

BRIAN MINGUS / UNDERGRADUATE THESIS

Learning to reach.

Developmental robotics

Psychology / CU Boulder / 2008

HONORS THESIS | APRIL 5, 2008

Computational Principles of Infant Reaching Strategies in Developmental Robotics

Brian Mingus

University of Colorado Boulder

An embodied model connects infant reaching, motor learning and biologically plausible reinforcement learning in a developmental robotics setting.

Collection cover added September 2026. Original manuscript follows unchanged.
AI-generated conceptual artwork; separate from the scientific figures.

brianmingus.com

Running head: REACHING IN DEVELOPMENTAL ROBOTICS

Computational principles of infant reaching strategies in developmental robotics

Brian Mingus

Advisor: Dr. Randall O'Reilly

Computational Cognitive Neuroscience Lab

Department of Psychology

University of Colorado at Boulder

April 5, 2008

Abstract

The field of developmental robotics stands to benefit from recent advances in the study of the development of reaching. These include verification of the minimum jerk hypothesis, the discovery that the adult nervous system finds excess degrees of freedom to be a luxury, clarifications as to the topographic organization of motor cortex, as well as a collection of datapoints that allow for the principled development and analysis of robotics models of infant reaching. In order to demonstrate the validity of these principles, a robot was developed using the Emergent Neural Network Simulator. The robot accounts for recent findings regarding motor cortex, solves the inverse kinematics of motion and learns using the biologically plausible Primary Value Learned Value reinforcement learning system, a first for the field. The flexibility of this framework allows for future, more realistic robots to be developed, including those that learn via autonomous mental development, the automatic selection of tasks based on intrinsic motivations.

Computational principles of infant reaching strategies in developmental robotics

Introduction

The first significant problem an infant must solve is learning to control its body, filtering and mapping low fidelity sensory input to only one out of an infinite number of possible motor outputs and further anticipating the effect of these commands on the environment. Such movements are not merely responses to stimuli, but quickly evolve into goal-directed behaviors that involve perception, motivation and planning (Hofsten, 2004). While it is well known that infants must learn to crawl before they can learn to walk (Adolph et al., 1998), until recently the developmental trajectory of infant reaching has been less well known, and discovering the control strategies employed to successfully articulate reaches remains an active area of research.

Research in these control strategies lags behind the development of biomechanical models of human arms and the optimal controllers used to guide them. For example, Kamen (2008) recently created an 18 degree of freedom (DoF) model of a human arm. This arm can be attached to the shoulder of amputees and connected to remaining neural tissue, giving the user a replacement that is, in some respects, better than the original. What remains unaccounted for is the development of hierarchical control strategies used by the central nervous system (CNS) to orchestrate the execution of top-down goals, modulated by perceptual feedback in real time.

It is known that at birth, neurons in the spinal cord and brain stem already contain sophisticated central pattern generators, useful synergies and postures. An infant that is held with their legs in warm water will vigorously attempt to walk, despite never having done so before (Thelen et al., 2002). A spinalized cat is able to run on a treadmill, and

when the treadmill gets faster or slower, the cat adjusts accordingly with no top-down influence (Lovel et al., 1990). These inborn elements of movement, combined with intrinsic drives, curiosity and a nurturing environment during development gradually lead to intelligence more generally. As Wolpert (2005) is quick to remind, the progression from babbling infant movements to sophisticated goal directed behaviors (such as typing) is the entire point of having a brain, allowing one to move the senses to the location of interesting information.

Despite the best efforts of decades of research, attempts to create a working system that rival the brain's generalized intelligence have been unsuccessful. The emerging field of developmental robotics hopes to change that by including the key factor that previous models of the sensorimotor loop have left out - that of development. Embodied models of the development of cognition, particularly motor control in human infants, stand to change the landscape of learning in robotics models. They promise to give researchers a principled approach to studying the developmental trajectory of intelligence in humans, allowing them to create robots that are better able to interact with us and learn a wider variety of tasks in a more natural setting, progressing from simple to complex not only in task difficulty, but also capability. Research in this field will benefit greatly from key discoveries in recent years. These include Graziano (2006), who discovered the existence of behavioral repertoires in motor cortex (M1), greatly expanding, for the first time since Penfield's experiments in the 1930s, our knowledge of its encodings. Yang et al. (2007) clarified our knowledge of the longstanding degrees of freedom (DoF) problem, demonstrating that excess DoF are not always seen as problematic to the nervous system; quite the opposite, in fact. Finally, Berthier and Keen (2006) used Electronic Motion Analysis (EMA) to provide modellers with a set of behavioral benchmarks against which they may evaluate developmental robotics models of infant reaching.

A prominent example that is representative of the problems that have been recently

solved in motor control research is the longstanding ambiguity as to what is represented in M1. Studying M1 has been extremely difficult due to the nature of the representations found there. The standard method of using recording electrodes to correlate firing rates with variables indicative of a particular hypothesis, such as position and velocity, has led to an ambiguous series of “positive” results. For example, Todorov (2000) cites a dozen studies, each looking for correlations between different variables, and each finding such correlations. These include “arm position, acceleration, movement preparation, target position, distance to target, overall trajectory, muscle coactivation, serial order, visual target position and joint configuration.” While recording studies seek to disambiguate, in M1 the opposite occurred, leading Thach (1978) to remark that, “all types of neurons that were looked for were found, in nearly equal numbers.”

Knowledge of the workings of M1 would have in turn led to reliable metrics against which to evaluate models of reaching. Instead, it generated a multitude of competing hypotheses and models. For example, Flash and Hogan (1985) suggest that M1 seeks to minimize jerk in order to create smooth movements, whereas Uno et al. (1989) hypothesize that the goal is to minimize torque, which further minimizes energy use. Harris and Wolpert (1998) conjecture that the goal is to maximize the precision of movements by accounting for the fact that larger movements are more variable due to increased noise in motor neurons. Due to the variety of correlations found in M1, researchers have had little difficulty in finding experimental confirmation for competing and often incompatible theories.

An alternate method of discovering important control variables is to perform a detailed behavioral analysis, as done by Kamm (1994), and then perform a factor analysis in order to discover the most heavily loaded factors. This approach, too, has met with limited success, although this time it has been due to inconsistent research paradigms. For example, Hofsten (1991) found that the number of peaks in the velocity profiles of infant

reaches decreases with age, and Fethers and Todd (1987) found it to be constant. A decrease would imply that infants initially produce several overlapping sub-movements, but eventually learn to produce a single, continuous movement. Halverson (1933) suggests that as infants become more skilled at reaching, they also begin to reach more quickly, a theory that was supposedly verified by Konczak et al. (1995). All of these findings have in common that they were found to be false by Berthier and Keen (2006), who used EMA to perform a rigorously controlled factor analysis of movement variables. In total, there have been four studies over a total of 30 subjects using this technique, but already it has revolutionized our understanding of infant reaching strategies, as the consistent methodology allows for conclusions to be drawn from the pooled data.

Here, I will focus on new insights into infant reaching strategies, such as how they cope with an excess of degrees of freedom, followed by recent findings regarding the encodings of motor cortex. I will then introduce several newly discovered principles of the development of infant reaching, followed by a discussion of how the synergy of these findings allows for a more principled approach to creating behaviorally and biologically plausible models and the learning principles that guide them.

Synergies and the degrees of freedom “problem”

Bernstein (1967), often considered the father of motor control, suggested that first and foremost, the CNS must learn how to cope with an excess of DoF. Models in developmental robotics must pass the tacit litmus test of solving Bernstein’s problem, which can be simply put: Given a task, such as reaching out to grab a banana, how should one configure their arm in order to position their hand, considering that they have more joints and muscles than necessary? This particular instantiation is also known as the inverse kinematics problem, as one must compute the necessary joint angles and torques needed to produce the specified hand location. Unlike the forward kinematics problem,

where the location of the hand is computed given the torques and joint angles, the former is ill-posed in that the solution is not unique. Even in a simple two DoF arm there are two possible solutions to every hand location, which can be obtained by rotating the elbow by 180 degrees. With 28 DoF in each arm, humans are faced with a much larger computational challenge.

Despite this difficulty, infants older than 16 weeks of age solve the DoF problem with apparant ease. Bernstein went on to revive the notion of synergies (Babinski, 1899) as a potential solution to the DoF problem. According to Bernstein, a synergy is a strategy that enables an actor to constrain the movement search space by restricting DoF that are not required (Schlesinger, 2000). This notion of synergies is similar to that of chunking in the cognitive literature (Miller, 1956), whereby memory, such as for a phone number, is facilitated by combining an arbitrary string of digits into three smaller chunks. By reducing the number of items to be remembered from nine to three, six DoF have been eliminated, thus improving the precision and recall of memory. This is readily observed in reaching infants, who partly lock their elbows until around six months of age in order to straighten their arm, preferring instead to reach using the shoulder and torso. The locking behavior emerges as infants learn to inhibit inborn reflexes which codify goal-invariant reaches and elbow movements. The immediate benefit is that they no longer have to solve the inverse kinematics problem because they have (literally) removed the elbow from the equation. This stiffening comes at the cost of increased energy expense, however, and by intention limits the complexity of the reaches that can be executed. These limitations provide a natural drive for slowly learning to apply control to the elbow rather than keeping it stiff, and after 2 years reaching converges to the skill level of adults.

While this account of synergies as the solution to the DoF problem has been dominant in the field for over 40 years, Yang et al. (2007) suggest a new interpretation. The traditional interpretation hinges on the idea that the CNS views excess DoF as

problematic and seeks, via stiffening and locking, to limit them. In a break with tradition, Yang et al. demonstrate that the CNS sometimes views excess DoF as a luxury, allowing it to focus attention on important control variables, such as hand location, ultimately increasing stability. They cite as evidence Bernstein, who monitored the motions of blacksmiths swinging their hammers, something they have done hundreds of times a day for years on end, and noted that individual joints varied more than the final location of the hammer. Bernstein interpreted this as meaning that the brain uses the apparent redundancy in the excess joints to produce less variable movements. More variability in joints somewhat counter-intuitively results in less variability in the end effector. Joints are not randomly perturbing, but are instead correcting for each other. They further elucidated this principle of sharing by having subjects repetitively produce 20 newtons of force in a finger tapping task while measuring the potentials produced by the muscles. It was found that if one finger produced 15 newtons of force, the other would produce 5 in order to compensate.

Representations in motor cortex

The challenge of motor cortex has been to find structure in what seemed to be a neural soup of information related to the physics of reaching. In an effort to unravel these relationships Graziano et al. (2005) connected a voltage source to the electrodes planted in the M1 of monkeys, artificially stimulating the neurons. Such stimulation experiments have been conducted in the past, such as by Penfield (1937), and produced minor twitches throughout the body.

In contrast to previous methodology, Graziano et al. (2005) left the electrodes in for successively longer periods of time. At 500 milliseconds, behaviorally meaningful movements began to emerge, such as complete reaches followed by prehension. The types of behaviors elicited were further found to be encoded differentially into subregions,

indicating the presence of a topographic map based on categories of behavior. These subregions seemed to be based on movements critical to survival and the frequency they were performed, such as eating, climbing, defensive postures and reaching. Intriguingly, as the stimulation duration was increased up to 4 seconds, the complexity of the behavior increased as well. The authors found this to be an indication that the stimulated area of M1 was not encoding for the entire repertoire of behavior by itself, but rather was implicated in such movements in virtue of its connections to other areas in motor cortex, lateral cortical connections, and connections to sub-cortical regions, spinal cord, and alpha motor neurons. This general method was verified by stimulating primary visual cortex, whose mappings are well known, and finding that the activation spread in a manner consistent with the known connectivity.

In a follow up observational study of monkey behavior, Graziano found that, similar to observations of the use of synergies in human infants by Schlesinger et al. (2000), monkeys tended to assume an arm posture and then work to stabilize it while the hand and fingers performed a task such as grooming. It was also found in stimulation experiments that the movements elicited by M1 do not dynamically update based on sensory inputs from the environment. For example, if an object is placed in the trajectory of a stimulation induced reach, the monkey will repeatedly run into it, whereas in a volitional movement the obstacle is easily avoided. These two pieces of evidence together indicate that, while M1 is capable of executing arbitrary updates to movement plans, it is dependent upon higher, lower and lateral influences through spreading activation to do so. While the higher level processes initially bias the monkey towards the banana, a modified path might be taken in order to avoid unexpected perturbations in the environment, as indicated by sensory information.

As Todorov (2000) suspected, the myriad of correlations between movement variables in M1 are due to the relationship between high level motor programs and any

and all physics variables that could be used to describe them. For example, the “serial order” correlation is due to the basal ganglia making that contribution to motor control and projecting to M1 (Mushiake & Strick, 1995). The “visual target position” correlation is due to bidirectional connections between posterior parietal cortex and M1 (Desmurget et al., 1999), while “muscle coactivation” is due to proprioception via alpha motor neurons which project to M1 and “movement preparation” is due to the contributions of premotor cortex (Wise, 1985), which is adjacent to M1.

After Graziano et al. (2005) discovered behavioral repertoires which seem to limit the amount of information that M1 directly encodes for, the only question remaining is what information *does* it actually encode for? A study elucidating this question was recently carried out by Wang et al. (2007), who placed an array of nearly a thousand electrodes on a population of neurons in the proximal arm region of motor cortex. Of these neurons, it was found that 8% were tuned to the position of the hand, 32% were tuned to the velocity, 24% to position and velocity and 36% to other variables that were not looked for. Depending on the instantaneous location of the hand, these neurons dynamically changed their preferred direction and velocity. Known as an attractor network, this arrangement conveniently allows the hand to be placed in any location from any location, and allows for simpler models of M1.

The development of reaching

Until recently, models of reaching have focused on the adult motor system, in part due to a desire to disambiguate the representations in M1, but also because the available data was inconsistent. This inconsistency led to the competing hypotheses described previously, each looking to infer the objective function that is optimized during movement, and each finding supportive evidence. Clarification of the computations of M1 does not in itself solve this problem, but rather allows for the creation of simpler, more principled

models of the hierarchy in the motor system. Evidence that the nervous system eventually learns to take advantage of the excess DoF available does not help the issue either, as there is clear evidence that infants do in fact initially view excess DoF as problematic and work to solve this problem by straightening the arm via stiffening the elbow.

In order to discover the variables considered important to the infant nervous system, Berthier and Keen (2006) employed EMA, a technique whereby infants have dozens of electrodes attached to the arm and are videotaped during reaching. Rather than attempting to find evidence for any single previously stated hypothesis, Berthier and Keen (2006) collected data on all of the variables that have been used to characterize reaches, including average speed, maximum peak speed, distance from the hand to the target at the start of the reach, length of the hand path, movement duration, number of speed peaks, squared jerk, straightness ratio, hand-shoulder distance change and time of peak speed. Principle Components Analysis was used to collapse over variables that are highly correlated and discover only the most heavily loaded factors, of which there were five: speed/jerk, reach duration/number of speed peaks, duration of peak speed, and distance covered/straightness. Thus, only five variables are needed to adequately describe infant reaches.

This study also resulted in a number of findings describing the development of infant reaching; that is, how reaches progress over the first few years of life. Initially, trajectories are very curved, but over the first few years slowly become straighter. It was also found that the maximum peak speed moves closer to the beginning of the reach with age. This is interpreted as a shift in goals from one where the infant only wants to get the hand near the object, to one where the infant is interested in actually grabbing it. The previous controversy as to the rate of change in the number of speed peaks was also settled in this study, as they found them to be constant over the first three years of life. It is still not clear, however, if these “submovements” are due to uncontrolled dynamics in

the arm (Thelen et al., 1996) or are discrete motor programs (Fishbach et al., 2007).

The authors suggest that together, these findings provide a new foundation for understanding reaching movements. At the onset of reaching, it is clear that infants simply do not have the experience needed to compute the inverse kinematics of motion, yet they successfully reach. The acquired data suggest that instead, infants select actions that simultaneously result in a quick reach that is also smooth, accurate and incurs a low energetic cost. The difficulty of this problem causes infants to simplify it by stiffening the elbow. Since this incurs an added energetic cost, within a few weeks of reach onset they learn how to control it. This incremental progression continues until roughly the age of three, during which they are slowly learning the inverse kinematics of motion, and after which the overall strategy has converged to something resembling that of an adult.

Autonomous mental development

Traditionally, there have been several competing and sometimes complementary approaches to training robotics models. These include knowledge-based methods utilizing search algorithms from artificial intelligence, such as robot controllers that use A* search for motion planning (e.g., Kuffner et al., 2005); genetic algorithms, where robots are evolved through survival of the fittest in a natural selection paradigm (e.g., Davidor, 1991); and machine learning frameworks, where carefully crafted sensory data is used as input to an optimal algorithm, such as the object recognition model of Pinto et al. (2008), which uses dozens of support vector machines in a one vs. many training regime. These methods often produce impressive results, and yet they lack some of the most fundamental constituents of intelligence seen in adults, such as online speech, language, and vision.

This gap between advances in engineering and progress in the modeling of intelligence is perhaps best understood by comparing IBM's Blue Brain Project (Markram, 2006) with Honda's ASIMO humanoid (Sakagami et al., 2002). Blue Brain is

an attempt to create a 3D molecular simulation of the human brain. Due to rapid advances in computing power, the project has already successfully created models at the scale of both mouse and rat brains, and expects to create models at the scale of a human brain within the next decade. The simulations have been highly successful to date, with the electrical properties of the simulated neurons producing better experimental data than measurements on real neurons in vitro, largely due to the highly controlled environment and the ability to literally zoom in on any neuron. Honda's ASIMO is an attempt to create a humanoid robot that can interact with people in a natural way and mimic human behavior on a variety of levels. Modern versions can climb stairs, run at speeds of up to 6 km/h with a strikingly human gait, recognize familiar environments and avoid people and robots while navigating. It has already been deployed to perform simple jobs such as acting as a receptionist.

Both projects have met with success in certain areas, leaving open the question as to why IBM's brain has not been put inside of Honda's humanoid. While Blue Brain may soon simulate the 100 billion neurons in a human brain, the cyto and myeloarchitectonics of the cerebral cortex are not yet fully known to neuroscience. Thus, the simulation may include every type of neuron found in the brain down to the level of molecules, but the large scale functional connectivity within and between brain areas will be absent. Present Blue Brain models focus only on the number of neurons that exhibit waves of activation, and while they may be rat scale, they are not sufficient to control a rat, whose nervous system has had millions of years to develop a sophisticated large scale brain area organization. Likewise, while Honda's humanoid can climb stairs, it can only do so when its machine learning algorithms have been carefully primed by humans. When it has fallen during public demonstrations, the results have been devastating, as the robot doesn't attempt to catch itself and breaks into pieces.

In order to eventually overcome these limitations and to advance the state of the

art, Weng et al. (2001) suggest that the three traditional approaches to robotics modeling be augmented by a new approach called autonomous mental development. In this paradigm, a brainlike agent is born and interacts in realtime with its environment, solely under the guidance of its intrinsic developmental program. It is nurtured by and interacts with its human robot sitters, has a body that is well suited to the types of tasks it will perform and environment it will live in, and lives a long life through which it can autonomously explore, learn and develop. This approach can be clearly contrasted with traditional robots, where the engineer first develops an understanding of a task, programs a task specific representation and then runs the program on the robot. In an autonomous mental development paradigm, the first step is to design a body with biophysics suitable to the environment the robot will live in followed by designing the developmental program. Next, the robot is born, raised by humans and interacts with its environment in a manner similar to human infants. This approach represents a convergence of methods, which includes both the bottom-up style of Blue Brain, where the individual neurons are simulated in mass, and also the top-down style of ASIMO, where human behavior is mimicked as precisely as possible. However, it also includes a third approach which has been largely missing in robotics models, where the robot is set off on a developmental trajectory, and rather than hardcoding the brain and body separately, they develop together naturally over the robot's lifetime, a product of both nature and the nurturing guidance of the robot sitters.

Several attempts have been made to create robots that undergo autonomous mental development, one of the first being SAIL (Weng et al., 1999). Although SAIL's body is not capable of developing, it is capable of automatically interacting with objects in its environment and being trained by its robot sitter via "good" and "bad" buttons which are attached to its body. The robot can be taken for a walk, automatically discerns discriminative features of the environment, and can then walk itself through similar

situations without aid. After toys are shown to the robot and its hand is guided towards them, it can interact with them in a play-like fashion in the future. The robot acts in realtime by automatically self-organizing discriminative features of interesting aspects of objects in a coarse-to-fine manner. Weng *et al.* claim that several of these capabilities have never been seen in robots before, giving credit to the autonomous mental development approach.

Another such robot was created by Sony engineers in what is known as The Playground Experiment (Oudeyer et al., 2005). Here, a small robotic dog was put in a play pen with a variety of toys and was allowed to interact with them. According to the researchers, the dog's developmental program was loosely modelled off of the intrinsic motivational systems seen in humans, allowing it to select tasks for itself which are not too predictable, but also not too unpredictable. At first, the dog would tend towards situations which were easy to learn, and was able to gradually increase in capability, all the while avoiding tasks in which nothing could be learned. The researchers were able to clearly demonstrate fully autonomous learning.

While both of these robots are groundbreaking, it comes at the cost of fully neglecting biological realism. For example, in The Playground Experiment, the dog stored every environment he had ever encountered as vectors. For each of these vectors, a machine learning algorithm such as Random Forests was trained in a one vs. all paradigm. For every task every random forest was called to predict an action and the predicted error was recorded. The action that demonstrated the greatest potential for learning was then executed. If an action was performed in an environment, the environment and action vectors were associated such that they could be ignored in the future, as they would then present little to no potential for learning.

Methods

Creating a robot that incorporates all of the ideas presented herein is the challenge facing the field of developmental robotics today. The present goal is to create a developmental robotics studio, an image of which can be found in Figure 1 on page 30, that allows for the modeling of the computational principles of infant reaching to proceed in much the same way as infant reaching itself does; that is, in a gradual fashion that slowly increases in complexity and, as a result, capability.

To this end, the Emergent Neural Network Simulator was chosen. Emergent is designed for creating complex, biologically plausible models of the brain and is thus aptly suited for the implementation of areas such as motor cortex. It has a 3D interface and an accurate simulation of rigid body physics via the Open Dynamics Engine, allowing objects to be precisely simulated and represented visually. This is critical for robotics models, in which an intuitive behavioral analysis of a reach is often just as revealing as the statistics that describe it. It also allows one to bypass a common difficulty in developmental robotics, that of creating robots whose morphology changes over time in order to accommodate the developmental program. In Emergent, unlike traditional robots, the arm can quickly and easily be programmed to grow over time at no cost to the researcher. Further, once a robot has first been prototyped in simulation, it can then be demonstrated in hardware.

In the following sections I will introduce the robotics studio, starting with the network architecture. In addition to explaining the motivations behind each layer and its associated projections, I will also discuss the biologically plausible learning mechanisms used and the parameters of the body, including its ability to quickly and dynamically change in morphology.

Network architecture

The network, listed in Figure 2 on page 31, is composed of a hidden layer which receives proprioceptive input concerning muscle length and rate of change of muscle length from the shoulder and elbow input layers. The hidden layer also receives visual input of the present hand and goal location from those respective layers. The hidden layer then projects to the elbow and shoulder output layers, which are decoded to produce torques on those joints. The hidden layer must simultaneously predict where the hand will end up at the end of the trajectory in the hand goal out layer.

Motor force layers.

The representations in the shoulder and elbow input and output layers have been modeled based on motor cortex. As Wang et al. (2007) found, the majority of neurons in the reach region of motor cortex encode for position, velocity or a combination of the two in a population vector and the tuning properties of any given neuron change dynamically with the instantaneous properties of the trajectory. In the specification for these layers, position (muscle length) is represented along the x-axis and velocity (rate of change of muscle length) on the y-axis. The units are topographically organized into groups based on their position and velocity tunings, which are decoded as a gaussian average into the force output. Although the simplified task given to the robot in this work has his arm start at the origin for each reach, this representation is sufficient to allow him to get to any position from any position. Because this same layer specification is used for both input and output, it also simulates proprioceptive feedback, which the robot receives throughout each reach.

Forward and inverse models.

Forward and inverse models represent mappings from one set of information to another and are two different ways of framing the computations performed in a reach

(Buneo, 2002; Jordan & Rumelhart, 1992). In a forward model the angular positions and velocities of the shoulder and elbow are known and the robot must compute the final hand position. In an inverse model only the hand position is known and the robot must compute the torques needed to produce a trajectory that positions the hand on the target. Inverse models are inherently ill-posed and give rise to the degrees of freedom “problem”. For any given goal position, there are millions of joint configurations that can place the hand there, and the nervous system must constrain the problem such that it only selects one.

Both forward and inverse models have been implemented in the current network architecture. At the beginning of a reach the Hidden layer receives input on the current status of his shoulder, elbow and hand from the Shoulder_In, Elbow_In and HandPos_In layers in addition to the location of the goal from the HandGoal_In layer. In order to solve the inverse model, the Hidden layer must activate the Shoulder_Out and Elbow_Out layers such that the decoded torques specify a trajectory that lands the hand on the goal. The forward model is solved by predicting this final location in the HandPos_Out layer, which, when reaches are successful, would ideally be identical to the HandGoal_In layer.

The Primary Value Learned Value reinforcement learning system.

Unlike traditional neural network models that utilize error driven learning, which is a full vector signal, this model uses the Primary Value Learned Value (PVLV - pronounced “pavlov”) reinforcement learning system, which provides only a scalar learning signal (O’Reilly et al., 2007). PVLV is able to replicate much of the Pavlovian conditioning data and is based on the detailed neuroanatomy of dopaminergic neurons in the midbrain, which fire proportionally to unexpected rewards. Due to the weaker signal, a network that uses PVLV (and any system using biologically plausible reinforcement learning) will learn much more slowly than a traditional error driven learning model.

While a detailed explanation of the PVLV algorithm is out of the scope of the present work, an understanding of the PV system will prove useful for interpreting the

results. The PV system provides a signal when there are unexpected rewards and punishments and is broken down into two subcomponents, PV_i and PV_e, for inhibitory and excitatory respectively. PV_e is a model of the lateral hypothalamus which indicates primary rewards (US) and has excitatory projections onto the dopaminergic nuclei in the midbrain and regulate dopamine production. PV_i corresponds to the patch neurons in the ventral striatum which in turn have inhibitory projections onto the dopaminergic nuclei. PV_i learns to expect primary rewards just ahead of PV_e, and through a simple neural subtraction procedure, it turns out that $Dopamine \propto PV_e - PV_i$. In other words, PV_i learns to expect when a US will occur and works to cancel it out by inhibiting dopamine firing. This further leads to weight changes in the network through the relationship $\Delta w_i = (PV_e - PV_i)x_i$ for each sending unit x_i .

The PVLV layers are represented in the lower left corner of the network diagram in Figure 2 on page 31. The associated projections are complicated and difficult to discern, so the projections associated with the PV system will be described here. PV_e receives projections from the output layers, including HandPos_Out, Shoulder_Out and Elbow_Out. PV_i receives projections from Elbow_In, HandGoal_In, HandPos_In, Shoulder_In and Hidden. According to the relationship $Dopamine \propto PV_e - PV_i$, the PV_i system sees the initial state of the network and attempts to predict the excitatory signal from the PV_e layer, which sees the outputs of the network and will drive dopamine release.

PVLV relies on a well-timed reward signal, and one can imagine any number of errors upon which to base performance. At a reward value of .5, the baseline dopamine firing rate, no learning occurs, anything below that is perceived as punishment and anything above that as reward. The reward function has been kept as simple as possible. In the first task where the network has only a shoulder and humerus, the reward is based on the angular distance from the humerus to the banana as $1 - \Delta$. Similarly, in the second task where the network has a shoulder and elbow, the reward is based on the

euclidean distance from the hand to the goal as $1 - \Delta$. In both cases the reward is normalized such that it lies between 0 and 1, providing the robot with rewards and punishments that are directly proportional to performance.

Body parameters

The masses and lengths of the robot's arm are based on that of a Rhesus Macaque, as described in Colman et al. (2005) and Chan and Moran (2006). The total arm mass is 1.5 kg and the total length is .35 meters. The length of the humerus is .125 meters and the mass is .53 kg. The length of the ulna is .14 meters and the mass is .6 kg. The length of the hand is .088 meters and its mass is .37 kg. The robot's torso is not based on experimental data and is .5 meters across. In order to support future development, the robot also has a head which is equipped with two video cameras. While it would have been possible to use the Computational Cognitive Neuroscience Lab's model of the visual system in place of his gaussian "visual" inputs, that model does not currently have the disparity maps needed for depth perception.

In addition to realistic body parameters several programs have been included in the Emergent simulation that allow for dynamic scaling of his arm. One program allows for the elbow and ulna to be removed completely, and bases reward off of the angular distance of the humerus to the goal after a reach has been attempted. Another program puts them back in, and bases performance off of the euclidean distance to the goal. Yet another program allows for the dynamic scaling of his arm, simulating morphological development as learning progresses.

Results

After the robot had been constructed, it proved difficult to coax it into learning a significant reaching task using reinforcement learning. In order to test that the architecture was in fact working, the framework was further enhanced to support error

driven learning using backpropagation. Initially, the elbow was locked to prevent it from moving and the Shoulder_Out layer was provided with the correct answer after every reach attempt. The network consistently learned the task in under 100 reaches, lending credit to the overall design.

The next phase of training was to remove backpropagation learning and rely solely on the PVLV algorithm. With the elbow and ulna still removed, the network refused to learn the task, despite several months of effort on my part. Finally, I decided to embark on an exhaustive parameter search using the CCN Lab's 27 node dual-core "Beowulf" cluster, and discovered that, with the correct learning parameters, some of them never seen in a PVLV model before, the robot was able to demonstrate learning, as depicted in Figure 3 on page 32. As you can see in the figure, the robot's performance starts near 0.225 rad, or 12.9°. After nearly 3000 reaches, the robot's average performance has gone as low as .08 radians (4.6°) and peaks at .375 radians (21.4°) from the goal. While this result is somewhat promising, as it seems to show clear signs of reinforcement learning for the first time, the robot's performance oscillates wildly, even when using a 50 reach rolling average for each point. If one stares at the chart for too long, they may even begin to doubt whether learning is occurring at all.

Undaunted, I reconnected the elbow and ulna and, using parameters very similar to those discovered without them (dopamine gain = 40, PV_i learning rate = .001), the robot immediately began to learn in a clear, unambiguous fashion, as depicted in Figure 4 on page 33. With the banana placed in random locations around his body, the robot initially starts at an average reach distance of 0.2 meters away, and within 1350 reaches is within .014 meters of the banana on every single reach attempt.

This result can be clearly understood in terms of the PVLV learning mechanism. A plot of performance in terms of PV_e, PV_i, dopamine and the euclidean distance from the goal can be found in Figure 5 on page 34. As shown in the plot, the PV_e value is 1 minus

the distance to the goal. The PVi value starts at 0.6, and slowly learns to predict primary rewards, cancelling out PVe as learning progresses. In the first half of learning, the dopamine dips below 0 dozens of times, corresponding to punishments, but is receiving positive reinforcement during the majority of the learning curve. Towards the end of learning dopamine has stabilized at a baseline level of firing, indicating that the robot has become adept at predicting when he will succeed and when he will fail based on the state of the network at the start of a reach.

Discussion

Here I have reviewed recent advances in the field of developmental robotics and demonstrated the feasibility of creating a robotics studio that might eventually account for these various data points. Emergent has proven to be a capable platform for developing such a robot, allowing for flexible modification of the network, robot and tasks to be learned when failures are encountered and progress is to be made. Despite this success, many questions have been raised by this model and there is much future work to be done.

One of the primary questions to be answered is whether or not the present model is able to recreate the reaching factors shown to be important to the infant CNS by Berthier and Keen (2006). These include verifying that the model obeys the minimum jerk hypothesis, has an appropriate number of peaks in the velocity profile and works towards producing a straight reach when possible, among others. Although I have been unable to investigate them in the present work, much of this data has already been made available in the simulator and waits to be analyzed. A key missing data point is the jerk of the movement. As the Open Dynamics Engine does not automatically compute it, it will be necessary to compute the discrete approximation and include that data in the studio.

Another question relates to the prediction of Yang et al. (2007) that the CNS eventually learns to take advantage of the excess DoF available, combined with existing

data that shows the infant CNS finds those same DoF to be problematic. If this is the case, a properly designed robot would display a U-shaped learning curve, where the innate ability to perform goal-invariant reaches is initially suppressed, and then slowly controlled as development progresses. While the present model displays a generally descending learning curve, it is encouraging that the model performs better with a shoulder and elbow than with the shoulder alone. A cursory analysis of the force and velocity profiles of those joints shows that they are clearly working together to produce reaches. Further, the model learns approximately 10 times faster when the shoulder and elbow are connected as opposed to the shoulder alone, which may eventually provide evidence for this hypothesis.

One area in which the present robot has excelled is in learning to control the Shoulder_Out and Elbow_Out layers using reinforcement learning, which were designed with the workings of M1 in mind. An analysis of the weights projecting from the Hidden layer to these layers shows that the robot has learned mappings that are akin to synergies. When the goal is placed in a location that is near a previously successful reach, the pattern of activations is largely similar, with units that have successful mappings showing crisp activations, and the rest of the units having activations that are nearly zero. In order to account for variance in the goal position, the activations of the units comprising the synergy change accordingly.

Finally, this robot is a first in terms of using a biologically plausible reinforcement learning mechanism, a feature that will be key to future versions that learn using autonomous mental development. Unlike the model of Oudeyer et al. (2005), which used an extremely artificial learning paradigm in a machine learning framework, this robot is able to learn to predict the presence of a primary reward based on the status of the environment. It learns to predict when it will succeed and when it will fail, providing a solid foundation for a future version that uses this knowledge to select its own tasks based on their potential for learning.

References

- Adolph, K. E., Vereijken, B., & Denny, M. A. (1998). Learning to crawl. *Child Development*, 69(5), 1299–1312.
- Babinski, J. (1899). De lasynergie cerebelleuse. *Rev Neurol*, 7, 806–816.
- Bernstein, N. (1967). *The co-ordination and regulation of movements*. Oxford, UK: Pergamo.
- Berthier, & Keen. (2006, March). Development of reaching in infancy. *Experimental Brain Research*, 169(4), 507–518.
- Buneo, C. A. (2002). The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia*.
- Chan, S. S., & Moran, D. W. (2006, December). Computational model of a primate arm: from hand position to joint angles, joint torques and muscle forces. *J Neural Eng*, 3(4), 327–337.
- Colman, R. J., Mckiernan, S. H., Aiken, J. M., & Weindruch, R. (2005). Muscle mass loss in Rhesus monkeys: Age of onset. *Experimental Gerontology*, 40(7), 573–581.
- Davidor, Y. (1991). *Genetic Algorithms and Robotics: A Heuristic Strategy for Optimization*. World Scientific.
- Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (1999, June). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nat Neurosci*, 2(6), 563–567.
- Fetters, L., & Todd, J. (1987, June). Quantitative assessment of infant reaching movements. *J Mot Behav*, 19(2), 147–166.
- Fishbach, Alon, Roy, Stephane, Bastianen, Christina, et al. (2007, February). Deciding when and how to correct a movement: discrete submovements as a decision making process. *Experimental Brain Research*, 177(1), 45–63.
- Flash, T., & Hogan, N. (1985, July). The coordination of arm movements: an

- experimentally confirmed mathematical model. *J Neurosci*, 5(7), 1688–1703.
- Graziano. (2006, March). The organization of behavioral repertoire in motor cortex. *Annu Rev Neurosci*.
- Graziano, Aflalo, & Cooke. (2005, December). Arm movements evoked by electrical stimulation in the motor cortex of monkeys. *J Neurophysiol*, 94(6), 4209–4223.
- Halverson, H. M. (1933). The acquisition of skill in infancy. *Journal of Genetic Psychology*, 43, 3–48.
- Harris, & Wolpert. (1998, August). Signal-dependent noise determines motor planning. *Nature*, 394(6695), 780–784.
- Hofsten, C. V. (1991). Structuring of early reaching movements: a longitudinal study. *J Mot Behav*, 23(4), 280–92.
- Hofsten, C. V. (2004, June). An action perspective on motor development. *TRENDS in Cognitive Sciences*, 8(6).
- Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16, 307–354.
- Kamen, D. (2008, February). *Ieee spectrum: "luke arm" prosthesis readies for clinical trials*. Internet.
- Kamm, K. (1994). *The influence of postural development on the development of skilled reaching*. Unpublished doctoral dissertation, Indiana University.
- Konczak, J., Borutta, M., Topka, H., & Dichgans, J. (1995). The development of goal-directed reaching in infants: hand trajectory formation and joint torque control. *Experimental Brain Research*, 106(1), 156–168.
- Kuffner, J., Nishiwaki, K., Kagami, S., Inaba, M., & Inoue, H. (2005). Motion Planning for Humanoid Robots. *Robotics Research: The Eleventh International Symposium*.
- Lovel, Y., Grego, R., Ro, Y., & Edgerto, N. (1990). Weight-bearing hindlimb stepping in treadmill-exercised adult spinal cats. *Brain research*, 514(2), 206–218.

- Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, 7(2), 153–160.
- Miller, G. (1956). *The magical number seven, plus or minus two: some limits on our capacity for processing information*. Indiana : Bobbs-Merrill.
- Mushiake, H., & Strick, P. L. (1995, December). Pallidal neuron activity during sequential arm movements. *J Neurophysiol*, 74(6), 2754–2758.
- O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2007, February). Pvlv: the primary value and learned value pavlovian learning algorithm. *Behav Neurosci*, 121(1), 31–49.
- Oudeyer, P. Y., Kaplan, F., Hafner, V. V., & Whyte, A. (2005). The Playground Experiment: Task-Independent Development of a Curious Robot. *Proceedings of the AAAI Spring Symposium on Developmental Robotics*, 42–47.
- Pinto, N., Cox, D. D., & Dicarlo, J. J. (2008, January). Why is real-world visual object recognition hard? *PLoS Computational Biology*, 4(1), e27+.
- Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., & Fujimura, K. (2002). The intelligent asimo: system overview and integration. In *Intelligent robots and system, 2002. iee/rsj international conference on* (Vol. 3, pp. 2478–2483 vol.3).
- Schlesinger, M. (2000, March). Learning to reach by constraining the movement search space. 67–80.
- Schlesinger, M., Parisi, D., & Langer, J. (2000). Learning to reach by constraining the movement search space. *Developmental Science*, 3(1), 67–80.
- Thach, W. T. (1978, May). Correlation of neural discharge with pattern and force of muscular activity, joint position, and direction of intended next movement in motor cortex and cerebellum. *J Neurophysiol*, 41(3), 654–676.
- Thelen, E., Corbetta, D., & Spencer, J. P. (1996, October). Development of reaching during the first year: role of movement speed. *J Exp Psychol Hum Percept Perform*, 22(5), 1059–1076.

- Thelen, E., Fisher, D. M., & Ridley-Johnson, R. (2002). The relationship between physical growth and a newborn reflex. *Infant Behavior and Development*, 25(1), 72–85.
- Todorov. (2000, April). Direct cortical control of muscle activation in voluntary arm movements: a model. *Nat Neurosci*, 3(4), 391–398.
- Uno, Y., Kawato, M., & Suzuki, R. (1989, June). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61(2), 89–101.
- Wang, W., Chan, S. S., Heldman, D. A., & Moran, D. W. (2007, June). Motor cortical representation of position and velocity during reaching. *J Neurophysiol*, 97(6), 4258–4270.
- Weng, J., Hwang, W. S., Zhang, Y., & Evans, C. (1999). Developmental robots: Theory, method and experimental results. *Proc. 2nd International Conference on Humanoid Robots*, 57–64.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., et al. (2001, January). Artificial intelligence. autonomous mental development by robots and animals. *Science*, 291(5504), 599–600.
- Wise. (1985). The primate premotor cortex: past, present, and preparatory. *Annual review of neuroscience*, 8, 1–19.
- Wolpert, D. (2005, December). *The puppet master - how the brain controls the body*. Francis Crick Prize Lecture. London, UK.
- Yang, J.-F., Scholz, J., & Latash, M. (2007, January). The role of kinematic redundancy in adaptation of reaching. 54–69.

Author Note

Thanks to my advisor, Dr. Randy O'Reilly, whose patience while I learned to use the Emergent simulator was endless. Without his advice this work would not have been possible. Thanks also to Dr. Harrison Carpenter for commenting on several drafts of the introduction. Thanks also go to my defense committee, consisting of Dr. Jerry Rudy, Dr. Matthew Keller, and Dr. David Sherwood, who were all willing to work with me on short notice despite their busy schedules.

Figure Captions

Figure 1. A screenshot of the robotics studio. The plot on the lower left displays the elbow velocity vs. the shoulder velocity, while the next two plots display the force and velocity of those respective joints over time. This is followed by the network diagram, a more detailed explanation of which can be found in Figure 2 on page 31. The top right plot displays the trajectory of the hand for an individual reach, and the plot below it displays the robot's reach performance. The robot itself sits in the foreground, with the banana that it is to reach for positioned in front of him.

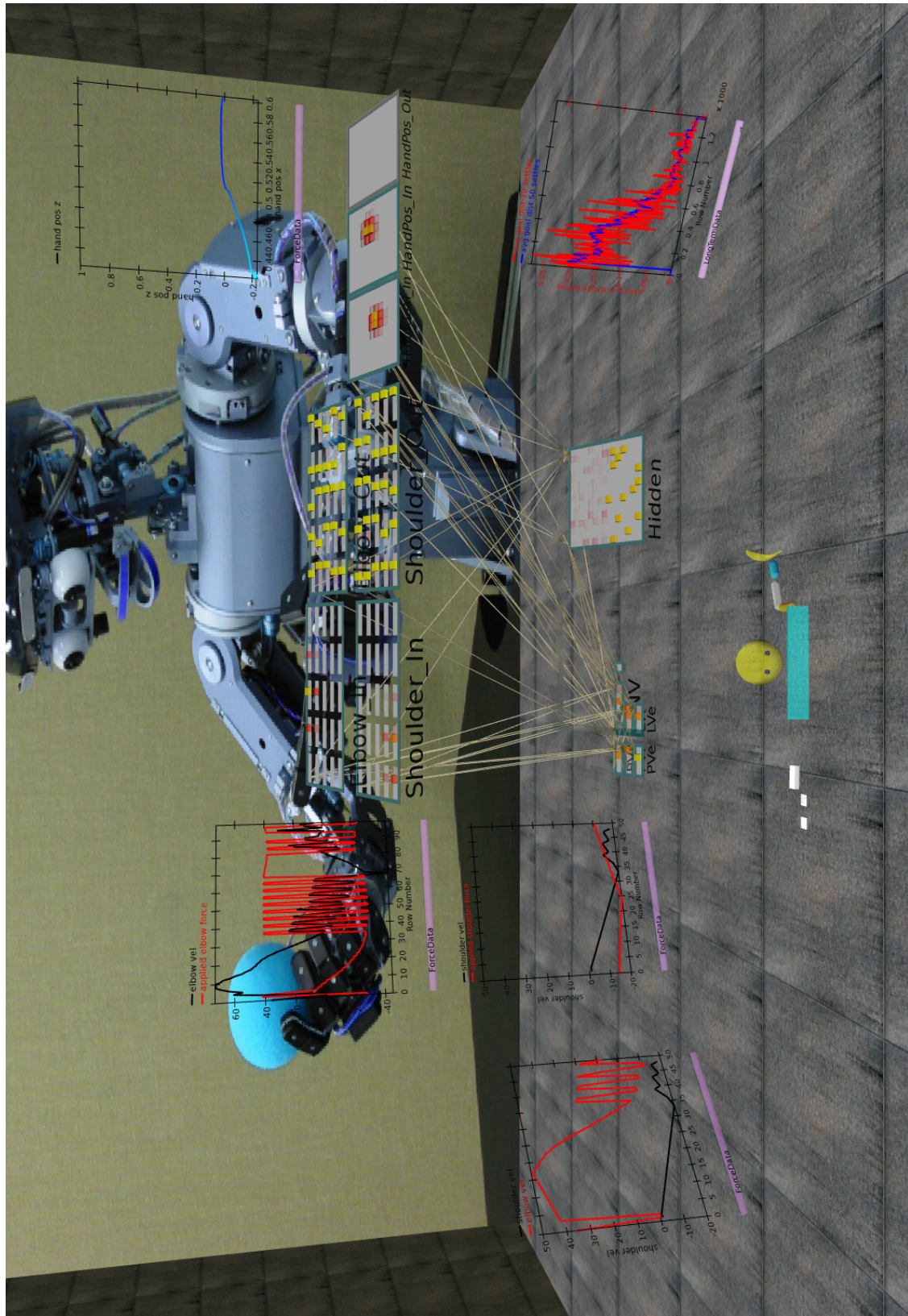
Figure 2. A diagram of the neural network. The Shoulder_In and Elbow_In layers provide information about muscle length and rate of change of muscle length to the hidden layer, which in turn projects to the Shoulder_Out and Elbow_Out layers, which are decoded to produce torques on those respective joints. The hidden layer also receives visual input concerning the current location of the hand from HandPos_In, and the current location of the goal from HandGoal_In.

Figure 3. Reach performance using only the shoulder

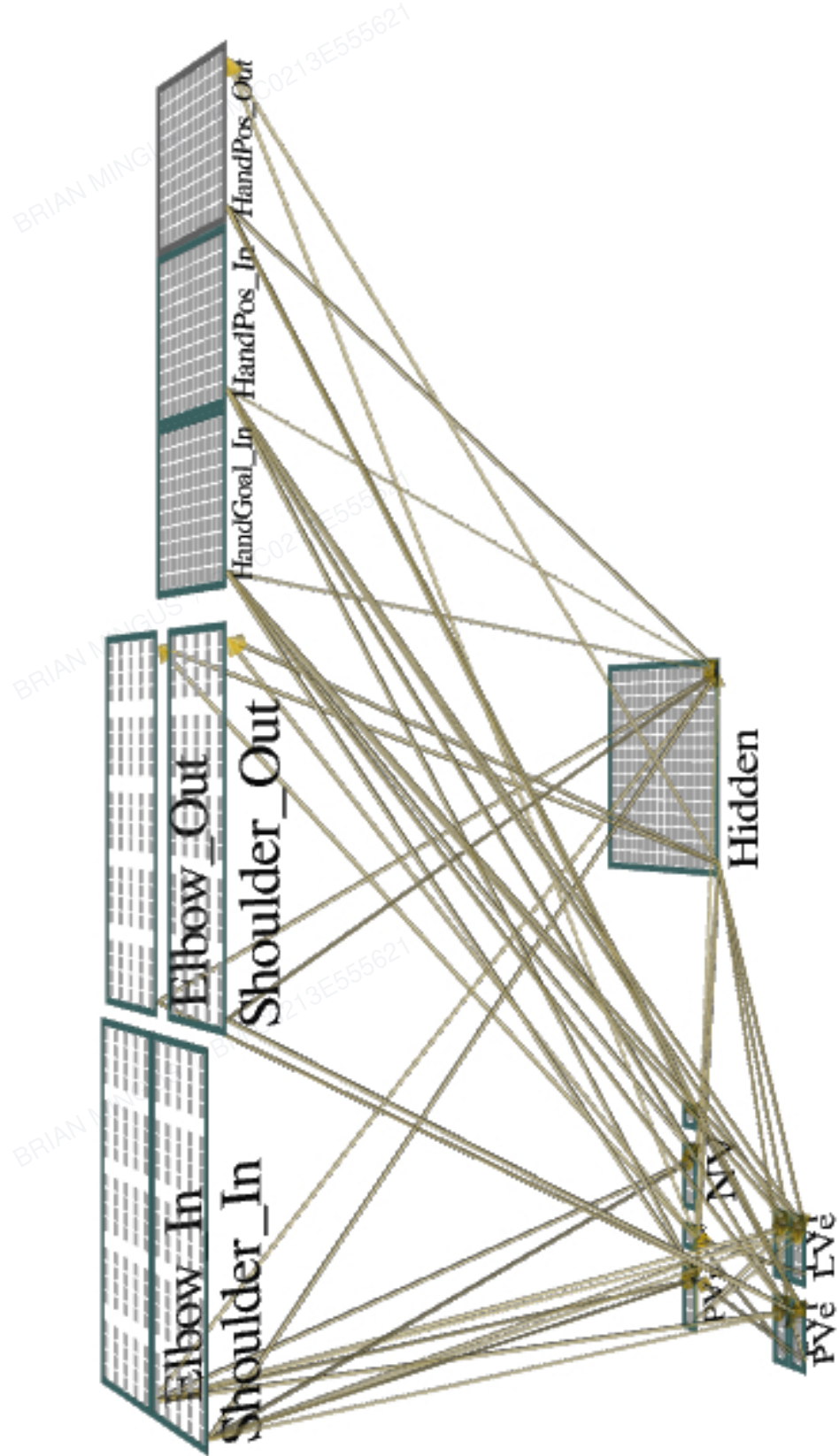
Figure 4. Reach performance using both the shoulder and elbow

Figure 5. Chart of PVLV values during learning with the shoulder and elbow

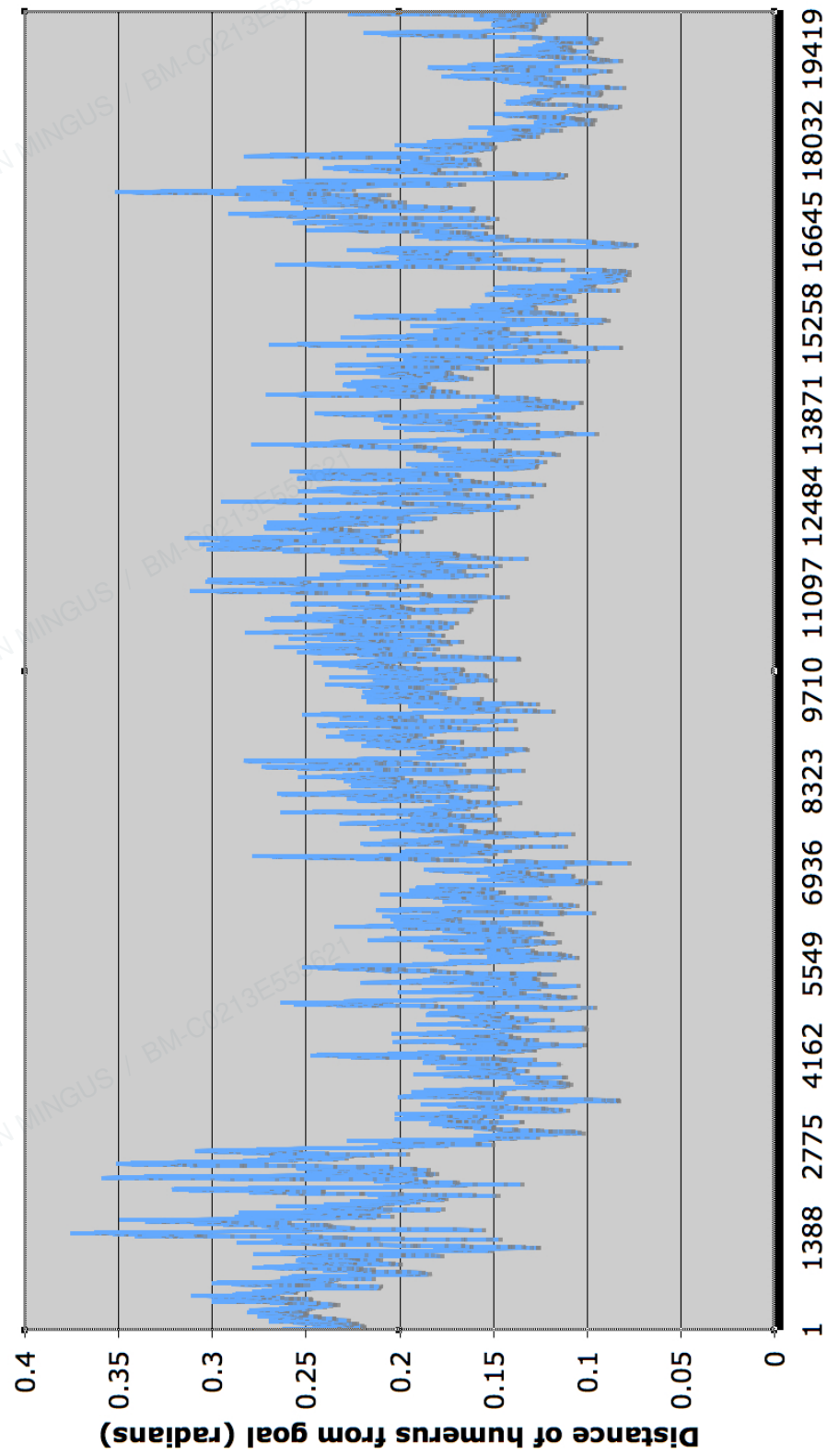
Reaching in Developmental Robotics, Figure 1



Reaching in Developmental Robotics, Figure 2

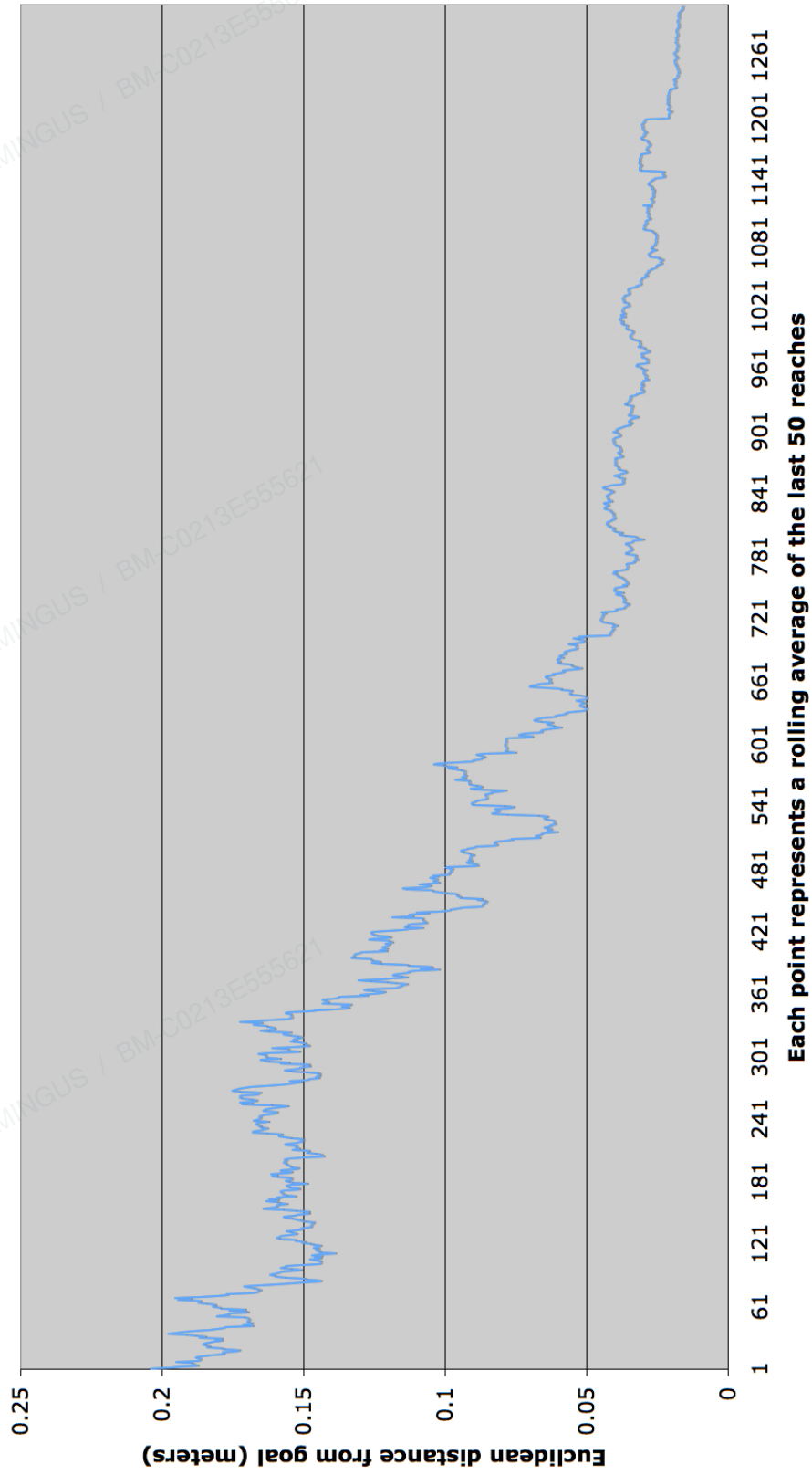


Reach performance with shoulder only over 20000 trials



Each point represents the rolling average over the last 50 reaches

Reach performance with shoulder and elbow over 1300 trials



Learning in the primary value system

