares in the center, which can see additional details in objects (these extra details can be used to differentiate objects enough to determine their identities). The universe also includes a hand which occupies a grid square, and can bump into

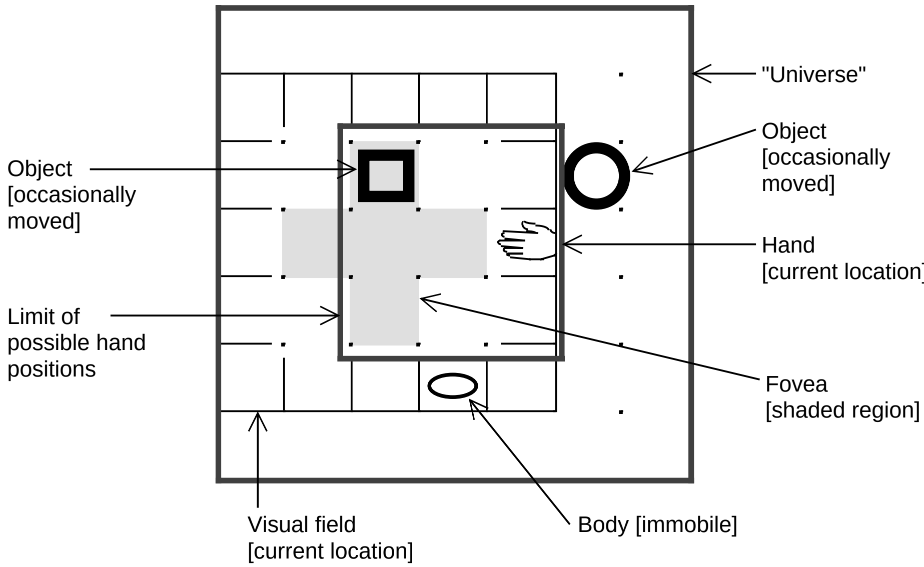

Figure 2: The eyehand/infant domain microworld

and grasp objects. (The infant's arm is not represented; just the hand.) An immobile body occupies another grid square.

The infant has 10 possible actions that it can take at any given timestep, consisting of moving the hand or eye forward, backward, left, or right (8 actions in all), and of opening or closing the hand. It takes one of these actions at every timestep.

The possible sensory inputs consist of all bits arriving from the eye, proprioceptive inputs from eye and hand (which indicate where, relative to the body, the eye is pointing or the hand is reaching), tactile inputs from the hand and body, and taste inputs from the mouth (if an object was in contact with it).[7] The eye reports only whether an object (not *which* object, only the presence of one) is in a grid square or not, except in its central fovea, where it reports many additional bits.[8]

7. The hand gets four one-bit details when it is in contact with an object on its left side, which can be used to differentiate objects by touch. It also gets one bit per side indicating whether an object is in contact with it on that side. Similarly, the body gets one bit apiece to indicate contact on any of its four sides; if an object is in contact on the front side of the body (the mouth), then four bits of taste information are also available. Finally, there is one bit each representing whether the hand is currently closed and whether it is grasping an object.

The simulated infant does *not* have a panoramic view of all 49 squares of the universe at once; at any given instant, it only knows about what the eye can see, what the hand is touching, or what the mouth is tasting, combined with proprioceptive inputs for eye and hand position. In particular, certain senses, viewed unimodally, are subject to perceptual aliasing, in which two distinct situations in the environment appear identical to the sensory system of the agent [Whitehead and Ballard 90]. For example, if a schema mentions only a particular bit in the visual field, without also referring to the visual proprioceptive inputs (which determine where the eye is pointing), then that schema may be subject to such aliasing—several different situations have been collapsed into the same representation, as far as the agent is concerned. Similarly, any schema mentioning any visual-field sensory item that is not in the fovea may alias different objects, since the non-foveal visual field reports only the presence or absence of an object in each position, rather than the exact identity of the object in question.

Typical knowledge that is learned about the world includes correlations between motion of the hand and motions of the image of the hand in the visual field; motions of the eye and motions of all objects in the visual field; correspondences between proprioceptive and haptic or visual information; whether or not an object will be graspable depending on whether or not it is felt to be in contact with the hand, and many more ([Drescher 91] and [Ramstad 92] report at length the many unimodal and multimodal facts about the microworld learned by this system).

8. Each of the five foveal squares provides 25 bits of detail information. Each different object sets a different combination of these 25 bits; hence, objects may be differentiated visually when they are in a foveal square, because different combinations of bits are perceived. Note that each foveal square covers one coarse visual square, and all motion is quantized by these coarse visual squares: this means that the details corresponding to some object will never be half on one foveal square and half on another. To put this another way, if an object is somewhere in the fovea, there are only five different positions (corresponding to the five different foveal squares) that any one object can be in. Objects are never rotated.

Typical strategic goals for the infant in this microworld include centering an object in the visual field, moving the hand into proximity to an object in order to grasp it, and so forth. Clearly, such goals cannot be accomplished in reasonable time without having learned about the effects of the infant's actions.[9]

2.2.2 The Hamsterdam scenario

The second agent and environment used in this research were based on the Hamsterdam system [Blumberg 94]. This system's primary use is for investigating ethological models of action-selection [Blumberg 94], and it is also used as a major component of the "magic mirror" virtual environment created for the ALIVE project [Darrell 94] [Maes 93].

The Hamsterdam system consists of a three-dimensional world in which simulated hamsters and predators may interact. The world also includes a floor, walls, food and water sources, and so forth. A second instantiation of the ALIVE project included a humanoid puppet, which ran under control of a simple finite-state machine, rather than the more complicated, ethologically-based controllers used for the hamsters and the predators. The entire microworld ran in real time, and was rendered as it ran using SGI Inventor on a Silicon Graphics workstation. Figure 3, on page 33, shows a typical scene from Hamsterdam, in which a hamster is in the lower left, and a predator is in the upper right. (The predator is currently unable to escape its box of walls, but a human outside of the simulation has the ability to slide the wall aside and enable the predator to reach the hamster.) Figure 4, immediately below Figure 3, shows the puppet, pictured standing alone in the world.

9. The phrase "reasonable time" is important: since the world is relatively small and the action set constrained, it might be possible to accomplish certain strategic goals by taking *random* actions and eventually reaching the goal by luck. However, as demonstrated in Section 4.4.3, on page 99, the average number of actions employed to reach a typical strategic goal are at least an order of magnitude shorter when the agent has learned the consequences of its actions than when the agent is not allowed to learn.

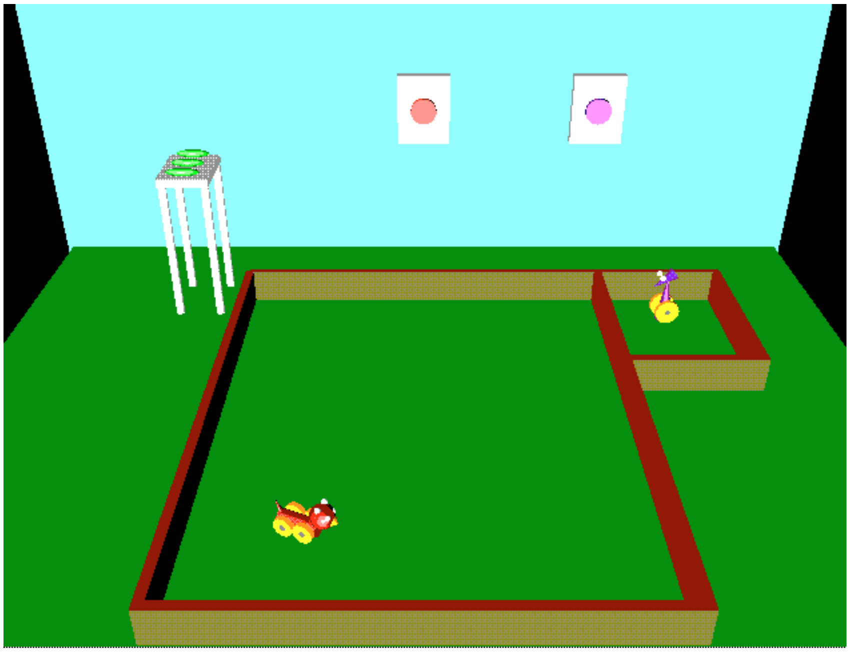

Figure 3: Hamsterdam with a hamster and a predator

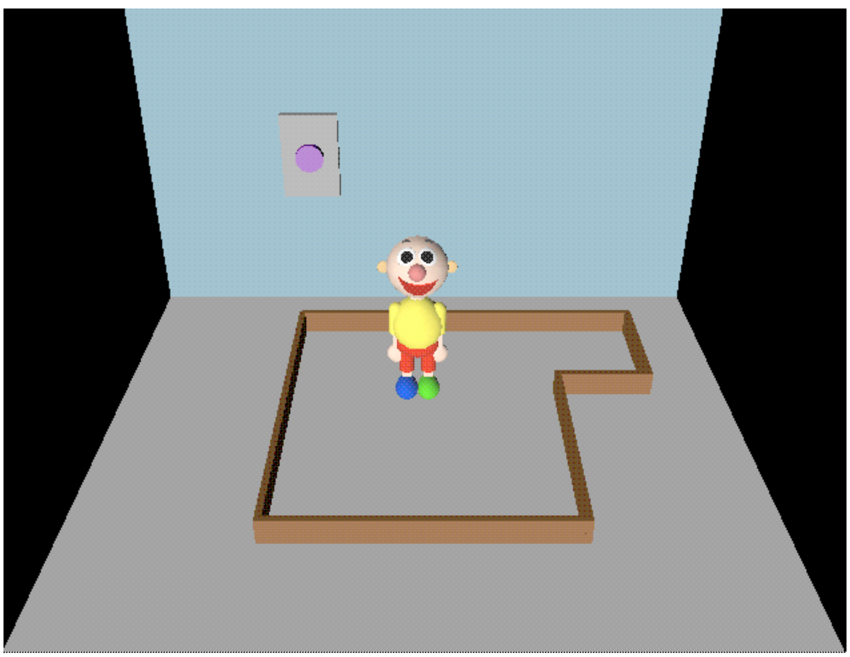

Figure 4: Hamsterdam with a puppet

The animals in the original Hamsterdam system have a sensory system resembling that of a robot sonar system. They shoot out 15 simulated rays in a horizontal fan at floor level, subtending 180 degrees total, and get back sonar-like echoes which indicate, along each ray, how close any given object is. The echo actually reveals the identity and type of any given object, rather than simply reporting that "something" is there. A typical representation of such a sensor fan appears in Figure 5, below, as rendered by Hamsterdam, in which

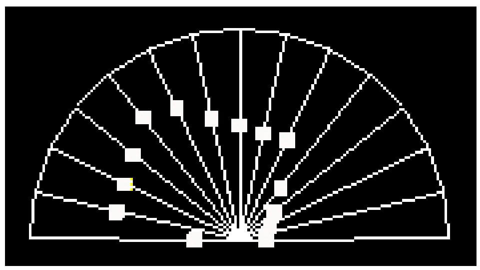

Figure 5: Hamsterdam sensor fan

the white squares at various distances along the radials are the sensor echoes of (in this case) part of a cul-de-sac where several walls meet. This sensory system, unlike that in the infant/eyehand scenario, is not particularly multimodal. The information returned for an object consists of the range r and angle θ at which the object is sighted: this is essentially purely visual. No proprioceptive information is available. The qualities reflected by taste and texture in the infant/eyehand scenario are reflected most closely by the tags returned by the sensor system, which indicate object identity and type.

Since the underlying learning system used in this research (e.g., the *schema system* [Drescher 91], without the composite-action system) requires all sensory information to be reduced to individual boolean predicates, rather than, e.g., numbers or ranges (see Section 3.2.1, on page 48), the results of the sensory fan must be discretized. This is accomplished as follows.

First, positions in the polar coordinate system defined by r and θ are quantized. The original, continuous value of r is reduced to one of five ranges, compressing r values to only five distinct quantities. These ranges are, in fact, scaled logarithmically, so that ranges which are farther from the agent cover more distance in the unscaled, original space; this provides more detailed range information for objects which are nearby without requiring a large number of additional ranges.

If an object is at some particular combination of r and θ, the sensory item (a single bit) corresponding to that combination is turned on. Since there are 15 radials and 5 ranges, this means that there are 75 sensory items devoted to *r*/*theta* information. Note that these sensory items cannot be used to differentiate one object from another, but merely to indicate that an object occupies that particular position. In this respect, they are similar to the coarse visual field items in the infant/eyehand scenario.

In addition to these inputs, the sensor fan is also "foveated" to yield higher-quality information at short ranges in the center of the fan. The central three radials, for the nearest four ranges, also return one bit apiece to indicate whether an object at that position is one of: a hamster, a predator, a wall, food, or water. The cross product of these 3 by 4 by 5 possibilities yields another 60 sensory items.

The discretized sensory system is illustrated in the schematic below. The foveation is

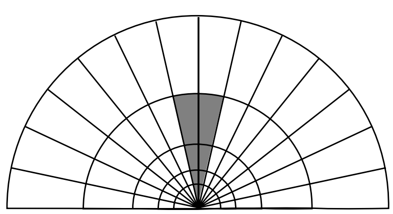

Figure 6: Discretization and foveation of the Hamsterdam sensor fan

shown by the gray region. Note that the logarithmic scaling of r is actually more pronounced than shown in this diagram

This sensory system has all of the possible problems with sensory aliasing described in Section 2.2.1, on page 29, in the description of the infant/eyehand scenario, and more. In particular, the infant/eyehand scenario at least has proprioceptive inputs from its eye, so it could (in theory) be able to build a complete world picture by taking into account the current portion of the world encompassed by its eye (as revealed by proprioceptive inputs), combined with current inputs from the eye itself. In the Hamsterdam scenario, even that level of information is unavailable, since the actors in the world are free to roam about it, and information about their current position or orientation is not available from the sensory system (and would have to be inferred either from available sensory information, dead reckoning from a known landmark, or something similar). If the infant/eyehand scenario did not make proprioceptive sensory items available from the hand and eye, it would resemble this aspect of the Hamsterdam scenario.

In the ALIVE system [Maes 93], the puppet never “shared the stage” with the hamsters and the predators; instead, users of the systems could switch between these two worlds. As modified for this research, the puppet and the animals are both allowed to share the same world.

Further, the puppet in ALIVE did not have an ethologically-based action controller; instead, inputs from the visual tracking system which tracked human participants drove a simple finite-state machine which in turn commanded various movements of the puppet. As modified here, the puppet’s actions are controlled by the learning system, and it has had the hamster sensory system “grafted on” to serve as input sensory items to the learning system.

The learning system may only control the puppet, and not any of the other animals or, e.g., the position of the walls. The allowable motions of the puppet include walking forward a step, turning in either direction,[10] standing up from a sitting position, and sitting down again.

As in the infant/eyehand scenario, learning the consequences of actions includes learning the correlation between actions such as rotating or walking and the observed movement of objects in the sensor fan. Goals include rotating until an object is centered along a foveal ray, walking until an object is at minimum range, and so forth.

Because the Hamsterdam scenario is significantly more active than the infant/eyehand scenario (e.g., hamsters or predators may be rolling around all the time, causing many changes to the environment regardless of what the agent is doing), the learning system has an explicit representation of a *null action* in this scenario, used to help model the *result of not doing anything*. See Appendix A for further details about null actions.

10. The puppet turns by an amount which matches the angular offset of a pair of rays in the sensor fan, e.g., about 12.857 degrees. This was chosen to conveniently match the action to the sensory system, in the way that moving the eye or hand in the eyehand/infant scenario causes all the other sensory inputs (e.g., proprioceptive, visual) to slide one position in some direction.