A Cognitive Affordance Model of Shoulder Rotation in ACT-R 3D.

by

Sterling Somers

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Cognitive Science

Carleton University
Ottawa, Ontario

Abstract

Presented is an implementation of a time sychronous middleware for Python ACT-R and the open-source robotics simulator, MORSE (Echeverria et al., 2012; Echeverria, Lassabe, Degroote, & Lemaignan, 2011), a novel vision system, and novel motor system, which I collectively call ACT-R 3D. A new 3D camera and a crude body-model robot was added to the MORSE system to facilitate modeling of afforance-based research on aperture passage (walking through apertures and rotating shoulders as needed).

Presented as a proof of concept are three affordance models based in ACT-R. The models tests a novel theory, the Theory of Geometric Affordances, that proposes that humans make geometric comparisons between apertures (depth, width, height) and stored representations of body postures (body schema). Both models are individually qualitatively compared against human performance for overall shoulder rotation while walking through apertures of various widths (Warren & Whang, 1987; Higuchi, Seya, & Imanaka, 2012) and overall safety margin while passing through apertures (Higuchi et al., 2012). The second model (Model 2) shows the best performance, with the same model exhibiting rotation similar to human performance across both experiments.

Model 2 supports the conclusion that an abstract geometric comparison mechanism is sufficient to support aperture passage judgment without the use of semantically-laden labels. This is the first known affordance model, modeled in a computational cognitive architecture, to match preliminary human performance data.

Acknowledgments

Instead of ending with, "last but not least", I'm going to start with my supervisor, Dr. Jim Davies. If it wasn't for Jim's willingness and enthusiasm to take on undergrads, I probably never would have been in the department at all. What started as a project shortly after my undergrad, turned into a working relationship that, as it happens, would result in a doctoral thesis. I am a little surprised, in hindsight, that I never once thought of quitting. But with weekly personal meetings and lab meetings, the whole experience was just one project after another and then at some point I was working on a thesis. And here it is.

Second is family. I'm not going to name them all because there are a lot of 'em. But parents, in-laws (which includes sisters and brother), and my three sisters, one of whom was my personal editor for the last seven years; have always been supportive, razzed me when needed, and inflated my ego (and helped me make English). Now they must call me doctor. Also, thanks to the psychic that said one of my mother's children would be a doctor. If it wasn't for you, I'd have been a philosophizer. And, of course, Jennifer. This whole thing took a little longer than anticipated... but the whole time we've been waiting to see what happens after I graduate as if that was a given. Thanks for the support throughout, I'm anxious to see what comes next.

Finally to the folks at Carleton: friends, committee members, administration, people responsible for finding me money, etc. I've made friends along the way, people who were in a similar boat, without whom this would have been dismal. I've had great support and had good supervision, especially from Dr. Robert West, who may seem like he's bumbling around the lab, but those hours in the hallway, and goofy sounding projects have led to an interesting set of papers.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Problem

In their classical aperture-passage research, Warren and Whang (1987) showed that there is a critical aperture ratio at which we decide that we have to rotate our shoulders. Their experiments showed that if the aperture width is in a ratio of approximately 1.3 or less with our shoulder width, then we judge that we cannot pass through those apertures without shoulder rotation. Since Warren and Whang's initial study into the passage of apertures, there have been a fair number of studies that explore aperture passage under a variety of conditions such as carrying bars through apertures (Higuchi et al., 2012; Higuchi, Cinelli, Greig, & Patla, 2006; Wagman & Taylor, 2005), walking through apertures in groups (Chang, Wade, & Stoffregen, 2009), passing through an aperture using a wheelchair (Higuchi et al., 2006), a variety of conditions for ducking through apertures (Stefanucci & Geuss, 2010), and varying visual properties in virtual environments (Fath & Fajen, 2011).

All the known research in aperture passage is tied, to different degrees of commitment, to the theoretical framework developed in Warren and Whang's (1987) original paper. The main aim of Warren and Whang's research was to provide evidence in support of *direct perception*, a notion at the centre of Gibson's theory of affordances, the theoretical drive behind his ecological approach (Gibson, 1986) to psychology. The main theoretical commitment of direct perception is that without inference based on internal representations, the perceptual system is able to extract sufficient information, directly from the environment, to make action decisions. In the case of aperture passage, Warren and Whang aim to show that we perceive apertures in units of body-width. Specifically, we perceive aperture widths as ratios of body widths such that a ratio greater than 1.0 signifies passability. They propose that this ratio (A/S ratio) is sufficient to modulate shoulder rotation, therefore not requiring any representation of the environment or of the body.

Although much of the research since Warren and Whang (1987) is viewed as an extension of their work, none of the evidence excludes a representational (non-direct) interpretation. Fath and Fajen (2011), for example, show that there are multiple visual features that contribute to aperture passage. In a series of experiments in a virtual environment, they implicate both head-sway and stride length while walking, in modulating shoulder rotation. While one interpretation is that the different visual features (A/S ratio, head-sway, stride length) all weigh-in, directly contributing to shoulder rotation modulation; a second interpretation is that all the visual features contribute to a representation of the environment that is then used to modulate rotation. Higuchi, Seya, and Imanaka (2012) show evidence for a process that produces a constant safety margin between the shoulder and the edge of an aperture for varying aperture widths. This, too, can plausibly be explained with a direct approach or with a non-direct approach. A direct perception approach, however, has trouble explaining aperture passage judgments when participants are grouped in dyad (Chang et al., 2009) because in these conditions there cannot be an A/S ratio to support the judgment since you have to take into account the person standing beside you.

That aperture passage can be explained with representations is not enough. What is aimed for in Cognitive Science is a processing description which describes the kinds of perceptual processes that create representation and the processing steps that use those representations to ultimately lead to the observed behaviour. Preferably these information processing descriptions are universal across multiple tasks. One of the approaches used to facilitate unified theories in cognitive science is the use of computational cognitive architectures that are constrained by findings across the various sub-fields in cognitive science.

One of the main advantages of computational models over verbal theories is that they must be functionally complete. That is, by necessity, in order to run, they have to have all components that are required: they must function. A perfect example of lacking completeness is actually apparent in the aperture passage research. Generally, the experiments in the aperture passage literature involve participants either walking through an aperture or making a static visual judgment of passability. Overlooked in the research, however, is the processing required to identify an aperture before ever walking towards it. This problem is immediately apparent in a computational model because there has to be some instruction to the artificial agent about what feature in the environment to judge. Without some symbolic or semantic information about, for example, the difference between a wall and the space between a wall, a naive agent will not even know what feature in the environment to make a judgment about. The first thing a participant has to be able to do is figure out where the aperture is. In order to walk through any space, the space must not be restricted in either height, width, or depth. A lot of the research into aperture passage discusses how the width might be directly perceived, one (Stefanucci & Geuss, 2010) discusses height, but none of them discuss how the depth of the aperture is judged. Assuming that there could be a direct perception explanation for perception of width, depth, and height, the agent then has to know where in 3-space the pass-through-able object is. It is not clear that a non-representational approach can offer an explanation[1]; but, regardless, no known aperture passage research even provides a theoretical account of how identifying an aperture in 3-space might be done.

Another advantage of computational cognitive modelling is that it forces a unified approach, forcing an implicit test of different theories. For example, in the present context it is unclear whether the findings of Warren and Whang (1987) that A/S ratio modulates rotation is compatible with the findings of Higuchi et al. (2012) who propose that the central nervous system attempts to achieve a constant safety margin between the shoulders and the edge of an aperture. Insofar as tasks are similar, models should predict performance measures across tasks. The theories can be implemented in a computational cognitive model and tested across conditions.

A final advantage worth discussing in the present context is that computational cognitive models can be, and often are, modular in scientifically relevant ways. Consider the problem of identifying an aperture. The simplest way to provide that information to your model is to provide the model with semantic knowledge of the environment. ACT-R models of human behaviour in 3D environments that have been built in the past, use just this approach (e.g. Best & Lebiere, 2003). A modular approach allows researchers to investigate cognitive processing for certain aspects of complex tasks, while largely ignoring other aspects. These ignored elements play a role as honest omissions that can be explored in future research.

The aim of the present work is to provide a computational cognitive affordance model of aperture passage. A task like affordance passage has a number of complex elements to it: perceptual processes, motor planning and motor control, that in turn requires some form of motor memory (even

[1]It is not, strictly speaking, inconsistent to maintain a direct perception approach and appeal to representations - although somewhat unparsimonious.

as simple as knowing which limbs to move). The resulting model is not expected to be the absolutely correct model of the task. The aim is, instead, somewhat exploratory, exploring whether a model built in a hybrid (symbolic and subsymbolic) architecture, like ACT-R, is plausible for scientific research into tasks like aperture passage.

1.1.1 ACT-R for Modeling Complex Tasks

ACT-R is a cognitive architecture used by scientists to model the mind. ACT-R is typically used to implement models of humans performing psychology experiments as a means of validating theories about the functional capabilities of the brain. The ACT-R architecture consists of several modules – meant to represent the the theorized modules of the mind (memory, perceptual system, a motor module, and a central production system) – that communicate with one another, via the central production system, through the use of buffers. These buffers facilitate communication through the use of *chunks*, slot/value pairs representing properties. ACT-R is, therefore, a symbolic architecture.

The central production system runs in a cycle of 50 (simulated) milliseconds and can ‘fire’ a production only once per cycle. This limitation helps to account for limitations in human mental processing. The central production system matches on the contents of the various buffers each production cycle, resulting in actions that could include: changing the contents of buffers, motor actions, and memory read/write requests. ACT-R also employs a sub-symbolic system to help account for human processing and memory limitations.

The 50ms production cycle, as well as other sub-symbolic equations in

ACT-R, is used to produce behavioural measures such as reaction time, error rates, and visual scanning patterns. Models built in the ACT-R architecture are validated when model predicts behavioural performance of human subjects who perform the same task, under the same conditions, as the model. Often a single model will be run under multiple experimental conditions. It is often the case that the model will be fit to a single experimental condition and then ran in new conditions without any parameter changes, in order to validate the model.

Althought ACT-R is typically used to model tightly controlled psychology experiments, in recent years, however, there has been a push to model more complex, real-world behaviour. For example, ACT-R models of driving behaviour have looked into how drivers maintain their lanes, negotiate curves, and perform lane changes (Salvucci, 2006). This research was extended to measure performance in the same categories in multi-tasking situations (Salvucci, Markley, Zuber, & Brumby, 2007). This work is particularly relevant in the present context because of its contrast with non-cognitive approaches. In (Salvucci, 2006), Salvucci cites research including an earlier work of his own (Salvucci & Gray, 2004), as an example of steering models not embedded in a cognitive architectures. Salvucci identifies three types of steering models: perception-and-action models (e.g. Fajen, Warren, Temizer, & Pack, 2003), control-theoretic models (e.g. Hess & Modjtahedzadeh, 1990), and machine learning approaches (e.g. Pomerleau & Jochem, 1996). While each of these approaches model the steering task at a computational level, none of them account for the perceptual processes that encode information in the environment, the motor responses that control the steering in response, and the cognitive processes that connect the two. Salvucci's solution is to model steering control in the ACT-R cognitive architecture within the embodied cognition, task, and artifact framework (Byrne, 2001).

The most relevant ACT-R-based system to this thesis is ACT-R Embodied (ACT-R/E) (Trafton et al., 2006; Trafton, Harrison, Fransen, & Bugajska, 2009; Trafton & Harrison, 2011; Trafton et al., 2012). ACT-R/E is, essentially, a robot controller based in ACT-R. Discussed more thoroughly in section 3.3, although ACT-R/E maintains a representation of body configuration, it is unclear whether it has an integrated motor controller as it seems to send high-level motor commands to the robot, that in-turn carries out the motor control (Harrison & Trafton, 2010). ACT-R/E also has a more advanced vision system, Specialized Egocentrically Coordinated Spaces (SECS) (Harrison & Schunn, 2003). Although impressive as a vision system, SECS does not address the need for non-semantically informed processing of space that is reminiscent in Ecological Psychology. Finally, it is unclear in the related literature whether ACT-R/E supports a memory for body postures that an agent can use as a motor plan.

This thesis introduces a new system, ACT-R 3D. The aim of the present work was to develop a simulation-time synchronized ACT-R implementation with a tightly-controlled motor module with sufficient motor control to model affordance-based research. ACT-R 3D updates Python ACT-R (T. C. Stewart & West, 2005) to Python 3, adding a simulation-time synchronized middleware to the robotics simulator, Mobile OpenRobots Simulation Engine (MORSE) (Echeverria et al., 2011, 2012). ACT-R 3D is comprised of: a Python 3 variant of Python ACT-R, a novel vision system, a novel motor control system, and a high-fidelity time-synchronized middleware for MORSE. This system was specifically designed to implement the aperture-passage affordance experiments by Warren and Whang (1987).

The vision system and the motor system, although similar at a high-level to those in ACT-R/E, are different in a number of respects. One of the key design commitments in the vision system, for example, was to minimize the semantically-laden tags used in 3D environments. One of the most compelling aspects of *direct perception* and affordances is the idea that we need not first semantically identify objects and environment features (such as doorways) to carry out actions towards them (such as walking with a shoulder rotation). This is particularly compelling in 3D simulation for two reasons: first, the practical problem of semantically labeling all objects and regions of space; and second, environmental features can sometimes function in a certain manner, without having a clear semantic label. Think of an aperture between bookshelves that can be used as a doorway but may not be labeled as such. Consider also an early inspiration for this thesis, close-quarters-combat (CQC) simulation. In modern 3D simulations solid walls can take damage and the resulting space can function as an aperture. Because the environment is dynamic, such CQC simulations cannot pre-label the environment. By restricting labeling in this way, it forces a discussion of the processing involved in identifying apertures. A system like SECS, though in many ways more advanced, simply does not seem to have those design commitments.

The motor system as well is designed to have complete time-synchronized control of simulated agent, controlling the 3D simulated body in simulated time (as opposed to real-time) to avoid any inaccuracy due to simulation processing and communication between ACT-R and the simulator. Since time course data is arguably the most used data in the cognitive science, maintaining high-fidelity time-synchrony was a major concern in developing this system. Although in principle the system can be used to control real-world robots, the aim of ACT-R 3D is to remain purely simulated. High-level real-time robot control has been demonstrated with ACT-R/E (e.g. Trafton & Harrison, 2011).

As a proof of concept the Information Processing Theory of Geometric Affordances (IPGA) is presented. IPGA is an attempt to situate the affordance-based research of Warren and Whang (1987) in a fully-representational cognitive framework for aperture passage. Essentially, IPGA proposes the existence of an abstract spatial mechanism that compares the geometry (width, height, depth) of an aperture against estimated geometry of the body to both determine if an aperture is pass-able and to control shoulder rotation when required. A model of aperture passage, developed in ACT-3D is presented. This model begins to address open questions in the aperture passage literature. The following is a list of open questions:

(Q1) How an agent identifies an aperture,

(Q2) The role of A/S ratio, head-sway, and stride length in planning (if any);

(Q3) and whether that planning involves:

- (a) a specification of when to start rotation,
- (b) a specification of kinemantic properties of the rotation (e.g. rotation speed),
    - i. and whether speed is an explicit factor in that calculation;

(Q4) Whether the rotation strategy used while walking through an aperture without carrying a bar is the same as the strategy used while carrying a bar.

Although there is very little discussion on these topics in previous literature, they become immediate questions when implementing a computational model of aperture-passage behaviour. This work presented in this thesis does not fully address any of those open questions. What it does do is present a model that addresses them sufficiently to provide an outline of required cognitive mechanisms to perform the aperture passage tasks.

The following hypotheses are explored in this thesis:

($H1$) a geometric comparison mechanism is sufficient to account for both the judgment phase and the rotation phase.

($H2$) The input to the geometric comparison mechanism differs between the judgment phase and the rotation phase. Specifically, that the geometric representation derived from the body schema is multiplied by a factor in-and-around the range reported in Warren and Whang (1987) during the judgment phase; and that a constant is added to the body schema during the rotation phase as described by Higuchi et al. (2012).

($H3$) That the motor plan used in situations where rotation does not include a bar is not the same as the motor plan used in situations where rotation does include a bar.

1.1.2 Contributions

The following contributions are introduced in order to address Q1 through Q4 and test H1 through H3:

(C1) an alternative affordance theory that aperture passage is a geometric affordance.

(C2) a computational-level processing description of the overall aperture passage process.

(C3) a body schema system used for motor aperture passage judgment and motor planning.

Chapter 2: Theory

2.1 Overview

The aim of this chapter is to present an information processing theory of a class of affordances I call 'geometric affordances'. Unlike previous affordance theories (Chemero, 2003; Chemero & Turvey, 2007; Gibson, 1986; Stoffregen, 2000; Sahin, Cakmak, Dogar, Ugur, & Ucoluk, 2007; Turvey, 1992), the information processing theory and formalism presented below is not an attempt to generalize across every type of action but merely an attempt to provide a framework to understand a class of actions that require achieving a body posture that satisfies the geometric requirements of entering or passing through some feature within the environment. They are here named, *geometric affordances* because it is theorized that body posture is selected via a match between the estimated geometric properties of the environmental feature (e.g. width of an aperture) and the estimated geometric properties of the body in the selected posture. An example - and the example used throughout this thesis - is of passing through an aperture smaller than the frontal width of the body.

2.2 Geometric Affordances

The term 'affordance' was first used as a noun by Gibson when he presented an approach to Psychology called, ecological psychology (1986). An affordance can be thought of as a property (or set of properties) of the environment[1] that an agent uses in order to determine what actions are available. Action, of course, is a broad category and there has been empirical research in support of affordances in a large number of domains including grasping (e.g.

[1]Their ontological status varies, depending on the author.

Tucker & Ellis, 1998), reaching (e.g. Carello, Grosofsky, Reichel, Solomon, & Turvey, 1989), stair-climbing (Warren, 1984), a number of sports abilities (see Fajen, Riley, & Turvey, 2009, for a summary), and examined in this thesis, aperture passage (e.g. Warren & Whang, 1987).

Geometric Affordances (GA) can be seen as a subset of affordances that require geometric reasoning. The main example used in this thesis is of aperture passage. According to the information processing theory being presented, the Information Processing of Geometric Affordances (IPGA), when we judge an aperture as passable or not, we do so on the basis of an abstract geometric comparison mechanism that compares the geometric properties of the environment (depth, width, height) against either the geometric properties of the body in its current posture, or against the geometric properties of a stored representation of the body. If an aperture is smaller than the frontal width of a current or retrieved body posture, we may decide to rotate our shoulders, provided that a shoulder rotation will successfully reduce our frontal width, at time of passage, to afford passing through. This view of affordances directly challenges most philosophical positions in Ecological Psychology (Chemero, 2003; Chemero & Turvey, 2007; Gibson, 1986; Stoffregen, 2003; Turvey, 1992), but, as argued in 2.2, IPGA offers a more cognitively penetrable and commensurate view. The dominant view in ecological psychology is that affordances are perceived directly, whereas IPGA proposes that affordances are derived through inference processes involving a geometric comparison mechanism.

2.3 Geometric Comparison Mechanism

The information processing theory of geometric affordances posits, at a functional level, the existence of a geometric comparison mechanism that compares the geometric properties (width, depth, and height) of an aperture against a current or stored body posture. The geometric comparison mechanism is used both when judging whether an aperture is passable as well as to control the extent of shoulder rotation when rotation is required.

IPGA presupposes the existence of body schemas that can be compared geometrically against the dimensions of an aperture. Although previous affordance researchers have rejected the notion of body schemas (e.g. Carello et al., 1989), there is growing evidence of their existence and their role in motor planning and performance.

2.3.1 Evidence for Body-Schemas

Schwoebel and Coslett (2005) identify three types of body representations: *body schemas* that represent the positions of body parts and is used to plan motor movements, the *body structure description* that is a topological map of body part locations, and a *body image* which is a lexical-semantic representation of the body detailing body part names, their function, and their relationship to related artifacts. Since body schemas are a major assumption in IPGA, a brief summary of evidence for schemas as well as an elucidation of their cognitive role will be described below.

Neural evidence provides support for the functional role of body schemas as real-time representations of the body. Firing-rates in parietal area 5 of primates support the idea of encoding arm posture both when the arm is occluded and when a realistic, fake arm is visible, suggesting both somatosensory and visual input is used to create body schemas (Graziano, Cooke, & Taylor, 2000). Of particular interest in this thesis is the role of body schemas in motor simulation.

As an example, the hand laterality paradigm has been used to study the link between imagined movement and actual movement (e.g. Parsons, 1987, 1994). There are two aspects of the laterality paradigm. The first involves making judgments of laterality (left vs. right) and the second involves simulating arm orientations. The reaction time for both tasks was relative to orientation differences between the participant's arm and the target arm. Simulated movements were strongly correlated with actual movement times.

There is strong evidence for physiological overlap between imagined and actual movements (e.g. Lotze et al., 1999; Decety, 1996). The fMRI work by Lotze et al. (1999) also supports the view that the main difference between imagined and actual motor movements is inhibitory signals from cortical motor areas to cerebral regions, inhibiting actual movement. Sirigu et al. (1996) also shows that only patients with parietal damage do not show a correlation between the times for actual and imagined finger movements.

Finally, Schwoebel, Coslett, Buxbaum (2001; 2008) provide evidence for Forward Modelling. Forward Models have been theorized to be used to develop representations of body schemas based partially from efferent copies of planned motor movements. What is particularly interesting in their work is the dissociation exhibited by a patient (JD) between body schemas due to purposeful movement and body schemas for passive movement. JD had accurate reach and pointing ability when moving her hand to a target in both occluded-hand and non-occluded-hand conditions, suggesting that she had an accurate representation of the position of her arm and hand. However, JD's pointing and reaching ability were impaired when her arm was moved by an experimenter (passive movement), suggesting that, in those cases, she did not build an accurate representation of her arm posture. As pointed out by Schwoebel and Coslett (2005), this dissociation suggests that JD had an intact ability to generate posture representations from an internal model based on predicted movements (Forward Model).

Arguably, the above findings in combination imply that humans have representations of the biomechanical constraints of our bodies. If reaction times for imagined movement mimics reaction time for actual movement, then this suggests that the simulated movement has similar kinematic and biomechanical properties as real movements. The fact that there is a strong neurological overlap between simulation areas and areas responsible for actual movement suggests that motor movements are encoded in the same format for simulations as they are for actual movements. It can then be inferred that some form of biomechanical representation has to exist to support biomechanically-accurate simulation. This offers compelling evidence that the biomechanical constraints of the human body are likely also represented (in order to support simulation). It stands to reason that simulation can produce predictions of body posture in simulated motor planning in much the same way as forward modeling does for active motor behaviour.

2.3.2 IPGA Formalism

The information processing theory of geometric affordances assumes that an agent has a goal state to carry out an action such as walking to some location. It further assumes that the agent is operating under some action constraints: the desire to reach that location by walking, for example. IPGA is a theory about how the agent determines if the environment can afford the desired goal (i.e., *if* the action is possible within the constraints of their environment). IPGA proposes the following formalism:

(*motor-plan*,(*entity*, *body schema*, *action requirements*)),

where *entity* is properties of the environment, *body schema* is an internal representation of the body from which estimated geometric properties can be extracted, *action requirements* are the desired kinematics (walking, running, etc.), and *motor-plan* is an end-state body posture retrieved (through simulation) that represents a symbolic goal state of a motor control mechanism. If there is no match between *entity* and *body-schema*, and *action requirements*, no motor plan is returned and, therefore, the action query fails (i.e. walking through the aperture is not possible). The existence of a motor-plan helps partially reduce complexity of the degrees of freedom problem.

2.3.3 Degrees of Freedom Problem

The human body is capable of taking an extremely large number of postures. In fact, in a motor control context this is often referred to as the degrees of freedom problem. Broadly, the problem is about choosing a posture when there are so many postures to choose from. Consider the problem of aperture passage. Suppose in such a scenario the desire is to walk (as opposed to running, crawling, etc.) through an aperture. Suppose further that the aperture is narrower than the shoulder width when the coronal plane is perpendicular to the opening. To facilitate walking through the aperture, a new posture has to be chosen. The chosen posture has to both meet the top-down requirements of affording walking (e.g. cannot involve crawling) and the bottom-up requirements of the geometric properties of the aperture (the frontal width of the posture has to be narrower than the aperture). On their own, body-schemas do not provide a sufficient mechanism because as described by Schwoebel and Coslett (2005), body-schemata are merely representational; they represent body-postures. The degrees of freedom problem is about choosing the appropriate body-posture for a given situation.

A geometric comparison mechanism reduces the complexity of the overall problem because the set of suitable postures is reduced to the set that fit the geometric requirements of the goal state. The space of postures is further reduced by the inclusion of action requirements, as certain postures might meet the geometric requirements but not afford the appropriate actions. Even with these limitations the degree of freedom problem remains quite large if there are seemingly infinitely divisible body postures (or at least a large number of body postures) that might fit both the action requirements and the geometric requirements. For example, clock-wise shoulder rotations, even assuming a rotation limit of 90 degrees, can be divided into any number of intermediate rotations, and is further complicated by direction (left vs right). In order to reduce complexity, the proposal being put forward is that we store body postures only for biomechanical body constraints, excluding postures between constraints.

2.3.4 Implementation of Body Schemas

The idea is fairly straightforward: when we rotate a joint and that joint reaches its maximum rotation (maximum both in the sense that it is the most rotation achieved, as well as the sense of the rotational limits of the joint) along the principle axes of rotation for that joint, then a body schema is produced and stored for that posture. Storing only the body schemas at biomechanical constraints limits the search space for body schemas at some future time.

In the information processing theory being presented, the biomechanical constraints are used as a motor plan or a goal-state of a the motor system. How to then achieve that goal state could plausibly be implemented as procedural knowledge represented in the motor system.

2.3.5 Aperture Passage Affordance

Aperture passage was first studied in a affordance context in Warren and Whang's (1987) classic aperture passage experiments. In those experiments, Warren and Whang attempted to identify the visual properties that modulate shoulder rotation when walking through apertures. Since Warren and Whang's classical aperture affordance paper, there have been a number of papers that explore different aspects of aperture passage. For example, Stefanucci and Geuss had people duck under apertures, including conditions where they had to wear blocks on their feet or helmets on their heads (Stefanucci & Geuss, 2010). Adding objects into the mix is a common theme, and a number of papers have explored aperture crossing for a person-plus-object system. For example, Wagman and Taylor (2005; 2007) and Higuchi et al. (2006, 2012) explore affordance judgment and passage while carrying an object. Higuchi, Takada, Matsuura, and Imanaka (2004) and Higuchi et al. (2006) studied passability judgments and performance of novel wheelchair users. Chang, Wade, and Stoffrogen (Chang et al., 2009) studied the passability affordance judgment for people grouped in a dyad. Fath and Fajen (2011) explored aperture passage in virtual environments in order to control for available visual input. A lot of this research has either explicit or implicit ties to Ecological Psychology and, in particular, Gibson's notion of *direct perception* (Gibson, 1986).

Direct perception is the claim that our interactions are not mediated by an internal, sensory-based, semantically-laden representation of the environment, about which we can make inferences, that ultimately lead to action. Instead, direct perception holds that the actions applicable within an environment or towards an object are part of the perceptual experience itself, or is directly perceived. Such perceptual processes are possible, according to their theories, because the perceptual system is tuned to pick up information relevant to action. One way to illustrate this might be in reference to the aperture passage work by Warren and Whang (1987) that explicitly aims to defend direct perception as a mechanism behind affordances.

A common theme in empirical ecological psychology research is to identify a body-scaled unit that is used by the perceptual system to realize an affordance. For example, in his work on stair climbability, Warren (1984) proposes that stairs are perceived in units of leg length. Perceived in this manner, there is no need to represent the environment in any extrinsic or mental units. The height of a stair is simply perceived to be some proportion of leg length. These dimensionless ratios, called *pi numbers* (π), relate some dimension of the environment (E) to some dimension of the the body (A, for animal). Essentially,

$$\pi = E/A$$

Researches have attempted to identify 'critical points'. *Critical points* are when π-ratios are greater than 1.0, or, when the relevant dimension of the body surpasses the sufficiency requirements of the environment. For example, a stair is efficiently climbable when its riser height is sufficiently proportional to leg height (Warren, 1984).

In relation to aperture passage, Warren and Whang (1987) performed a series of experiments aimed at showing that aperture passage is directly perceived. In their first experiment they had participants walk through apertures, rotating their shoulders as needed. Participants were grouped according to size: small or large. Unsurprisingly, small participants rotated their shoulders less than large participants when passing through the same aperture. However, when expressed as an aperture-width to shoulder-width ratio (A/S), the group differences were eliminated, suggesting that absolute degree of shoulder rotation is modulated by the A/S ratio. This experiment was also able to establish a critical ratio (π number) at which participants, regardless of their size, would change from maintaining a forward posture to a posture that included rotation. This π-ratio (1.3) they maintain is a constant, used by an agent to determine when shoulder rotation is required.

The second and third experiments were meant to explore what optical information contributes to passability judgments. In these experiments, aperture passage was only judged and not performed. Participants were asked to judge if they could pass through apertures without shoulder rotation. Experiments two and three manipulated the optical information by modifying binocular and monocular vision, static (no movement) and dynamic (walking towards the aperture), as well as introducing an Ames-room-like illusion. Both of these experiments were meant to establish that the perception of passability is scaled to body units as opposed to absolute size judgments (presumably in some extrinsic dimension). The Ames-room experiment is particularly compelling.

The theory that Warren and Whang develop in the course of the paper is that aperture passage can be directly perceived in units of eye-height that, for each individual, exists in a constant ratio with shoulder width. Instead of perceiving the width of the doorway in extrinsic units (such as is proposed by IPGA), the doorway is perceived in intrinsic units of eye-height. Aperture passage judgments and performance need not involve any form of representation or inference because all the information is specified through the interaction. This position, however, makes it very difficult to ask important cognitive questions.

Consider, for example, the rotation results in Warren and Whang (1987). Although they did not report exact figures[2], large participants rotate approximately to an average of 78°for apertures of $40cm$. Large participants have a shoulder width average of $48.4cm$ ($SD = 0.7cm$). Assuming a shoulder width of $50cm$, At 78°, the frontal width of a $50cm$-shoulder-width participant is approximately $12cm$ ($cos(78°) * 58cm = 12.06cm$), which is a large over rotation for an aperture passage of $40cm$. It is very difficult to raise a question about the over-rotation in an Ecological Psychology framework because it is unclear what mechanism controls the rotation, what the input and output of that mechanism is, and what the temporal dynamics of that system might be like. Clearly the motor system plays a role but it's unclear whether the rotation is planned (top-down) or whether it is reactive (bottom-up) or whether it is some mix of both. For example, it is unclear if this is a planned-over rotation to maintain a large safety margin, or an over-rotation due to a temporal delay in the motor system. Alternatively, the over rotation could be due to

[2]Dr. Warren was contacted for data but only a limited amount of data was available since the experiments preceded this thesis by 29 years.

perceptual processes or some complex combination of any number of reasons. Most troubling is that it is difficult to even raise these and similar questions from a non-representational and non-information-processing framework.

2.4 Information Processing Description of Aperture Passage

This thesis proposes that the overall aperture crossing process can be usefully divided into two phases: a judgment phase, and a performance phase. In the judgment phase, we first determine whether we can pass through the aperture at all. The performance phase occurs once we have judged an aperture passable and begin to walk through it. The performance phase can be subdivided further into three sub-phases: rotation initiation, rotation, and rotation termination. The following will provide a detailed outline of the proposed processes in each stage.

2.4.1 Judgment Phase

Although it is discussed very little in the previous aperture-passage literature (Warren & Whang, 1987; Wagman & Taylor, 2005; Wagman & Malek, 2007; Higuchi et al., 2006, 2012; Stefanucci & Geuss, 2010), before we ever attempt to pass through an aperture, we must first make a judgment of whether passage is at all plausible. Anecdotally, this must be the case because we simply do not often find ourselves trying to squeeze through apertures smaller than our bodies. This process has to be more complex than the direct perception theory (2.3.5) proposes because passage cannot be judged purely on current posture.

In Warren and Whang's (1987) second experiment, participants were asked to judge the passability of an aperture under the constraint that shoulder rotations are not allowed. With these constraints, participants judged passability at a mean width of $47.4cm$ for small and a mean width of $55.7cm$ for large participants; whereas in the first experiment (where shoulder rotation was acceptable), all participants passed through apertures of $40cm$. This suggests that participants based their judgment of passability on a representation of a future body state. It is unclear how a *direct perception* approach could account for this finding as a direct relationship between the aperture width and the frontal width cannot plausibly exist until the shoulders are rotated. The findings are also suggestive of separate processes: one that determines if passage is possible (judgment phase) and a second that performs rotation (performance phase).

The information processing theory of geometric affordances proposes instead that a positive passage judgment results from two possible situations. In the first case, current body geometry is estimated from body schema of the current body posture (Schwoebel & Coslett, 2005). This information can then be used top-down to filter information in the visual system to find apertures of an appropriate size. If the vision system is able to return a feature in the environment that meets those constraints, the returned apertures are considered passable. If no environmental feature is returned by the vision system, the second case proceeds. In the second case, a potential series of memory requests are made for stored body schemas that match the present body schema for action capabilities but differ with respect to joint rotation at their biomechanical constraints; for example, a walking posture with the shoulders fully rotated. As described above, storing only the biomechanical constraints vastly reduces the search space for suitable postures. If a body schema is returned, the geometric properties of that schema are used to filter visual results in the same manner described above for the first case. In the second case, the retrieved body schema functions as a goal state for the motor system during the rotation phase.

2.4.2 Rotation Phase

Another aspect of aperture passage with no known discussion in the affordance literature is the need for some trigger that starts the rotation. One possibility is that the agent plans to rotate at some specific point and initiates rotation upon arrival. A second possibility, and the one explored in this thesis, is that there is a bottom-up environmental trigger that is responsible for initiating the rotation. The theory proposed here is that the visual system performs bottom-up obstacle avoidance and that the presence of the edges of the aperture triggers the rotations. When the edges of the aperture are within a multiple of the agent's rotation radius, the vision system pushes information into the visual buffer, which can then be responded to.

Recall that during the judgment phase a stored body schema memory may be recalled and used as a goal state for the motor system to achieve the affordance. In Warren and Whang's first experiment (1987) there is a multi-second delay between what I am describing as the judgment phase and the rotation phase (while the participants walks to the aperture), therefore I propose that once a body schema is retrieved it is maintained in the imaginal buffer. When the presence of the obstacle (aperture edges) is pushed in to the visual buffer combined with the presence of a body schema in the imaginal buffer, the body schema representation is pushed into the goal buffer of the motor system. Note that for shoulder rotation, the goal state will be a biomechanical constraint, i.e. fully-rotated shoulders. However, we know from Warren and Whang (1987) and intuitively that we do not rotate our shoulders to maximum rotation every time we rotate. Instead the theory assumes that rotation completion is controlled by a vision-action loop in the dorsal visual stream.

2.4.3 Rotation Completion Monitoring

In their *Two Visual Streams Hypothesis* Milner and Goodale propose a functional distinction between the dorsal visual stream and the ventral visual stream (2011; 1992). They propose that the ventral stream composes what they call *vision-for-perception* and that the dorsal stream composes what they call *vision-for-action*. This distinction is recognized by Anderson in his book on cognitive modeling and ACT-R (Anderson, 2007, p. 48). This distinction is also implemented in the ACT-R vision system, SECS (Harrison & Schunn, 2002). In the present work the distinction between the streams exists on a functional level.

Two Visual Streams Hypothesis

The default vision system in ACT-R 6.0 is influenced by the two visual system hypothesis presented in Ungerleider and Mishkin's series of lesion studies (Ungerleider & Mishkin, 1982). Ungerleider and Mishkin present a convincing story but more recent evidence suggest that the functional distinction between the *what* (ventral) and *where* (dorsal) streams is not accurate. In particular, evidence presented by Milner and Goodale (Goodale, 2011; Milner & Goodale, 2008; Goodale & Milner, 1992) suggests that the division of labour between the dorsal and the ventral stream is *vision-for-action* and *vision-for-perception*.

After their initial introduction of the functional role difference between the dorsal and ventral vision pathways (Goodale & Milner, 1992), Milner and Goodale (Milner & Goodale, 2008) spend a fair amount of time clearning up the distinction made between how they use the term 'action' and how they use the term 'vision'. By vision-for-perception they mean the properties of an object that might be used to parse a scene including describing and identifying objects. Vision-for-action, conversely, refers to the visual properties of an object which, *during the course of carrying out an action*, program the motor movements for carrying out the action. In terms of action, the ventral (vision-for-perception) stream is used for planning action and carrying out unpracticed action whereas the dorsal stream (vision-for-action) is used for moment-to-moment visual updating of actions that are comparably more automatized.

There is a variety of different evidence in support of the action/perception functions, though much of the evidence comes from dissociation between brain damaged patients with either optic ataxia (impairment of action performance, but keeps object recognition in tact) or visual agnosia (impairment in discrimination/identity of visually presented stimuli with action performance in tact). In particular, a number of studies involve the patient, DF, who cannot recognize faces, cannot perceive object dimensions (inability to report or pantomime), or their orientation, however is able to make the appropriate grasp action when picking up objects. This dissociation suggests that her motor system is aware of some physical object properties that she reports not seeing.

Some of the more interesting non- dissociation/double-dissociation studies including studies that use optical illusions such as the Ebbinghaus illusion, where the percieved size of discs are influenced by the discs surrounding them. In a grasping study by Aglioti, DeSouza, and Goodale (Aglioti, DeSouza, & Goodale, 1995), participants reported the sizes of target discs to be different, but when reaching to grasp the discs they showed no difference in their maximum grip aperture. Illustrating the distinction between action planning (ventral stream) and moment-to-moment action programming (dorsal stream); van Doorn, van der Kamp, and Savelsbergh (van Doorn, van der Kamp, & Savelsbergh, 2007) perform a Müller-Lyer experiment combined with both action planning and active grasping. Participants in this study showed the effect of perceived rod length difference in both judgment and grasping choice but showed no effect of the illusion when performing a grasp.

When viewing through an ACT-R lens, an important aspect of Milner and Goodale's theory is that the information projected in the dorsal (action) stream is limited to the moment of action. The information conveyed in the dorsal stream is seemingly not present when performing actions from memory. In a study by Hu, Eagleson, and Goodale (Hu, Eagleson, & Goodale, 1999), participants performed grasps in three different conditions. In the first condition, participants grasped targets after a 5*s* delay; in the second condition, participants grasped targets without a delay; and in a third condition participants grasped objects while both their hand and the target remained visible. It was only in the delay condition that grasp aperture was significantly different, suggesting that size estimation information is different when encoded dorsally than when recalled. Interestingly, under similar conditions, but combined with the Müller-Lyer illusion, participants show the effect of the illusion when a delay is present (Westwood, Heath, & Roy, 2000), further evidencing that information from the ventral stream is encoded in memory and information from the dorsal stream is not.

The two visual streams hypothesis suggests a tight coupling between the vision system and the motor system while carrying out actions. It also supports the idea that the type of information being used to program and update the motor system is precise geometric properties. The information processing theory of geometric affordances leverages the two visual streams hypothesis by proposing that affordance passage completion is a result from a constant monitoring of the geometric properties of the body (via body schemata) and the geometric properties of the aperture. This tight coupling between the visual system and the motor plays a functional role of cancelling the motor plan once sufficient rotation is achieved. By monitoring the rotation and cancelling once sufficient rotation is achieved, the motor system need not maximize rotation for every case, despite the original plan retrieved as a posture at a biomechanical constraint.

The evidence in support of the two visual streams hypothesis supports the idea that geometric information is encoded in the vision-for-action system, evident in the grasping aperture's resilience to illusion. This resilience to illusion may also help to effectively ignore Warren and Whang's fourth experiment where participants reported illusory aperture widths when confronted with a false floor. It is perfectly reasonable, assuming the division between vision-for-action and vision-for-perception, that participants reported based on ventral stream processing and perfectly plausible that, had they passed through apertures, we might expect similar rotation angles to non-Ames-room conditions.

Chapter 3: Method

3.1 Overview

The aim of the present chapter is to provide a thorough description of ACT-R 3D and the aperture-passage model developed therewithin. In order to situate the system and model appropriately within previous research, a brief summary of related affordance models will be presented in 3.2. Because the modeling system is based in ACT-R (Anderson, 2007), a brief review of recent relevant ACT-R modifications will be presented in 3.3.

3.2 Affordance Model Review

The purpose of this section is to describe two models of affordances: a dynamic systems affordances model (Fajen et al., 2003) and robotics affordance model (Sahin et al., 2007; Ugur & Sahin, 2010).

3.2.1 Dynamical Systems Model of Obstacle Avoidance

Including a dynamical systems model is important to set up a useful contrast to the modeling efforts presented in this thesis. Dynamical systems modeling is, perhaps, the most appropriate form of modeling when modeling affordances. This is because many Ecological Psychologists make ontological commitments that affordances are emergent from the relationship between an animal and its environment (e.g. Chemero, 2003; Chemero & Turvey, 2007; Stoffregen, 2003) and not derived through inference processes as claimed in IPGA.

One such model, though not related to aperture passage, is a navigation model by Fajen et al. (2003). Their model was designed to model agent navigation to a goal while avoiding obstacles. They aim to show that seemingly intelligent path planning behaviours can emerge from control laws which define the relationship between properties of the environment and the agent. Their model essentially consists of a goal component, an obstacle component and a dampening term. The goal and obstacle components act as attractors and repellors (respectively), influencing the agents heading accordingly. The dampening term lessens the impact of turning.

The first two model simulations are run in virtual conditions similar to their first two experiments. Essentially, the first two simulations are to test the equations and set parameters. Unsurprisingly, their simulations were accurate to human performance (r^2= 0.982 for experiment 1, $r^2 = 0.975$ for experiment 2). In their third simulation, the model performed in a novel environment setup for which an obstacle was present between the starting position and goal. Under a variety of placements, the model would chose a path that was either to the left or right of the obstacle. With a minor change in parameter settings, the model successfully predicted behaviour of human participants in experiments similar to simulation conditions (no r value reported).

The final two simulations in the Fajen et al. study (2003) were not compared to human data but were presumably meant to illustrate the applicability of the model across various situations. In the fourth simulation Fajen et al. included a close and distance obstacle. What they found was that the model took the most efficient route. Important in their paper, the route exhibited by the model emerges as a consequence of the control laws (equations). Finally, their fifth model was exposed to randomly positioned obstacles. The success of this model is described as its ability to always reach the goal while avoiding all obstacles.

3.2.2 Affordances for Autonomous Robotics

The present subsection presents a novel work in robotics that aims to learn affordances. Although there are other affordance-related works in robotics such as (Shneier et al., 2008) whose robot learns traversability (can the robot traverse over), or the PDP work by Haazebrock, van Dantzig, and Hommel (2011) which associates features codes with certain actions, the work by Ugur and Sahin (2010) is the most relevant to this thesis since one of the affordances learned by the robot is an aperture passage affordance. While the work is impressive, their approach aims to mimic direct perception by limiting the representations it uses. Once reviewed, I will discuss how the work in robotics differs from the work proposed here within.

Ugur and Sahin (2010) present a robot highly inspired by the work of Gibson (1986). One of the biggest problems faced in autonomous robot navigation is programming a robot with the ability to navigate novel environments with novel obstacles. Instead of pre-programming their wheeled robot to navigate an environment with a specific set of obstacles or perform standard obstacle avoidance techniques, that often result in myopic behaviour, Uger and Sahin programmed their robot to learn a mapping between visual features and actions.

The action capabilities of the robot were simplified to consist of seven movement categories. The movement categories consisted of a rotation of either 0°, ±20°, ±40°, ±60°; followed by a forward movement of 70cm. The goal of the robot was to learn when the environment afforded a particular movement category. The robot's perceptual system consisted of a 720 x 720 (reduced to 360 x 360) range camera. The images were placed in a grid and analyzed for distance and shape. A single, one-dimensional feature vector was then used to represent an image.

The robot learns to associate the perceptual vectors with the action categories using a novel affordance formalization. One of the problems they needed to solve was how to make the learning general enough and how to distinguish a successful action (which should be repeated) from an unsuccessful action (which should not be repeated). To facilitate affordance learning, their formalization represents the relation in a nested triplet of the form:

$(effect, (entity, behavior))$

in which the *entity* represents the environment component, the *behavior* component is an action capability of the robot, and the $effect$ is a result. Importantly, each of the components in the triplet are represented as an *equivalence class*. The learning algorithm attempts to generalize across effect, entity (perceptual vector), and behavior when there exists multiple mappings between any of the components. For example, a rotation of 20° and a rotation of 40° might be generalized.

The robot was able to learn traversability in a number of different situations. It was able to learn aperture passability, gap crossability, and ramp climbability. In novel environments populated with obstacles in the original training set, the robot was able to successfully identify traversability with a 90% success rate. However, when placed in an environment with three novel, complex, real-world objects, correct traversability judgments varied with a marked drop in performance.

For example, in a virtual environment where the height of a wooden table was adjusted such that it was possible for the robot to pass underneath 64.9% of the time, prediction success was 85.5%. When the position and orientation of a metal table with a bar connecting a pair of legs, which was adjusted in height, that was passable in 74.2% of the cases; correct judgment of passability was 81.7%. Finally, when the orientation and position of a wooden chair was varied such that passability was at a rate of 68.4%, the robot made correct passability judgments with 94.7% success.

3.2.3 Affordance Models Discussion

The purpose of the above two model reviews is to set up a sense of contrast as to the aims of the model presented in this thesis. Although both models have their merits in terms of how well they model their respective tasks, the major objection to both of them is the lack of insight into the information processing involved in affordance judgments and motor control.

Consider first dynamical systems model by Fajen et al. (2003). Their model consists of an equation with a term for the goal (attractor), a term for obstacles (repellor), and a dampening term. In their model, behaviour –rotational acceleration– is a product of the interaction of two *control laws*. The model assumes semantic identification of the goal as the goal and the obstacle as an obstacle. It is impossible to then ask cognitive questions about how an agent might identify a goal or an obstacle because they are pre-coded in the environment and in the control laws. The equations model the path of the average participant but, leaves an open question as to how the path was derived. Because there is no mechanism for derivation, it is impossible to ask questions about individual differences. While most models are meant to capture the performance of the average participant, if its performance is mechanistic, then researchers can at least ask questions about what aspects of the mechanisms may account for individual differences. Being unmechanistic, we cannot break down the system any further, without adopting a brand new model. We therefore do not gain any insight into cognitive processing and cannot penetrate the model in hope of gaining further insight.

Considering the robotics model (Ugur & Sahin, 2010) is not as straightforward. Although not cognitive, their model does have some resemblance to a cognitive model. Their robot has perceptual processes that get encoded into a visual vector, and builds a memory over time through its learning processes (a support vector machine). The memory system is easily likened to a motor memory, providing mappings between what the robot sees and an appropriate action. Unlike the dynamical systems model above (Fajen et al., 2003), it is possible to interrogate and modify the model. It is possible, for example, to ask questions and modify how the vision system works and, perhaps, make different assumptions about the kinds of visual components to encode in the perceptual vector. So, in many ways, their model has many features and merits of modern cognitive models.

The robotics affordance model (Ugur & Sahin, 2010), however, is not a cognitive model, and it is unsurprising that it lacks the mechanisms we expect to find in a modelling architectures such as ACT-R (Anderson, 2007). One of the prominent mechanisms that exists in ACT-R that does not exist in the robotics model, is a limited capacity central processing bottleneck. In particular, ACT-R's production system limits the number of processes that can be computed in a given time frame and as such has been one of the main sources of human-comparable performance, allowing researchers to validate their models against behaviour measures of reaction-time, learning rates, and error rates (among others). While the robotics model has strong merits as a robotics model and possibly some as a cognitive model, it is a lack of cognitive architecture that detracts from the kind of validation that can be made as a model of human cognition. As a robotics model, however, this was never the expectation.

The remainder of this chapter will be committed to outlining a model of aperture passage made for ACT-R 3D. Although the model described will have similar shortcomings to the robotic's model (Ugur & Sahin, 2010) because new, invalidated architectural components are added to ACT-R for the purposes of the model. However, the hope is to explore modeling phenomena that are not traditionally modeled in an otherwise highly successful cognitive modeling architecture.

3.3 Recent ACT-R System Modifications

As alluded to in the opening paragraphs of Section 1, one of the main motivations for this thesis is to explore cognitive modeling of complex behaviour. The cognitive architecture, ACT-R, is an attractive modeling architecture because of its success in lab-based experiments. Lab-based experiments, however, are often designed to be simplified in order to control for a multitude of factors that may influence performance on a given task. Using ACT-R as a tool for researching more complex behaviour is compelling and has enjoyed some success with fairly recent modifications. One of the most impressive examples of ACT-R being used to model complex behaviour is the model of driving behaviour by Salvucci (Salvucci, 2006). Another recent extension to ACT-R is ACT-R/E (Trafton et al., 2006, 2009; Trafton & Harrison, 2011; Trafton et al., 2012) which uses ACT-R as a high-level robot controller. Because a similar system was developed for the purposes of this thesis, the present section will provide a brief overview of ACT-

R/E, including a brief overview of use cases. Because of its strong integration with the ACT-R vision system, Specialized Egocentrically Coordinated Spaces (SECS) (Harrison & Schunn, 2002), both will be described together as needed.

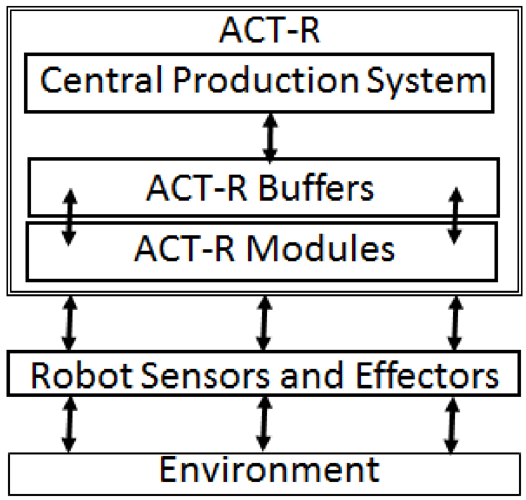

Figure 3.1. A simplified depiction of ACT-R/E, resembling image from (Trafton and Harrison, 2011).

ACT-R Embodied (ACT-R/E) (Trafton & Harrison, 2011) uses ACT-R as a robot controller with an updated visual system, SECS (Harrison & Schunn, 2003). SECS has three main systems: *visual*, *manipulative*, and *configural*. The visual system uses fiducial and face trackers to provide object identification for video camera images. The manipulative system represents objects as 3D geons (stored in a database) as well as position and orientation information. The manipulative system also supports spatial transformation such as rotations in a manner similar to that proposed by Shepard and Metzler (Shepard & Metzler, 1971), supporting motor planning. Finally, the *configural* system provides egocentric range vectors to objects.

ACT-R/E also extends the basic ACT-R motor system that is very limited. Although somewhat vague in description, Trafton et al. (2011) suggests the motor system maintains real-time limb representations and restricts movements based on muscle groups and Trafton et al. (Trafton et al., 2009) and Harrison and Trafton (2010) suggest that motor control is handled externally to the ACT-R architecture once the ACT-R selects a motor command. For example, given the description by Harrison and Trafton (2010), once a representation in the manipulative module is associated with the objects semantic representation, the central production system in ACT-R issues an appropriate grasping command to the robot controller, which carries out the grasp. ACT-R/E has been used in a variety of models. Harrison and Trafton (2010) used ACT-R/E to model response times of grasp actions, Trafton and Harrison (2011) used it to model gaze-following and level one perspective taking, and SECS (the spatial representation system used in ACT-R/E) was used to model an egocentric navigation task (Harrison & Schunn, 2003)

3.3.1 Potential ACT-R/E Shortcomings

ACT-R has been used as a high-level controller for agents dating back to the Close-Quarters-Combat (CQC) work by Best and Lebiere (?, ?). These models were very early and largely a test of ACT-R from a functional perspective: whether or not ACT-R as a system could scale up to complex tasks. What the CQC model lacked, however, was any strong mechanisms for cognitive control. For example, much of the perceptual and navigational aspects relied heavily on the simulation environment they were built in. The simulation environment was equipped with both perceptual and path planning tools that enable much of the details of the cognitive processing involved to be glossed over. One of the shortcomings resulting from the glossing over was that, as system, it was unable to support many of the predictive abilities that have made ACT-R such a widely used and validated cognitive architecture.

ACT-R/E goes far beyond the early CQC model in terms of the model cognitive plausibility. SECS (Harrison & Schunn, 2003) alone provides a powerful vision system for 3D, real-world environments. The ACT-R/E motor system also seemingly provides realistic motor performance constraints by limiting action by muscle group. One of the main differences between ACT-R/E and ACT-R 3D is focus. Although both systems have similar theory bases: ACT-R/E is based in the embodiment movement in psychology and ACT-R 3D is based in affordance theory, which are strikingly similar theories from a high-level perspective. The intended focus of ACT-R 3D, however, is to maintain some of the tenants of affordance theory and limit the amount of assumptions used in the modeling of affordance behaviour. The main assumption is to avoid the use of pre-coded actions. To illustrate, a brief overview of an ACT-R/E experiment will be presented below.

Harrison and Trafton (2010) present a model of the visual-motor compatibility effect (Tucker & Ellis, 2001). In these experiments, participants are trained to respond to everyday objects, sorted into two groups, either with a precision grip (thumb and forefinger) or a power grip (full hand grip). Importantly, objects depicted in either response group will either be objects that one would grasp using a precision grip or objects that one would grasp using a power grip. Tucker and Ellis show a reaction time advantage when the natural grip is compatible with the trained response grip. Harrison and Trafton model reaction time by implementing spreading activation. Motor response is learned through co-occurrence between the semantic content of an object and and the grasp used to manipulate that object. While impressive, and one of the best known examples of why a systems like ACT-R/E and ACT-R 3D are useful, there is a manner in which the ACT-R/E grasp model does not strictly adhere to affordance theory (and it need not as their research has a different focus).

However, one of the main aims in affordance research is to explain how an action might be chosen. Although there it is somewhat unclear, it seems safe to assume that the encoding between an object (e.g. pencil) and the appropriate action (e.g. precision grip) is pre-coded in the ACT-R/E productions. That is, the ACT-R/E agent likely does not autonomously chose what grasp to use but rather the association between object and action are pre-programmed. As described in 3.4, ACT-R 3D uses a body schema system to enable the agent to autonomously map between features of the environment and the appropriate action.

Also present in Harrison and Trafton's paper (2010) is a brief suggestion that the robot's motor system has a slow execution time. While that in-and-of itself is not necessarily an issue in their research, it speaks to an important technical issue in related reaction-time research. Time-based simulators often run in simulation time, which, in many cases, means that they run faster than real-time. However, many simulators, where measures like reaction-time are not an issue (e.g. flight simulators, driving simulators, robotic simulators like MORSE), the simulators attempt to run in real-time. In a situation where a simulated-time simulator needs to work in-synch with an real-time simulator, in many cases, approximate synchronization can be achieve by slowing simulation-time simulator to real-time. One of the concerns, however, in developing ACT-R 3D was with processing delays in either the real-time simulator (i.e. MORSE), the simulated-time simulator (ACT-R), or time delays in communication between the two.

To illustrate, suppose we want to model visual processes with a robot simulator. Assume that there are appropriate computer vision algorithms to carry out whatever function we require. Assume further that we know approximate processing times for, say, moving visual attention from a blue triangle to a red square. Suppose that time is 80 milliseconds. Suppose such a request involves some form of visual search and different types of recognition by the vision system on our robot. In normal cases, the clock-time of that processing is going to depend on the computational power of whatever system is doing the vision on our robot. Whether the robot's vision system can carry out the task faster or slower than 80 milliseconds is of no consequence. What is required if our measurement is going to be accurate is a system such that, no matter how long it takes in real-time to process the vision, the appropriate amount of simulation time elapses. In research where tens of milliseconds can have a lot of implications, simulation-time synchrony seems like a strong requirement. Since none of the publications related to ACT-R/E address these concerns, it is assumed that they have not been addressed.

3.4 ACT-R 3D

I designed the software used in this thesis for modeling an embodied model in the Python variant of ACT-R (T. C. Stewart & West, 2005). The system is designed to enable modelers to implement ACT-R models in a 3D, simulated environment. ACT-R 3D takes advantage of a pre-existing robotics simulator, MORSE (Mobile OpenRobot Simulation Engine) (Echeverria et al., 2012). The system is entirely open source and modelers can implement their own 3D simulation to suit their needs. Custom simulated robots can be developed for simulation in MORSE using Blender[1] and these simulated robots can, in theory, have any number of limbs and articulating joints. Sensors on the joints can provide feedback to ACT-R and ACT-R can be in complete control of the joints, thus simulating cognitive motor feedback and control.

3.4.1 MORSE

The MORSE simulator is a robotics simulator based in Blender. MORSE comes with a number of standard sensors and actuators, pairable to a number of robotic bases. Custom environments, robots, and sensors can be developed in Blender. MORSE is written in Python 3 and the Python library supports full control of robots through Python scripts. Communication between the robot, sensor, and actuators; and the control program is handled through sockets. The intention of this project was to make ACT-R function as the simulator control script (albeit a complex one). The following subsections will describe the MORSE-relevant developments required for this thesis. For a description of the ACT-R-relevant developments see 3.4.6.

3.4.2 MORSE-based Human-Body Model

In order to model human-like shoulder rotation control, a low-fidelity, human-like software robot (softbot) was developed using Blender. The softbot, Manny, is presented in Figure 3.2. The robot consists of a rectangular cuboid base mesh with a single armature that, in turn, consists of: a rectangular

[1] https://www.blender.org/

cuboid pelvic region, a rectangular cuboid torso, a spherical joint between the pelvic region and the torso (not visible) two rectangular cuboid shoulder, two spherical shoulder joints, two cylindrical upper arm segments, a spherical neck joint, a spherical head, and a custom camera. Figure 3.2 is a polygon reduced version for quicker graphics processing. Each segment of the armature has full degrees of freedom. Although Blender does support inverse kinematic solvers for armature control, there are no kinematic solvers associated with any of the armatures for this project. All joint limb control and position representation is controlled by ACT-R 3D, synchronously, as described below. Collision sensors are placed on the shoulder joint, the shoulders, the upper arm segments, and the torso. An alternative version of Manny also includes a length adjustable bar protruding from chest region to simulate an agent holding a bar. Collision sensors are also implemented on the bar.

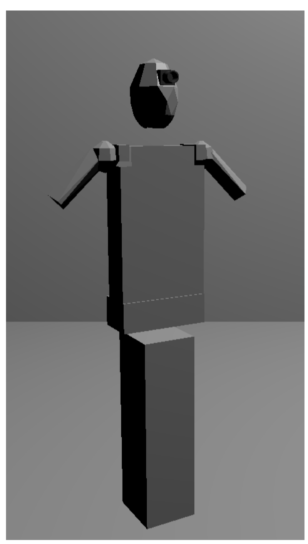

Figure 3.2. Screen shot of softbot, Manny

3.4.3 Simulation Environment

The simulation environment is presented in Figure 3.3. The environment mimics the experimental setup in the study by Warren and Whang (1987). The environment consists of a ground plane, two front wall segments separated by a gap (of varied aperture width according to experimental conditions), and a distant wall. The linear distance between the forward wall segments and the rear wall is 4.25 meters, matching the setup in Warren and Whang.

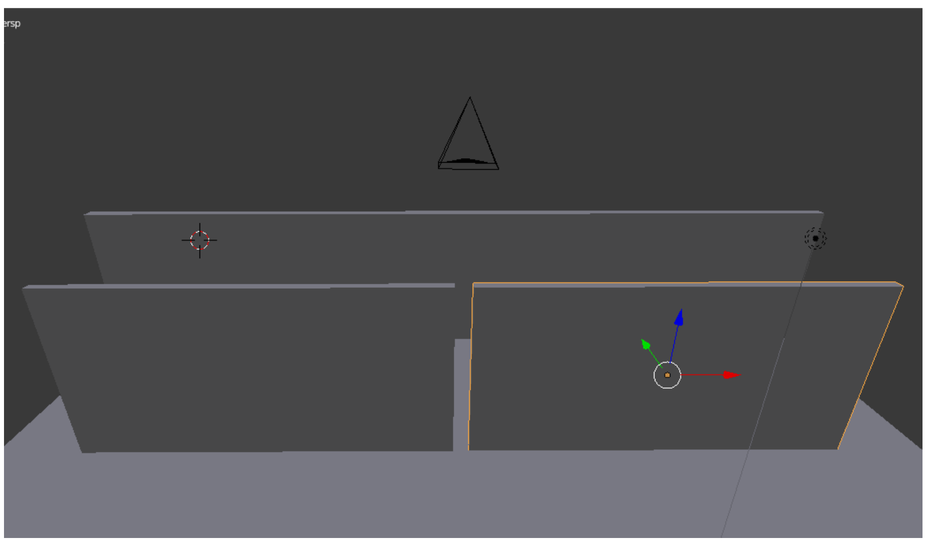

Figure 3.3. Screen shot of the simulation environment

As is evident in Figure 3.3, the walls and floor, unlike the case in the Warren and Whang experiment, are untextured. This fact is not an issue for the models because the vision system is not sensitive to texture (as will be described in 3.4.6). The walls are physically fixed in place and are not affected by collision. The walls have an ‘obstacle’ property that allow them to be detected by the agent’s collision sensors and is used to filter visual processing, but is otherwise unused for the purposes of the simulation. Importantly, apart from the floor (to reduce processing time in the Geometric Camera), there are no semantic labels used in the environment. The environment does not specify any semantic information to be detected by the agent. Determining where and what the aperture is, is done by the model at the time of simulation.

3.4.4 MORSE-side Middleware

MORSE natively supports scripting from Python scripts through socket interfaces. To manage the problem of synchrony as described in section 3.3.1 the plan was to use a 10 ms simulation cycle on both the ACT-R side and the simulator side. To synchronize ACT-R with MORSE, 10 ms of simulated time occurs in ACT-R, communication between ACT-R and the simulator occur, and two successive cycles are run in MORSE which is set to run at 200 frames per second. MORSE is then paused, and waits for the next tic signal to come from the ACT-R side[2]. The two successive cycles helps avoid a limitation in MORSE where only one request for data can occur per tic cycle. A second method for avoiding that limitation was also developed involving a queue that handles requests to the simulator. Essentially, this function is mapped to the original function calls. A handler is built on the robot side to carry out an arbitrary number of calls, effectively disguising multiple calls as a single call, that function then returns a dictionary, that, on the ACT-R side, is re-routed to the function that originally called it.

[2]This process did reveal a bug in MORSE with methods that perform the external tic triggers. A fair amount of effort went into first finding this error and having it fixed. As far as I am aware the solution for external triggers has changed in the most recent versions of MORSE.

3.4.5 Geometric Camera

I developed a custom camera class, Geometric Camera, for MORSE for the purposes of this project. The intention behind the camera is to provide a single, structured, retinotopic description of the scene from the perspective of the agent (Manny). The camera was placed at approximately the midpoint of Manny's head, facing forward. The camera moves appropriately with robot locomotion, but currently does not adjust to armature movement. For example, rotation of the head does not affect the position of the camera. However, when Manny moves forward, the camera maintains a relative egocentric position. The camera has a focal length of 14mm, giving the agent an approximate 104° field-of-view. Although not formally tested, the focal length of the camera can be adjusted as low as desired. 14mm was chosen as a trade-off between a realistic field-of-view and processing speed.

The camera outputs a dictionary description of the scene primarily organized by screen-coordinate y-values (where y is the vertical plane). Each y-value entry is organized by label of object visible in the scene. Each label is a key associated to a list which contains values for the extension of that particular object in screen-coordinate x-values, the egocentric angle from the camera to the objects at the x,y pair, and the approximate egocentric distance between the camera and the objects intersected at the x,y screen position. Although not formally tested, the accuracy of the camera can be adjusted by passing alternative parameters. The values used as defaults are qualitatively chosen for a trade-off between accuracy and speed. It is important to note that that all the values are approximations and slight inaccuracies are expected. During development, for example, it was common for the wall on the left (see 3.4.3) to be represented as slightly closer to Manny than the wall on the right or vice versa, when in fact, the experimental setup is such that the walls are equidistant from the agent. There is no known theoretical basis why such inaccuracies should affect the results.

Geometric Camera Programming Description

The Geometric Camera is designed to run either in single-processing mode or in multiprocessing mode. Though not tested beyond eight processors, there is no known limit to the number of processors that can be used [3]. The Geometric Camera is the primary source of slow processing as compared to real-time. The vision system was not designed to mimic human vision but, rather, to provide to ACT-R sufficient information about the environment. Therefore the processes in the Geometric Camera are not theoretically driven. The entire scene is outputted to ACT-R so that all timings associated with different types of processing can be handled on the ACT-R side. The Geometric Camera is a child of the Video camera class in MORSE, that is, in turn, a child of the Camera class in Blender.

The Geometric Camera scans the image of the video camera output indexed first by the y-axis (vertical axis) in a retinotopic coordinate system. It uses a built-in Blender function *getScreenRay*, to send out a ray at the x, y coordinate using the Camera's projection matrix. If the ray intersects with an object (it will, unless the camera is looking at nothing), the Blender name of that object is stored (each 3D mesh in blender has a unique ID). The egocentric vector is stored with the name, and the x-coordinate. A cus-

[3] All experiments use 8 processors to reduce vision processing time. No formal test of speed was performed on the vision system. A complete description of the environment described 3.4.3 takes approximately 50 ms on the test computer.

tom depth estimation function determines the approximate distance to that object, providing the length of the egocentric vector. Depth estimation is achieved by increasing the length of a *getScreenRay* vector until an object's Blender name is returned. The scanning function loops across the horizontal (x) plane until either the reaching the end of that plane, in which case the algorithm will proceed to the next y-value; or until a new object ID is detected. If a new object ID is detected, the scan will jump back the appropriate amount of units and carry-out a fine-grained scan to find an accurate edge of the original object. The algorithm repeats until every object visible by the camera has an outline of egocentric vectors that define its edges.

3.4.6 ACT-R Updates

Because the Python variant of ACT-R (T. C. Stewart & West, 2005) is written for Python version 2.7, I updated Python ACT-R to Python 3 in order to support MORSE integration. Although there are tools in Python to convert from Python 2 to Python 3, a fair amount of re-coding had to be done to support functions that are no longer supported or function differently than in previous version. As of yet, no formal tests of Python 3 ACT-R have been completed.

ACT-R-Side Middleware

In order to manage the simulation loop and communication between Python ACT-R and MORSE, I developed a new middleware library for Python 3 ACT-R. The ACT-R side simulation loop controls the simulation tic cycle for both MORSE and ACT-R. An ACT-R-side tic consists of $10ms$ of simulated time, during which all commands for communication between ACT-R and

MORSE are gathered in a dictionary that maps that calling functions to their doppelgangers on the MORSE side. The end of the ACT-R cycle triggers two consecutive MORSE cycles as described in section 3.4.1. During that time, the middleware waits for response from MORSE and sends back the appropriate data to the ACT-R system once MORSE completes its cycles.

ACT-R-Side Vision Module

I developed a new visual module for Python 3 ACT-R (*BlenderVision*) that assumes an input from the Geometric Camera in MORSE. *BlenderVision* stores the scene description provided from the Geometric Camera privately. The dictionary specifying what is visible from the Geometric Camera (see above, 3.4.5) is stored in an private dictionary, *objects*. BlenderVision has arguments for delay time as well as a standard deviation for delay to account for appropriate cognitive processing delays. The module will be in a busy state during the delay period and will return errors to a module that requests information from it while in a *busy* state. These design decisions are based on the Simple Operating System vision module for Python ACT-R (West & Emond, 2005).

The data in the *objects* dictionary is unavailable to the agent until a request for information is published. Because the data has no useable semantic information, methods for detecting 'features' relevant to the project were developed. Different feature detectors can be written for the module but currently only a detector for obstacles and openings is present.

The Obstacle and the Openings Detector The obstacles detector is straightforward, creating visual chunks indicating the egocentric angle and distance of an object, pushing that chunk into the visual buffer. The obstacle detection is used in this project to initiate the rotation. To facilitate that process, a parameter that specifies the proximity of obstacles before they are pushed to the visual buffer is used. For this project, the parameter has been set as a ratio value that is multiplied by the agent's turning radius. Larger agents, therefore, would begin a rotation earlier than smaller agents, relative to the agent's width. Currently, the obstacle detector assumes a flat floor and any size obstacle will trigger the obstacle detection, even if that obstacle could easily be walked over. In order to detect an obstacle, the detector uses as input the value used in the vision request and does an exhaustive search of the dictionary from the geometric camera. The output is chunks, describing the location of the object, in the vision buffer.

The opening detector is designed to parse the *objects* dictionary to find apertures either between separate objects or within an object (e.g. a hole in a wall). It does this by searching the objects dictionary for object edges and evaluating whether the depth between edges is greater than the depth criteria, passed in as a chunk in the vision request. Edges that do not pass the depth criteria are skipped. If an edge meets the depth criteria, a second edge is scanned for. This time the distance between the edges is calculated. If the width between the edges is greater than the width criteria, passed as a chunk in the vision request, the edge is stored as a candidate opening. Once candidate spaces are found, the visual system returns randomly a chunk representation of spaces greater in width, depth, and height, as a specified set of parameters. At a high-level the openings detector is designed to act as a filter and it is assumed that, at a neural level, a highly parallel process would perform such a function. As it is, the openings detector does an exhaustive search through the *objects* dictionary in order to find all openings. In a busy environment this would include the space between small objects, for example, the space between a keyboard and a mouse, a monitor and a mouse, etc. Future or alternate versions of such a detector could be implemented to reduce computational complexity in busy environments.

Motor Module

In section 2.3.1 evidence for *body schemata* was presented. As explained, body schemata play a dual role of representing the body during motor movements and providing biomechanical constraints and representations to support motor simulation. The aim of the motor module is to support body schemata so that an approximation of motor simulation can be implemented.

The motor module maintains a hierarchical, symbolic and numerical representation of the body parts currently being modeled and is synchronized with the 3D body-model (see section 3.4.2). For each body part, there are representations of their degrees of freedom. The motor system monitors for when limbs reach biomechanical constraints, which are set internally, privately in the module. If a biomechanical constraint is reached or a limb has rotated more than any other previous rotation, a chunk relating the posture, the bounding box of the agent, and the action capabilities (stored as chunks in the motor system) is encoded into the agent's declarative memory. Each bone has a set of chunks which describe it's rotation both numerically and symbolically. There are three terms that describe the rotation of the bone: 'none', 'min', and 'max'. At any time one of the descriptive terms is assigned to the bones chunks describing the three principle axes. The terms, 'min' and 'max' are used when the bone are rotated to their biomechanical constraints on a given axis of rotation.

The motor module uses a single buffer for relaying information such as a posture description for storage of body schema or bounding box estimates. There is currently no strong cognitive theory implemented for how the bounding box is calculated. Implementation-wise, bounding boxes are simply calculated for the 3D agent model, and relayed to ACT-R, stored privately in the motor module until requested. Requests for body schema chunks can be made in a similar manner to declarative memory, though this feature is not used in this project. Requests can also be made with slots-value pairs *type:properioceptive* and *feature:bounding_box* to retrieve width, depth, and height information. This is how the body's size information is retrieved for geometric comparison.

3.5 Aperture Model

This section will describe the ACT-R model designed to carry out the aperture tasks found in Warren and Whang (1987) and Higuchi et al. (2012). Generally the model is the same across all tasks modeled. Any differences between the models will be discussed in the sections specific to the experiments that required changes. Note that, while above, a description of the ACT-R system is given, this section describes the model that uses the modules describe therein. Programmatically, the model has two visual systems, one for directed attention, a top-down model; and one for buttom-up alerting when no specific request for visual information is coming from the production system.

Figure 3.4 presents a diagram of the overall process model. The geometric properties are extracted from visual processes and compared, through the central production system in ACT-R, to the geometric properties of the body schema. This process can be cyclical, leading to greater relaxation of the desired body schema. If a body schema is selected, that body schema represents a complementary action. The body schema can also be further constrained by action constraints coming from the goal module (such as the desire to walk, run, etc.).

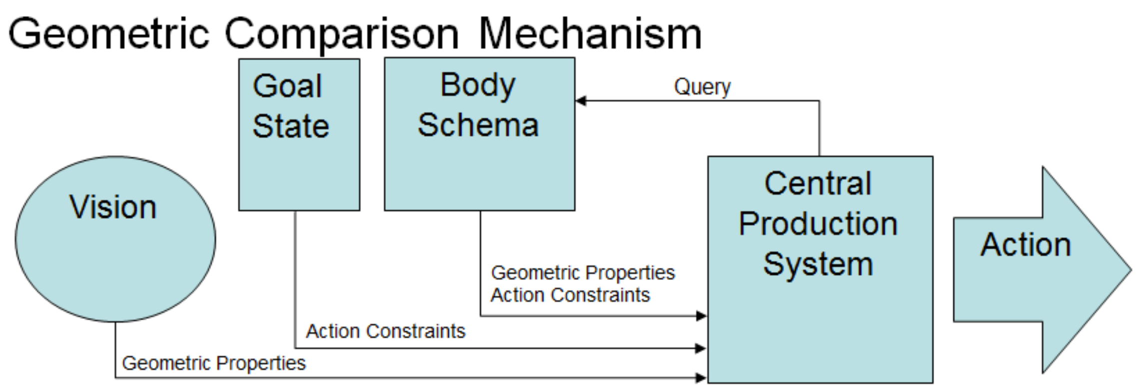

Figure 3.4. Diagram of the process model.

3.5.1 SGOMS

The model uses the SGOMS modeling framework (Somers & West, 2013; West & Somers, 2011; West & Pronovost, 2009; West & Nagy, 2007) for modeling complex tasks. SGOMS is an extension of Goals Operators Methods and Selection rules (GOMS) (Card, Moran, & Newell, 1980) for sociotechnical systems by including a planning unit (West & Nagy, 2007) as the highest level of cognitive control. In its implementation in ACT-R, SGOMS imposes a hierarchical task structure implemented as buffers. SGOMS is realized in ACT-R through the inclusion of planning unit, unit task, and operator buffers. A brief description of the functional role of those buffers is provided below.

Operators

Operators represent low-level actions, functionally equivalent to simple productions in ACT-R. There is a unique operator for each unique production. Implementation-wise, there is a separate buffer to support operators. The contents of the operator is set top-down from unit tasks.

Unit Task

A unit task in SGOMS has a functional role of tying together coherent sets of operators (or subtasks) and are the highest level of control in GOMS (Card et al., 1980). Operators of a single unit task are coherently grouped based on expected interruption. A set of subtasks that are expected to be carried out in sequence form a coherent group. In GOMS and SGOMS both, the unit tasks have the benefit of breaking down complex tasks in a manner that avoids overload and downtime (Card et al., 1980; West & Nagy, 2007). Like operators, unit tasks are realized as productions that match a unit task buffer. The unit task buffer is set top-down from the planning unit.

Planning Unit

The planning unit is the highest and slowest level of control in the SGOMS framework. They consist of a set of unit tasks. Like operators and unit tasks, planning units are implemented with a dedicated buffer. Collectively the operators, unit tasks, and planning units play the same functional role as the goal module in ACT-R 6.0. What makes the planning unit distinct from unit tasks is that the mapping between planning unit and unit tasks is located in the declarative memory, whereas the mapping between unit tasks and operators is considered to be procedural knowledge (encoded in the productions).

One of the key benefits of implementing the SGOMS framework is to support both task switch (changing planning units) and interruption/resumption. Changing the contents of the planning unit buffer will result in a new plan being carried out. The contents of the planning unit can also be stored in declarative memory upon interruption. A task can be resumed so long as the planning unit can be retrieved from memory. SGOMS is used in the model for this thesis under the assumption that the start of the motor behaviour is triggered bottom-up from vision as visual obstacles come close enough to require reaction.

3.5.2 Bottom-Up Vision System

The buttom-up vision system, inspired by previous work by Somers and West (2013) is implemented as a production system [4]. Somers and West used such a system for motion detection in a model of counter steer. In the present work it is used to detect obstacles that are within an area of immediate concern.

From a cognitive perspective there is little need to represent distant obstacles until that obstacle is in a proximity that requires some reaction. Following that assumption, the bottom-up visual system's functional equivalent of a goal buffer is set by the central production system to scan for task-relevant items. In this model, once the agent begins to walk to the aperture, the buffer is set to scan for obstacles (in this case, interpreted as the walls that make up an aperture). The knowledge of what features to scan for is

[4] Note that Python ACT-R allows for productions in modules and this should be thought of as an alternative way of programming modules, not a departure from standard ACT-R.

then encoded in the productions.

In this case, the bottom-up vision system's productions scan for obstacles that are within a multiple of the rotation radius of the agent. There is a *RadiusMultiplier* parameter that is used in this model to affect the distance at which the model starts to rotate. The bottom-up vision system gives a request to the vision module for obstacles in the visual field, within the distance specified by the rotation radius (scaled by *RadiusMultiplier*). The obstacle detector described in section 3.4.6 will push chunks into the visual buffer (shared between the central production system and the bottom-up system). As currently programmed, the bottom-up visual system will modify the central production system's planning, unit task, and operator buffer, leading to an interruption, when an obstacle is detected.

3.5.3 Motor Methods

Motor control is implemented in this model using multiple production systems, dedicated to different muscle groups. This approach is inspired by the approach in ACT-R/E (Trafton & Harrison, 2011) that uses different buffers for different muscle groups. For this project there are only two motor method production systems: one to control the walking and one to control the shoulder rotations[5]. Depending on the model, the motor module productions for rotation have different levels of complexity. At at high-level of detail, all versions of the model send rotation commands to the motor module and continually cycles until the buffer is changed.

The motor methods are not entirely autonomous. In fact, they are not

[5]Like the visual system, the fact that the motor methods are production systems is just an implementation detail. Python ACT-R supports embedding production systems into modules.

very different from standard ACT-R 6.0 implementation as they still require the central production system to start them (by setting their buffers). Python ACT-R makes it very easy to control the speed of the production cycle. In this project, both motor methods' cycle at a $10ms$ production cycle (which is the fastest supported by the middleware). For models 1 and 2 (see section 4) there is very little theoretical cognitive commitment regarding such a fast production cycle because movement is essentially passive as the initiation and cessation is entirely controlled by the central production system. Model 3 explores a tight coupling between the vision system and the motor system for autonomous monitoring of the rotation degree. This tight coupling is inspired by grasping research (Glover & Dixon, 2002; Aglioti et al., 1995) where patient, DF, initially programs the aperture of her fingers incorrectly for grasping, but changes the grasping aperture in-flight to the appropriate grasping aperture.

3.5.4 Overall Design

Each model carries out a pre-experimental phase during which full-rotations are carried out quickly left to right and vice-versa multiple times, encoding the body schema, action capabilities, and bounding box information at the biomechanical constraints, in each direction, in the declarative memory. Without a memory of past rotations, all models will judge as passable only those apertures that are larger than its frontal width because the model has no declarative knowledge of its action capabilities.

Given the experimental setup in Warren and Whang (1987) and in Higuchi et al. (2012), it is assumed that the agent will have a goal state of walking through an aperture. However, at the beginning of the simulation experiment, the agent does not know whether there is an aperture or where it is in the simulation environment. Assuming, however, that the instructions are to walk towards the the aperture, the agent begins walking and motor methods continue the agent walking. During this phase, the planning unit is to walk through the aperture, and walking initiation and scanning for the aperture are unit tasks represented in the unit task buffer.

There are three possible scenarios that the agent has to face in the experiment and I will describe as the: *large*, *medium*, and *small* aperture scenarios. In the *large* scenario the aperture is sufficiently large to accommodate the agent, not requiring the agent to rotate its shoulders. In the *medium* scenario, the aperture's width is in a range to accommodate the agent should the agent rotate its shoulders. Finally, in the *small* scenario, the aperture is too narrow to accommodate the agent, regardless if the agent fully rotates its shoulders. The high-level goal of all agents before presented with the scenarios is just to walk forward. Each scenario represents the conditions under which different unit tasks and operators are selected. How the geometric affordance is realized is a result of those conditions.

In the *large* case, a vision module request for an 'opening' feature is made, using as input the agent's current body size, as retrieved from the motor system. The vision module filters openings (there is only one in the experimental conditions) for openings greater than the input parameters (the agents size) and, in the scenario described, chunks describing the aperture are pushed into the visual buffer. No motor plan is required and the agent walks through the aperture without incident.

In the *medium* case, a similar vision module request is made, however, because the aperture is smaller than the agent's frontal width, the vision request fails. It is in this scenario where the body schemata are required. In this scenario a request to memory is made for a posture that affords walking but has a minimal width value. This process will result in chunks describing a body schema with the shoulder fully rotated either to the left or the right being pushed into the motor buffer. A vision request for an 'opening' is made a second time with the dimensions of the recalled body posture as input to the filter. Chunks related to the aperture are pushed into the vision buffer and the agent continues towards the aperture.

The *small* scenario precedes as the *medium* scenario with the exception that the second vision request (after a body schema is retrieved) fails. This failure results in the agents judgment that they cannot pass through the aperture, under the conditions of the experiment. The planning unit buffer is cleared, the agent stops walking, and the trial ends.

Regardless of whether an aperture is found as a result of the agent's current posture (*large*) or a recalled posture (*medium*), the central production system pushes chunks into the bottom-up vision methods to perform bottom-up obstacle detection as the agent approaches the aperture and the planning unit is changed to manage the walk through the aperture. To perform bottom-up object detection, vision methods loop requests to the vision module to find obstacles within a multiple of the agent's turning radius (after some exploration, this value was kept at 3.0 for all experiments). A single production that will catch an obstacle in the visual buffer is constantly running during this period.

Detection of an aperture begins with the detection of the aperture's edges as obstacles. Detection of an obstacle to either the left or the right will trigger the central production system to re-scan for an aperture, following the same process described above for originally detecting the aperture in scenarios *large*, *medium*, or *small*. So long as an aperture is still detected, chunks are pushed to the buffer in the motor module which triggers the start of rotation. Direction of rotation is counter to the detected obstacle, which is randomized to the left or right, depending on if both edges of the aperture are detected simultaneously.[6]

Kinematic Assumption

One of the main assumptions across all models (unless otherwise specified in the relevant section), is that they rotate with a constant and instantaneous velocity of 120° per second. This value was approximated from walking through apertures and estimating shoulder rotation. The only known aperture-passage study to report on the kinematics of shoulder rotation through apertures is the work by Fath and Fajen (2011), where participants were immersed in virtual environments. Even in that study the timing of shoulder rotation is only reported with approximate values. Fath and Fajen reported participants initiating shoulder rotation between 0.5 and 0.7s before reaching the aperture with ranges of rotation approximately between 20°and 60°. Rotation of 60° in 0.5s is equal to 120° per second.

The kinematic details of shoulder rotation are a constraint on the model and therefore can be considered a large assumption. The mechanics of the rotation have two factors: rotation rate and time. Rotation rate can be measured in different conditions (e.g. aperture width, agent speed) to see how those conditions affect the rate (if at all). Time is affected by the amount of distance between the agent and aperture at the start of rotation

[6]Note: For simplicity, this feature was turned off during all experiments. During experiments, the agent only rotates toward the left.

as well as the agent's speed.

The kinematic details of shoulder rotation can be considered a major assumption because of the timing accuracy possible in ACT-R 3D. A more accurate kinematic model of shoulder rotation can plausibly aid in making inferences about the cognitive processes involved by providing constant while modifying the processing steps in the model.

Rotation Cessation

The strategy for rotation cessation is one of the features explored between models (Section 4) in this thesis. Part of the reasoning behind that is that it is unclear what is involved in ending the rotation. Regardless of model, like the instantaneous velocity, rotation is stopped instantaneously once the motor method production fires. There is little discussion in the literature as to what mechanism controls the extent of the rotation. Higuchi et al. (Higuchi et al., 2012) theorize that the central nervous system creates a constant spatial margin between the shoulders and the edges of the aperture. How that spatial margin is achieved is unclear.

One possibility is that the same process that controls the judgment of whether an aperture is passable is used to judge whether rotation should continue. Recall from above that there is a parameter which is multiplied by the agent's width to over estimate body size. This parameter is inspired by Warren and Whang's (1987) finding of a critical ratio of 1.3. If the same process is used to judge aperture passability then the parameter is re-used when judging rotation termination.

A second possibility extends from Higuchi et al. (2012) in which the central nervous system creates a safety margin. In this scenario, the processes for initial aperture judgment and rotation termination would differ such that instead of multiplying the body with by a factor, a constant can be added the body width. Higuchi et al. propose that perhaps the central nervous system calculates the gap, however, a functionally equivalent approach is to simply add a constant to the body width with the model staying otherwise the same.

Note also that there are a number of modeling options in terms of implementing either of the above two strategies. Either option can be handled through the central production system or plausibly from a lower level in a vision-action loop, as inspired by the grasping literature (e.g. Aglioti et al., 1995).

One possibility not explored in this thesis is used by Ugur and Sahin (Ugur & Sahin, 2010) in their autonomous robot. Over a period of learning the robot associates perceptual vectors to action equivalence classes. A plausible story is that certain widths become associated over time with degrees of rotation that resulted in successful passage. This type of approach is also functionally similar to a highly planned approach where motor system planning simply plans in advance the degree of rotation. These possibilities will not be explored in this thesis.

Chapter 4: Experiments

4.1 Model 1

As described in section 3.5.4, there are a number of options of how to model rotation control. This model is designed explicitly to model Warren and Whang's (1987) first experiment. This model is a partial test of ($H2$), that the input to the geometric comparison mechanism differs between the judgment phase and the rotation phase. In this model the input to the geometric camera is the same in both cases: it uses a model parameter, *VisionMultiplier*, that multiplies the agent's width by a factor, and uses the result as input.

4.1.1 Experiment 1

The purpose of the first experiment is to test how well the model predicts the rotation results in Warren and Whang's (1987) first experiment. As described in section 2.3.5, in their first experiment, Warren and Whang had participants (groups: small vs. large) walk through apertures of different sizes in two speed conditions (normal vs. fast). They found that participants rotated more in response to smaller apertures, that larger participants rotated more than smaller participants, and faster speeds resulted in larger shoulder rotations (all main effects).

4.1.2 Method

Parameter Search

There are three main parameters used throughout the models. The first, *RadiusMultiplier*, is used to control the distance from the aperture before initiating rotation. The bottom-up vision system's obstacle detection requires some metric to filter out obstacles that the agent need not attend to. One of the early planned evaluation criteria of the model was a qualitative assessment of the overall look of the agent's performance with respect to when rotation begins. An unrealistic look would involve the agent first rotating it's shoulders and then walking through the aperture as opposed to rotating at the moment of passage. However, as the project increasingly became about matching rotation measures, the focus became more concentrated towards judgment, both of passage and rotation termination. While distance to aperture is undoubtedly a factor in human performance, the only known source of data is the research by Fath and Fajen (2011). However, because their research is more a test of the vision system, their experiments are not being modeled. Future work could involve measuring distance-to-aperture as a factor of agent size and speed.

The second parameter is *VisionMultiplier*. This parameter is used as a scaling factor to force an overestimation of the agent's width. This is to account for the reported rotation means in Warren and Whang's study that shows agents of $40cm$ rotating their shoulders for apertures of $50cm$. This parameter was fit by having small and large agents pass through and rotate for apertures of $40cm$ and $55cm$ in width.

The final parameter was RotationRate that controls the agents rotation rate in degrees per second. A small set of rotation rates, within the range reported by Fath and Fajen (2011) were used in the overall parameter search (80°per second to 120°per second).

A number of *RotationRate* x *RadiusMultiplier* x *VisionMultiplier* parameter searches were run starting with large gaps between parameter values in order to narrow down reasonable parameter ranges. The overall absolute, mean rotation difference between model and human data in the *normal* speed condition of the final parameter fit is presented in Table 4.1. In the parameter search there were 5 agents per group condition and agents sizes were chosen from a normal distribution around the means for each group ($40.4cm$ for small and $48.4cm$ for large) with a standard deviation as reported ($SD = 2.0cm$ for small and $SD = 0.7cm$ for large). The aim of the parameter search was not to achieve best fit possible but just to bring rotation values to an reasonable ball park as actual rotation rates and rotation model could not be empirically validated. The parameter search was limited to aperture widths $40cm$ and $55cm$ and included $RadiusMultiplier = 2.5$ and 3.0, $VisionMultiplier =$ 1.135, 1.137, and 1.139; normal walking speed ($1.28m/s$); and $RotationRate$ = 100 and 120 degrees per second. Agents went through each aperture 10 times, for a total of 10 (trials) * 3 (VisionMuliplier) * 2 (RadiusMultiplier) * 2 (apertures) * 10 (agents) * 2 (RotationRate) = 2400 simulations.

Because the simulations are time-synchronized, and MORSE runs in real-time, overall the simulations run slower than real-time. Because of this, there was some practical concerns regarding running experimental simulations. Byrne (2013) suggests a method for estimating model runs to achieve desired confidence interval, based on the models standard deviation. One issue with this method, brought up by Byrne, is that we often cannot know *a priori* what the standard deviation of the model will be. He suggests two possible solutions: the first, to estimate based on the human participants and, the second, to run simulations (he suggests to start with 20) and estimate the standard deviation from there. Even that approach is problematic in this case it is not clear that there are identical distributions at each aperture ratio, which is one of the major assumptions in his approach.

Furthermore, his approach assumes that the standard deviation scales with the mean, such that there is more variance with greater means. For example, reaction times of 50ms will likely have less variance than reaction times with means of 500ms. It is actually clear, in hindsight, looking at Figure 4.3 that the complete opposite is true. Because human data was not available from the Warren and Whange (1987) experiment, it was decided that calculating the number of simulation runs was not necessary.

A further complication in this study that it is not just a matter of performing Monte Carlo simulations. Presumably one of the major sources of variance in the human studies was the size of the agents. We know, for example, that small agents in Warren and Whang's study (1987) have a smaller degree of rotation than do large agents, for the same apertures. This gives reason to believe that even within a size group, a fair amount of variance is due to the contribution of different sized agents.

Simulation Runs

Due to practical concerns of simulation running time, models were only run through apertures up to 70cm. When apertures are 50cm or greater for small agents or 60cm or larger for large agents, there is little variation in rotation degree (due, presumably to a floor effect). Given that it is reasonable to expected that the last two apertures would produce flat results, there would be no information gained from running the simulations.

The experimental conditions (including participant sizes) were created as accurately as possible in the simulation environment. The environment and the agent descriptions are given in section 3.1. In the simulation there were 5 agents per group condition and agents sizes were chosen from a normal

Table 4.1

Summary of Parameter Values and Evaluation Criteria, Model 1 Experiment 1

Rate (*deg./s*)	RM	VM	AP	Small	Large	Abs. Diff.
100	2.5	1.135	40	51.1	59.5	
100	2.5	1.135	55	14.9	29.26	33.4
120	2.5	1.135	40	54.3	66.6	
120	2.5	1.135	55	17.1	31.4	29.4
100	2.5	1.137	40	50.6	59.4	
100	2.5	1.137	55	15.2	30.2	32.1
120	2.5	1.137	40	53.5	68.2	
120	2.5	1.137	55	18.3	31.6	27.9
100	2.5	1.139	40	51.4	60.2	
100	2.5	1.139	55	14.8	30.9	31.0
120	2.5	1.139	40	53.5	66.7	
120	2.5	1.139	55	34.6	49.9	53.3
100	3.0	1.135	40	35.9	51.2	
100	3.0	1.135	55	16.0	27.4	53.5
120	3.0	1.135	40	56.0	75.0	
120	3.0	1.135	55	18.1	30.6	24.5
100	3.0	1.137	40	54.1	68.6	
100	3.0	1.137	55	14.4	27.7	28.2
120	3.0	1.137	40	55.5	74.9	
120	3.0	1.137	55	18.0	31.7	22.9
100	3.0	1.139	40	55.5	69.5	
100	3.0	1.139	55	15.1	30.0	27.1
120	3.0	1.139	40	56.0	75.7	
120	3.0	1.139	55	18.8	35.4	19.6*

Note. RM = RadiusMultiplier; VM = VisionMultiplier; AP = Aperture Width (cm); Abs. Diff. = The combined absolute difference between the mean rotation of the model and the mean human rotation combined for both large and small at apertures $40cm$ and $50cm$. The smallest difference is the most favorable. *Rotation rate of 120 deg./s was chosen as the overall default rotation rate.

distribution around the means for each group (40.4cm for small and 48.4cm for large) with a standard deviation as reported ($SD = 2.0cm$ for small and $SD = 0.7cm$ for large). Agent sizes by group are presented in table 4.2. Agents walked at the average speeds per group as reported in Warren and Whang: 1.29m/s and 1.61m/s for the normal and fast conditions (respectively) in the small group and 1.28m/s and 1.77m/s for the normal and fast conditions (respectively) in the large group. Each walked through each aperture a total of 60 times, for a total of 60 * 10 (agents) * 2 (speed) * 5 (apertures) = 6000 simulation runs.

Table 4.2

Agent Sizes by Group, Model 1 Experiment 1

Small (cm)	Large (cm)
39.81	48.07
39.95	48.16
41.14	49.02
41.33	49.36
41.41	49.53

4.1.3 Analysis

Because the original data from Warren and Whang's original paper was not available[1], only a minimal comparison between the model and human data was possible. A visual comparison between the Warren and Whang's results and the model results are illustrated in Figures 4.1 (normal speed condition) and 4.2 (fast speed condition).

The model data (blue and red lines) show a reasonable fit with the human data, with an obvious over rotation at larger apertures for both the small

[1]Because the paper is nearly 30 years old, this was expected. William Warren was contacted and was surprisingly able to provide some material. The material, however, was not relevant for comparative statistical analysis.

Mean Rotation Rate by Aperture Width (Normal Condition)

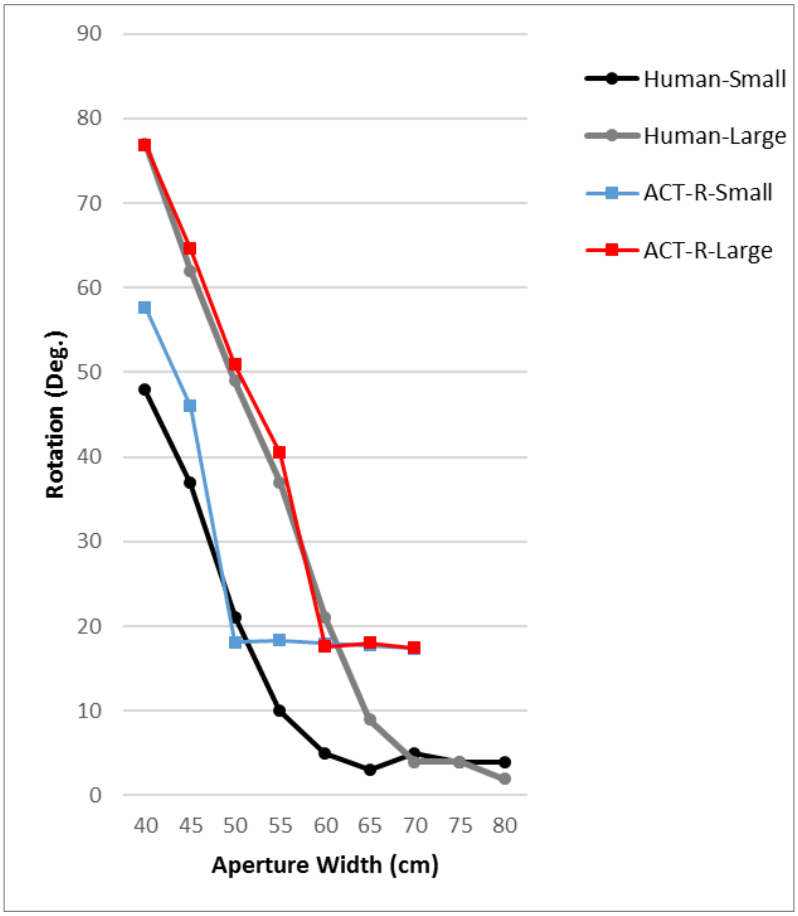

Figure 4.1. Human vs. Model rotation angle (normal speed condition). Black line and gray line represent small and large human performance (respectively). Blue and red lines represent small and large ACT-R performance (respectively).

(blue) and large (red) agents.

A Pearson's correlation between the mean of the human data and the means of the model was conducted. In the normal speed condition, the squared correlation was 0.89 for the small agents and 0.97 for the large agents in the normal condition. In the fast speed condition, the squared correlation was 0.89 and 0.92 for the small and large agents, respectively. An

Mean Rotation Rate by Aperture Width (Fast Condition)

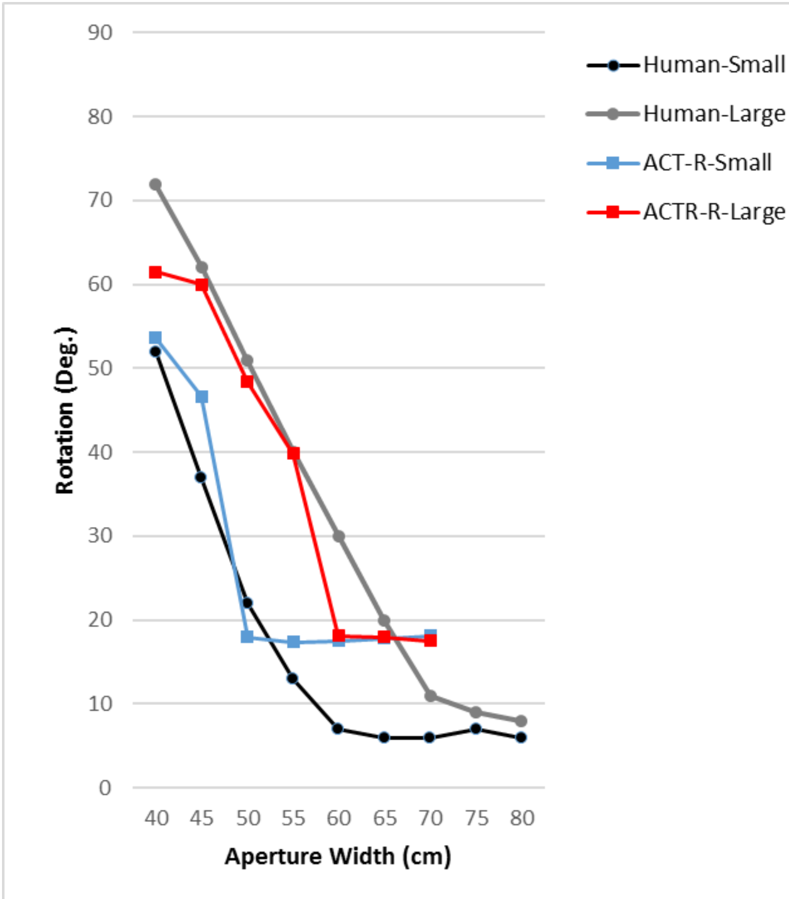

Figure 4.2. Human vs. Model rotation angle (fast condition). Black line and gray line represent small and large human performance (respectively). Blue and red lines represent small and large ACT-R performance (respectively).

RMSE of the means revealed a average error of 8.82 degrees of both large and small agents in the normal condition; and 7.74 degrees of both large and small agents in the fast condition. In addition to the comparative statistics, an ANOVA was ran on the model to see if the main effects as described in Warren and Whang's experiment were replicated. The ANOVA showed that large participants had larger degrees of rotation than smaller participants.

Participants rotated more for narrower apertures. Unintuitively, and opposite to that found in Warren and Whang (1987), agents rotated less in the fast speed condition compared to normal speed condition. All main effects (Group, Speed, and Aperture) were main effects, ps < 0.01.

Despite the main effect of speed in the opposite from expected direction, as illustrated in Figure 4.3, that effect appears to be dominated by an under rotation at aperture width of $40cm$. Otherwise, the means are more or less congruent.

Discussion

Visually, the model does a decent job predicting the data in both cases, though it exhibits an obvious over-rotation at large aperture widths. This over-rotation is likely due to the *VisionMultiplier* parameter. The agent likely overestimates its width, and rotates when it need not. Rotation reaches nearly $20°$ before rotation is stopped.

Although the ANOVA shows a main effect of speed, as mentioned above, that main effect is in the wrong direction, when compared to human data, with the model under rotating in the fast condition. However, when examining the data further, aperture by speed shows congruent results across speed conditions, except for in the smallest aperture ($40cm$). This suggests that the model does not account for speed effects. In hindsight, this is not at all surprising, as there is no factor that explicitly accounts for speed. This begins to address ($Q3_b$), suggesting that speed is an explicitly planned factor. However, without more empirical evidence, it is unclear whether overall rotation angle at higher speeds is a result of faster rotation rate or some other factor affecting rotation termination.

Mean Rotation at Different Speeds (Model 1)

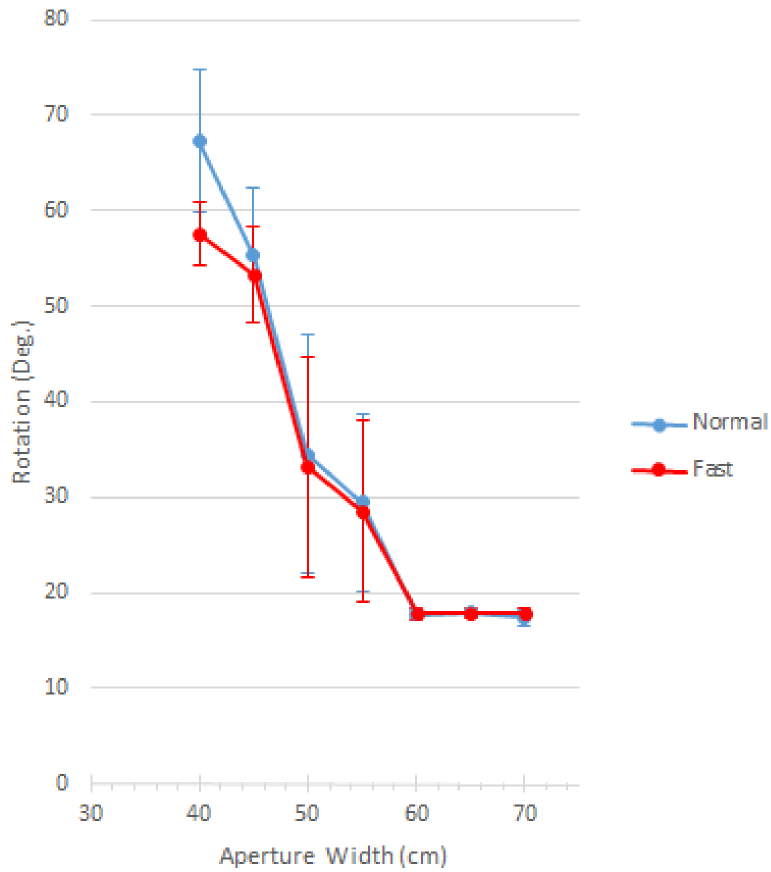

Figure 4.3. Combined (small and large) rotation by speed (Model 1) shows that both normal and fast speeds are congruent except for the smallest aperture (40cm). Error bars indicate 95% confidence intervals.

4.1.4 Experiment 2

The second experiment is meant to be a test of a common implicit or explicit assumption in previous aperture-passage literature. Many of the authors assume the theoretical commitments of Warren and Whang (1987), that aperture passage is directly perceived in units of eye-height. Some authors (e.g. Higuchi et al., 2012; Fath & Fajen, 2011) even go as far as using aperture-width to shoulder-width ratios as their independent variables, choosing ratios in and around the range of the critical ratio (A/S) reported by Warren and Whang (1.3). In Higuchi et. al (2012), this is somewhat puzzling because they offer their own theory on how rotation termination is modulated.

Warren and Whang (1987) imply that rotation control is heavily top-down. Although they do not offer any description of the processing involved in rotation control their theory suggests that rotation control is a result of a plan that is largely modulated by the A/S ratio. Higuchi et al. (2012), however, propose that the central nervous system controls rotation by maintaining a constant safety margin between the shoulders and the edges of the aperture. This is somewhat puzzling because they also use as one of their independent variables aperture ratios instead of absolute aperture widths. Their motivation is based on the idea that using A/S ratio to modulate rotation becomes inefficient with really large agent widths.

Higuchi et al.'s (2012) study is somewhat more complex and involves a somewhat different task than in Warren and Whang (1987). In the Higuchi et al. study, participants walk through apertures while carrying bars of varying lengths. Higuchi et al. consider the participants and the bar to form an agent-plus-bar system (a common view within the literature). Their hypothesis, that the central nervous system maintains a constant safety margin, is tested in this study by varying bar length as an independent variable. By exaggerating the length of the bar they are able to test whether rotation is extremely exaggerated (as would be the case if an A/S ration of 1.3 was used). The reasoning is as follows:

Consider an agent, $40cm$ in width. From Warren and Whang (1987) we know that an agent would rotate their shoulders at an A/S ratio of 1.3, meaning that a 6 *cm* spatial margin (on both sides) is required for passage (presumably a safety margin). If that same person was carrying a bar 100 *cm* in length and rotated based on the same ratio, then they would create a 15 *cm* safety margin ($(100 * 1.3 - 100)/2 = 15$). This over-rotation is markedly inefficient (Higuchi et al., 2012)

Higuchi et al. propose instead that the CNS controls the rotation in each case to maintain a constant safety margin. Assuming a safety margin of 6 *cm*, an A/S ratio of 1.12 ($(100 * 1.12 - 100)/2 = 6$) should only be required for safe passage. That is, participants should only begin rotating their shoulders when the ratio between the aperture and themselves (including a bar extending their length) is 1.12. Their hypothesis predicts the following:

1. The amplitude of rotation should become smaller as width increases and A/S ratio maintained.

2. Spatial margins should remain constant regardless of absolute size or A/S ratio.

In their experiment they manipulated aperture ratio and agent width by having the participants carry bars that modify their frontal width by either a factor of 1.5 or 2.5 (as well as a control condition). Aperture widths are set to create ratios of 0.9, 1.0, and 1.1 to encourage a large rotation. With respect to amplitude of rotation, the authors found a main effect of bar length such that the angle of rotation was smaller as bar length increased (addressing 1). With respect to spatial margin the found a main effect of bar length.

Model and Parameters

The model is the exact same model used in Experiment 1. The 3D model is somewhat different in that it has a bar at a constant position in front of the agent. The bar is attached to the agent such that the bar rotates with the agent, around an arch centered on the agent. Other than the experimental differences (the sizes of the agent are different in Haguchi et al. as compared to Warren and Whange (1987)), the agent and model are exactly the same, including the parameters as described in Experiment 1. The hypothesis for this experiment is that the rotation control when walking through the aperture is the same regardless of whether a bar is being held or not (addressing ($H3$)).

Simulations

Analysis

Figure 4.4 shows the absolute rotation angle for both the model and human participants in the three bar-length conditions (control, 1.5 times, and 2.5 times) and the three aperture ratio conditions (0.9, 1.0, 1.1). (recreated from data from Higuchi et al., 2012). Visually there is an evident over-rotation in comparison to the human data. A Pearson's correlation between the human mean rotations and the model mean rotations was 0.68. A Pearson's correlation between the human mean safety margins and the model mean safety margins was 0.07. There was a RMSE of 23 degrees. The poor fit is also evident from visual inspection. Visually it does seem as though agents rotate less at higher aperture ratios as would be expected from Experiment 1. An aperture ratio by bar ratio ANOVA confirms the effect of bar length (p <

0.01).

Figure 4.5 presents a comparison between the model safety margins and the human safety margins (recreated from data from Higuchi et al., 2012).

Mean Rotation Angle

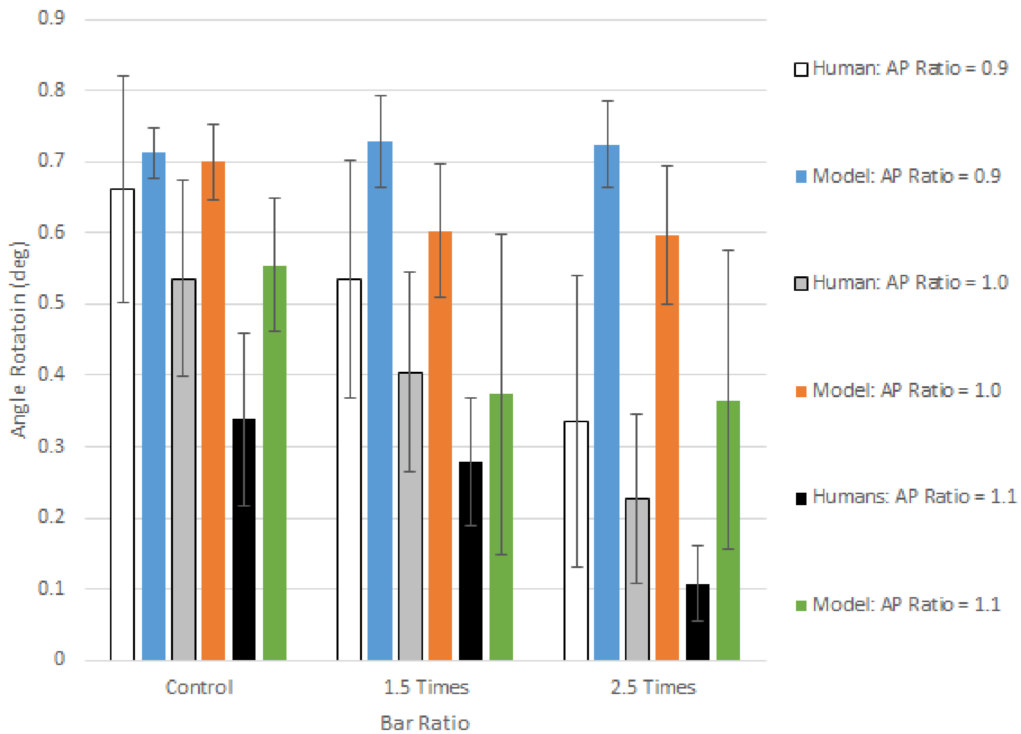

Figure 4.4. Human (grayscale) vs Model (colour) absolute rotation angle for bar ratios: control, 1.5, 2.5 and aperture ratios: 0.9, 1.0, and 1.1. Error bars indicate 95% confidence intervals

4.1.5 Discussion: Experiment One and Two

The model shows a reasonable fit in experiment one for absolute rotation, showing similar effects of aperture and similar effects of body size. The statistics show also an effect of speed in the wrong direction but, largely, that effect is dominated by one aperture size (40cm). In Experiment Two,

Mean Safety Margins

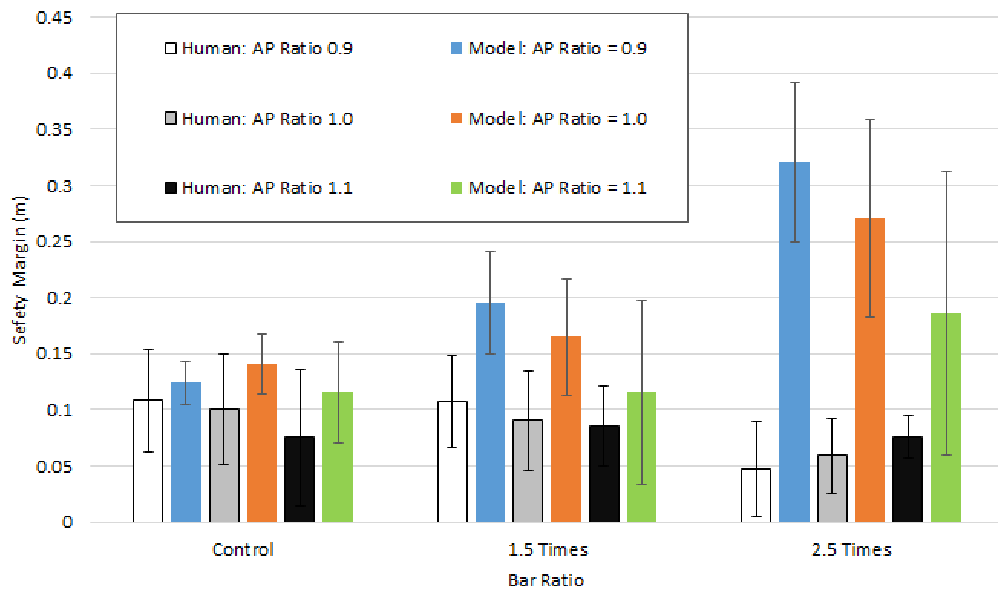

Figure 4.5. Human (grayscale) vs Model (colour) safety margins for bar ratios: control, 1.5, 2.5 and aperture ratios: 0.9, 1.0, and 1.1.

the model performance barely resembles the human data and in regards to mean safety margins shows massive variance in spatial margin, relative to human participants. While variance is also quite high in Figure 4.4, visually the variance is comparable to human participants.

Effect of Speed

Warren and Whang (1987) found that faster participants rotated more than slower participants where as Model 1 showed the opposite (though, largely congruent results in both speed conditions). This was explored to see if the temporal dynamics of the system would function such that the effect of speed would not have to be an explicit factor of the model. As it stands in the first experiment, it looks as if speed might have to be accounted for as part of the motor plan. No model was built to explore this possibility as doing so would simply be fitting a new parameter.

Transfer of Rotation Strategy

The aperture passage literature is ambiguous as to whether rotation strategy changes for different conditions. Although, intuitively from experience we might expect there to be differences when carrying an object through and aperture versus when not carrying an object, the related literature has no clear stance. Haguchi et al. (2012) propose that the constant strategy in rotation cessation is to terminate rotation once a safety margin is reached. Running Model 1 in Experiment 2 was largely a test of whether a rotation strategy based on A/S ratios can explain their results. Although a long-shot it was at least plausible given the stochastic nature of the model. Assuming the model is at all accurate, using an A/S ratio strategy does not transfer between tasks. That does not mean that strategy transfer between tasks is not supported. A second strategy is explored with Model 2.

4.2 Model 2

Model 2 attempts to address the over rotation evident Experiment 2 of Model 1. Specifically, in the second model, there is a change in the representational content used to stop shoulder rotation. Model 1 assumed that the judgment for deciding if it could pass through an aperture (based on A/S ratio) was the same as the judgment used to stop the rotation (i.e. A/S ratio used for judging passability and for judging rotation termination). Model 2 is a modified version of Model 1. While it still uses *VisionMultiplier* as a parameter, it is only used when judging whether an aperture was passable or whether an aperture would require rotation. To implement the theory proposed by Higuchi et al. (2012), a new parameter was introduced, *VisionConstant*. *VisionConstant* is added to the agent's width acting as a mechanism to over estimate body size.

There was also implemented a minor change in how the *feature:openings* detector works in the vision system between Model 1 and Model 2. The over rotation evident in Figures 4.1 and 4.2 at higher apertures would ideally be at zero. Although human participants exhibit a rotation of approximately 5°-10°, Warren and Whang (1987) suggest that those rotation values are likely due to body sway while walking. Since the model does not experience body-sway, rotation values at higher aperture widths should effectively be zero.

4.2.1 Experiment 1

This experiment is exactly the same as described in section 4.1.1

Parameter Search

Because of the model changes a new set of parameter searches was done for this model. Like Model 1, this was done in a series of searches. *VisionConstant*, however, was estimated for practical reasons. A rough estimate of *VisionConstant* was informed by Higuchi et al. (2012) who hypothesis a safety margin of $6cm$. A *VisionConstant* of $3cm$ (per side) is used throughout Model 2. Parameter fit methodology was similar to that reported in section 4.1.2. Because there was less concern about whether or not the agent rotated near the door, *RadiusMultiplier* remained at a value of 3.0. *VisionMultiplier* was set to 1.36. Recall that Model 1 used a *VisionMultiplier* parmeter of

1.139. Just as in Model 1, *RotationRate* was set at 120 degrees per second.

Simulation Runs

The simulation setup was largely the same as Model 1, Experiment 1. The breakdown of simulation runs was different in this experiment than in Model 1, Experiment 1. There were a total of 20 agents, 10 per size group (large and small). Each agent walked through the aperture 15 times. Remaining conditions were otherwise the same for a total of 15 * 20 (agents) * 2 (speed) * 5 (apertures) = 3000 simulation runs.

Analysis

A visual comparison between the Warren and Whang's results and the model results are illustrated in Figures 4.6 (normal speed condition) and 4.7 (fast speed condition). A Pearson's correlation indicate a fit of 0.98 and 0.91 for the small and large agents in the normal speed condition; and 0.98 and 0.92 for small and large agents in the fast speed condition. A combined RMSE for both large and small agents was 8.73 degrees in the normal speed condition and 8.27 in the fast speed condition. In addition to comparative statistics, an ANOVA was run on the model data to see if the same main effects were present in the model data as in the human data. Large participants had larger degrees of rotation than smaller participants. Participants rotated more for narrower apertures. Just as in Model 1, agents rotated less in the fast speed condition than in the slow speed condition. All main effects (Group, Speed, and Aperture) were significant, $ps < 0.01$. The difference in the speed conditions is extremely similar to Model 1, Experiment 1, and deemed unimportant. No data is therefore presented.

Mean Rotation by Aperture Width (Normal Condition)

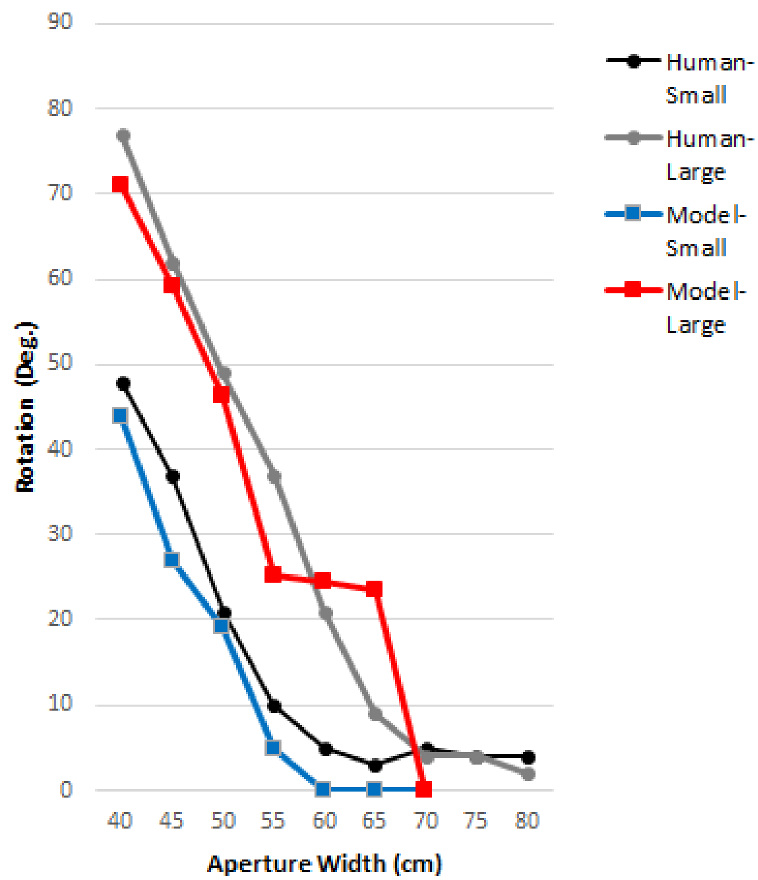

Figure 4.6. Human vs. Model rotation angle (normal speed). Black line and gray line represent small and large human performance (respectively) in the normal speed condition. Blue and red lines represent small and large ACT-R performance (respectively).

Discussion

Model 2 has, visually, a reasonable fit to the human data in both the normal and fast conditions. Group differences between small and large agents is clearly evident. Notice in this model, large aperture values for small agents show no rotation in both speed conditions. Notice also that large agents have

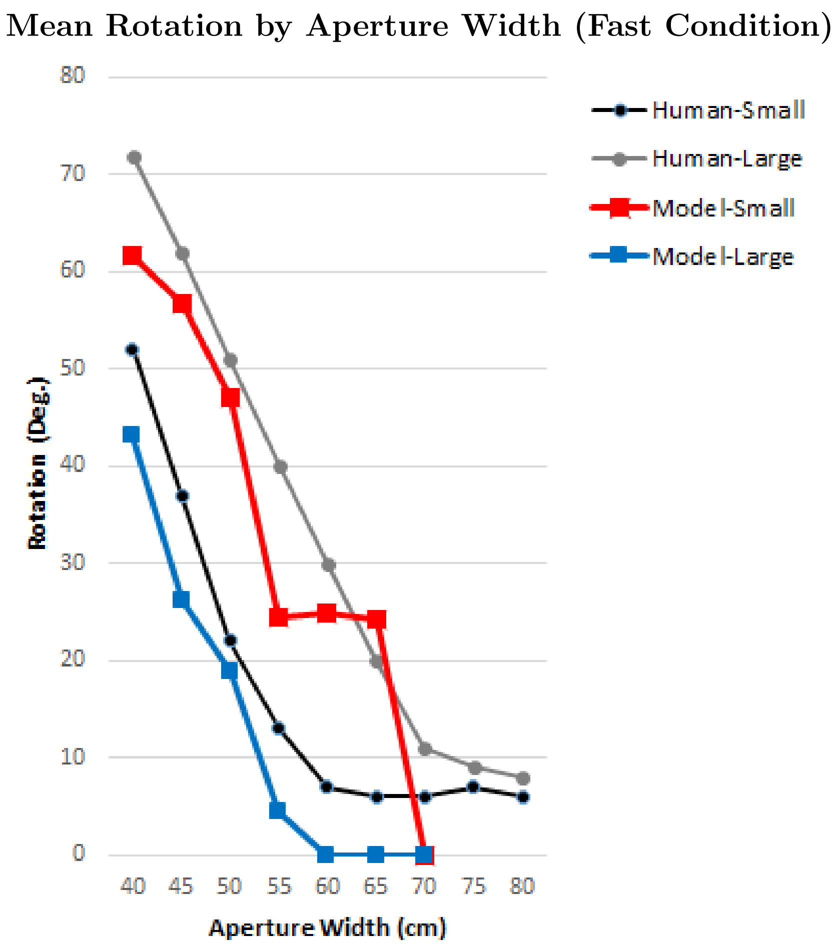

Figure 4.7. Human vs. Model rotation angle (fast speed). Black line and gray line represent small and large human performance (respectively) in the normal speed condition. Blue and red lines represent small and large ACT-R performance (respectively).

a flat mean rotation at apertures 55cm, 60cm, and 65cm (in the range of 25°) before no longer rotating at an aperture of 70cm.

The flat mean rotation between 55cm and 65cm for the large agent is reminiscent of the over rotation from Model 1 which had similar mean values (near 18°, see Figures 4.2 and 4.1). The flat section suggests another source of over rotation. After some investigation it appeared as though the delay in the production system, combined with the admittedly high rotation rate (at the high-end in (Fath & Fajen, 2011)). At 120 degrees per second, even a delay of 200ms can lead to 24 degrees of rotation.

4.2.2 Experiment 2

As with Model 1, performing Experiment 2 for Model 2 was to provide a test of the assumption that overall rotation strategy between the two tasks is the same.

Model and Parameters

As with Model 1 (4.1.4), Experiment 2; Model 2 is exactly the same in Experiment 2 as in Experiment 1, with the exception of bar attached to the 3D agent model.

Simulation Runs

Each agent performed 15 trials of each aperture * bar combination. There were a total of 10 agents, and agent size was selected the same way as described in 4.1.4.

Analysis

Figures 4.8 and 4.9 illustrate absolute mean rotation and mean safety margins for human (grayscale) and model (colour) (respectively) for Experiment 2.

As is evident in Figure 4.8, agents rotated less with larger bars, and rotated less at higher aperture ratios. The effect of bar and aperture ratio are both significant ($ps < 0.01$). A Pearson's correlation indicates a fit of 0.80 for rotation angle and 0.21 for safety margin. Note, however, that the model exhibits a large over rotation in the 2.5 times bar condition. Excluding those conditions, the Person's correlation is 0.84 for absolute rotation and 0.89 with respect to safety margin. For the control and the 1.5 times bar ratio condition, mean absolute rotation is highly comparable to human participants. In bar ratio condition *2.5 times*, there is a large over rotation.

As shown in Figure 4.9, the model has a fairly good fit for mean spatial margin in both the control bar condition and the 1.5 times condition. Agents leave less safety margin between themselves (or the bar) and the aperture edges. Agents in Model 2 leave greater spatial margins when carrying larger bars. All effects are significant ($ps < 0.01$).

RMSE for model 2, excluding the 2.5 times bar condition was approximately 9 degrees of absolute rotation and approximately 2 cm with respect to safety margins.

4.2.3 Experiment 1 and 2 Discussion

Model 2 comes a lot closer to fitting both absolute mean rotation angle and safety margins in Experiment 2. The patterns in mean rotation angle are similar to human data (see Figure 4.8). There remains a rather large over rotation in the *2.5 times* bar condition. One source of over rotation might simply be that agents rotate more cautiously when carrying such a long bar, which would be over a meter in some cases. In light of the over rotation in Experiment 1 for large agents, it is possible that some over rotation might be due to production lag (as discussed in Section 4.2.1).

The spatial margins are comparable in the first two bar conditions (control and 1.5 times) though the mixed results do not indicate anything obvious. The over rotation evident is somewhat surprising. That said, the margins

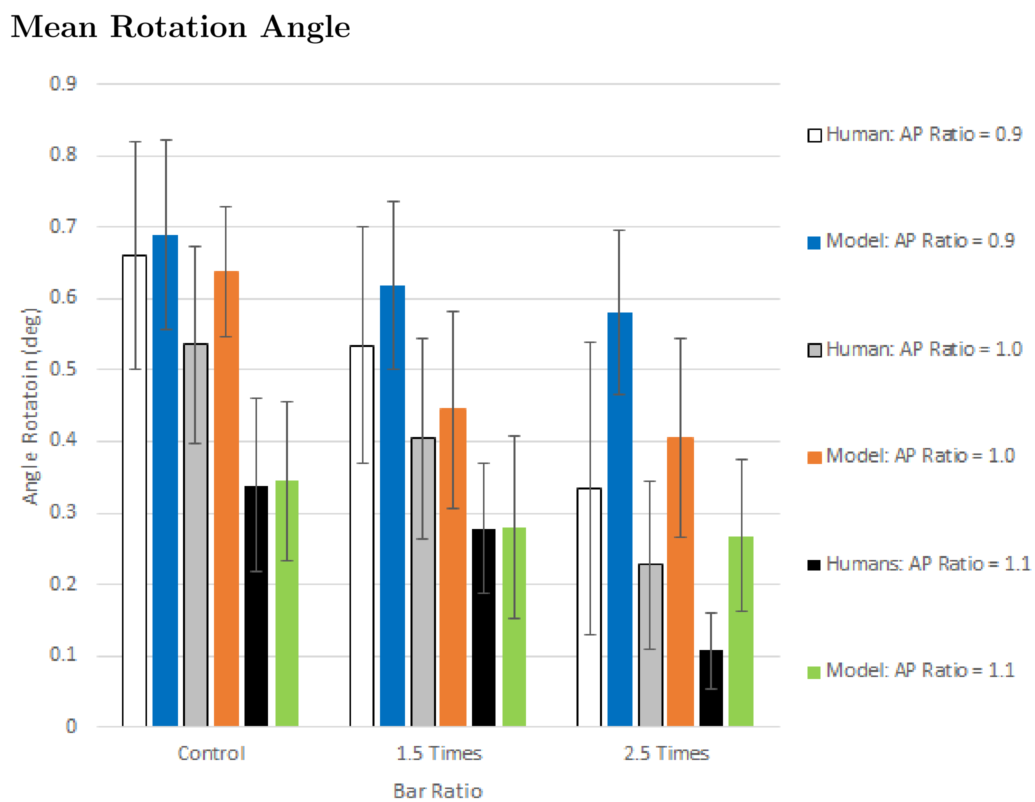

Figure 4.8. Human (grayscale) vs Model (colour) safety margins for bar ratios: control, 1.5, 2.5 and aperture ratios: 0.9, 1.0, and 1.1.

are also remarkably small for human participants at that range, suggesting possibly a change in strategy or perceptual differences. Note absolute aperture widths are in the region of $100cm$, as opposed to a maximum of $70cm$ in Experiment 1. It is possibly that the vision system of the agent is less accurate at those widths, especially as the agent gets closer to the aperture and the visual angle between the edges of the aperture increase dramatically.

4.3 Model 3

I developed a third model as part of the exploratory nature of the thesis. The third model differs from the other two in terms of how the motor system

Mean Safety Margins

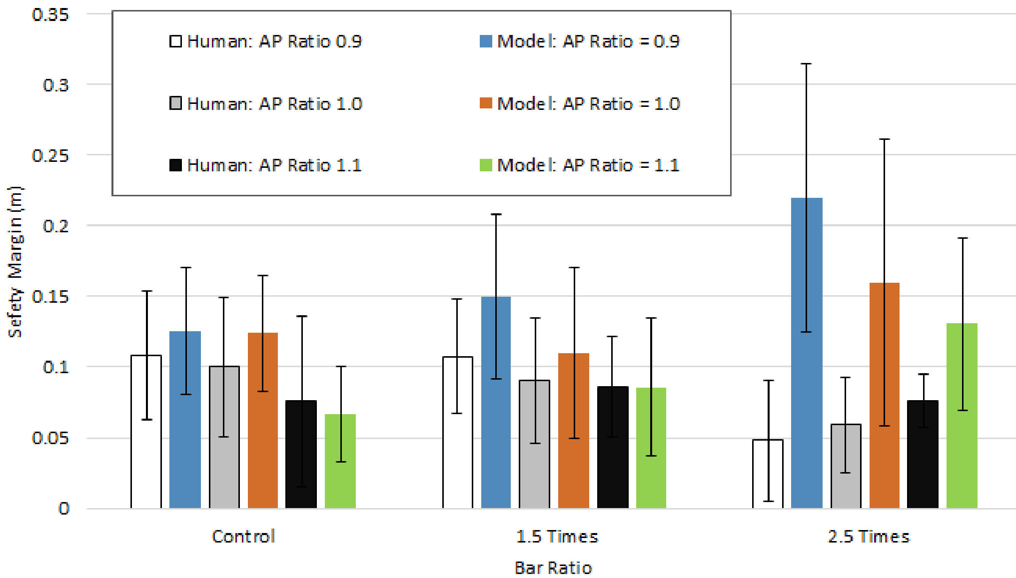

Figure 4.9. Human (grayscale) vs Model (colour) safety margins for bar ratios: control, 1.5, 2.5 and aperture ratios: 0.9, 1.0, and 1.1.

and vision system interact during control of rotation. As described in Section 3.5.3, the third model explores a tight coupling between the motor system and the vision system. The design of Model 3 is inspired partially by Milner and Goodale's *Two Visual Streams Hypothesis* (2011; 1992); particularly supporting research that suggests a high-speed interaction between the vision system and motor system (Glover & Dixon, 2002; Aglioti et al., 1995).

Models 1 and 2 follow standard ACT-R theory insofar as monitoring of the rotation control happens in the central production system. Model 3 is a break from that tradition by giving the motor system direct access to the vision module and performing geometric comparison as part of the motor system's production loop. Since the production rate in the motor system is set at $10ms$, it was expected that the resulting model would have a highly accurate rotation control. Part of the idea was to help reduce the amount of over-rotation evident in Models 1 and 2.

With regards to the geometric comparison mechanism, Model 3 is informed by Model 2 and uses different input between the judgment phase and the rotate phase. Just as in Model 2, it uses the *VisionMultiplier* parameter during the judgment phase and the *VisionConstant* parameter during the rotation phase.

Analysis and Discussion

Figure 4.10 shows some parameter search data for Model 3. Higher *VisionConstant* (VC) values are required for the data to perform at human rotation levels. Although more parameter searches and both Experiment 1 and Experiment 2 were performed, upon closer analysis of the data Model 3 was rejected as an implausible model. Unlike Models 1 and 2, Model 3 shows no variance in rotation performance for the individual agents. That is, although there is variance for a group (e.g. *small agents*), a single agent has the exact same performance from one trial to the next (for a single aperture).

This result suggests that having a direct connection from the motor system to the vision system, at least given this initial investigation, is hard to justify. Although the motor control is potentially more accurate it is simply too accurate. A noise term could introduce some individual variance but doing so would be unparsimonious. Future work could pursue increasing the production rate of the motor system as well as modifying retrieval rates for motor-to-vision-system requests.

Mean Rotation by Aperture Width (Parameter Search, Normal Condition)

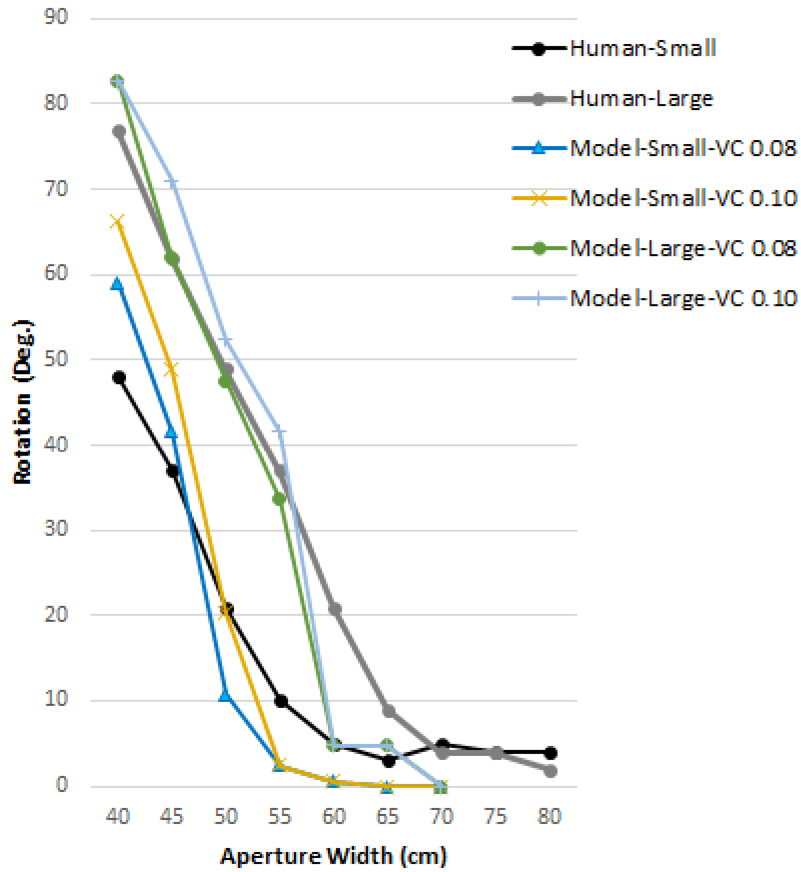

Figure 4.10. Human vs. Model rotation angle (normal speed) parameter search. Black line and gray line represent small and large human performance (respectively) in the normal speed condition. Blue line represents small agents at *VisionConstant* (VC) 0.08, yellow line represents small agents at VC 0.10, green line represents large agents at VC 0.08, and light blue line (cross-hatch) represents large agents at VC 0.10.

4.4 Overall Discussion

4.4.1 Rejection of Model 3

Rejection of Model 3 is based on the argument that adding noise to the system would be unparsimonious. The parsimony, in this case, is evaluated in comparison to both Model 1 and Model 2 that exhibit individual variance in their performance while having a reasonable fit to the data. While adding a noise term, for example, to the vision system, could potentially introduce variance, doing so adds a free parameter that makes no prediction. Exploring adding noise to Model 3 is an option for future work but the predictive success of the noise term should be evaluated in other studies.

4.4.2 Comparative Statistics

Pearson's correlation and RMSE were conducted for Model 1 and Model 2. While these measures are less than ideal (as there is very little account for confidence intervals), they do provide means of comparing the competing models. However, equivalence testing (T. Stewart & West, 2010), would be a preferred method for evaluating the models, especially in determining if they equally predict the human data. However, as it was not clear whether confidence intervals for the Warren and Whang (1987) experiment could be estimated, correlation, error, and an ANOVA were done as an alternative.

From a philosophical perspective, however, it should be noted that these statistical methods are meant to demonstrate that the model's performance is a reasonable fit of the human data. While it may have been possible to make a stronger fit to the model by extending the parameter search, doing so would have been premature.

One of the main reasons a strong parameter search would be premature has to do with the assumption about the rotation rate (120 deg./sec.). Assuming Model 2 is a reasonable model, the rotation rate is clearly to high in the 2.5 times bar condition. Some more complex processing may be involved in planning the rotation. The rotation rate is clearly a large factor but because there is no well reported measure of rotation rates it is a factor that should be measured empirically. Because the rotation rate is such a large factor, a drastically different rate could easily invalidate the model. Future work, therefore, should measure rotation in order to more suitably inform the models.

Chapter 5: Discussion and Future Work

5.1 Competing Theories

In Gibson's original presentation of affordances(1986), he proposes that the environment presents possibilities for action. In his view, affordances are perceived by animals with complementary abilities to perform the action the environment presents. Expanding on the overall Gibsonean view, Michaels and Carello (1981) describe affordances as being analogous to how different animals, with different sensory apparatuses, are sensitive to different physical properties of the environment. For Michaels and Carello, affordances are properties of the environment. Other philosophers (Chemero & Turvey, 2007; Chemero, 2003; Stoffregen, 2003) have since proposed that affordances arise as emergent properties of systems composed of both the animal and its environment. While each of these philosophers have a number of minor differences in their affordance ontologies, a common theme among all of them is a rejection of a strong representational view that is characterized by inference processes that operate on semantically-laden representations, internal to the agent.

Section 3.2.1 describes a typical Gibsonean model of a navigation task (obstacle avoidance while navigating to a target) (Fajen et al., 2003). Their model consists of a continuous equation where different terms affect how strongly the target acts as an attractor and how strongly obstacles act as repellers, influencing the path of the agent. Although their model is able to predict paths, it has minimal import as a cognitive processing theory. Following Salvucci's argument against similar models (Salvucci, 2006), it is impossible to tell how to map the terms of the equations on to standard cognitive mechanisms and it is unclear how such equations can account for cognitive constraints such as memory capacity.

The purpose of outlining the model by Fajen et al. (2003) was to provide an analogy regarding opposition to other affordance-based research, namely, affordance passage. Although no computational or mathematical model has been developed, a rough verbal-conceptual model can be assembled from the research in the area (Higuchi et al., 2004, 2006, 2012; Chang et al., 2009; Fath & Fajen, 2011; Stefanucci & Geuss, 2010; Wagman & Taylor, 2005; Wagman & Malek, 2007; Warren & Whang, 1987). Describing the model, however, is not straightforward because there are no processes as such, simply a set of factors that influence shoulder rotation when passing through the apertures, much like the dynamical systems navigation model (Fajen et al., 2003).

One of the main factors thought to be an influential modulator in the control of shoulder rotation while walking through apertures is the ratio between the width of the aperture and the width of the shoulders of the agent walking through it (A/S). Warren and Whang showed (Warren & Whang, 1987) that an A/S ratio of 1.16 - 1.3 (depending on conditions) is the ratio at which their participants judged an aperture to be impassible without shoulder rotation. They show further that although participants of different sizes (*small* and *large*) rotate to different degrees for various aperture widths, when expressed as an A/S ratio, the group difference is eliminated; showing very convincingly that the degree of rotation is modulated by A/S ratio. Their proposal is that the A/S ratio is perceived directly without performing any internal inference processes. They theorize that because eye-height and shoulder width are in a consistent ratio, participants use intrinsic units of eye-height to perceive aperture width. Then, according to their theory, since eye-height has a proportional relationship to shoulder-width, the aperture is perceived as a ratio such that a ratio greater than 1.0 indicates passability. Another finding of note in Warren and Whang's research is the effect of speed that has been replicated in other aperture passage studies (Higuchi et al., 2011; Wagman & Malek, 2007)

A related study by Fath and Fajen (2011) manipulated the visual properties of the environment (in a virtual environment) in order to show that the A/S ratio is not the only factor influencing shoulder rotation. Participants in their study viewed 3 different virtual environments through a head-mounted display, while walking through an empty room. The participants were instructed to walk through the aperture and rotate shoulders as required to avoid collision.

The virtual apertures were presented in three different conditions. In the first condition, the 'post' condition, cylindrical poles appeared in a textured ground plane. In the 'tall post' condition, no ground plane was visible and the polls spanned the visual field from top to bottom. In the third, 'wall' condition, no ground plane was visible, a wall structure defining the aperture spanned the visual field top to bottom as well as from each edge of the aperture to the edge of the visual field. Importantly, each of the conditions was purported to provide decreasingly rich visual cues that could be used for passability judgment. For example, without a ground plane, an estimate of eye-height is impossible.

Results showed that rotation across the three conditions was identical, implicating different sources of factors modulating rotation. They propose that, along with eye-height, head-sway and stride length are separate factors in shoulder rotation. Both head-sway and stride length are intrinsic body-scaled units that require movement. Head-sway units of aperture width are visually present in the 'tall post' condition and stride length units thought to be the only remaining intrinsic unit in the 'wall' condition (for full explanation how they are optically specified see, Fath & Fajen, 2011).

Finally, Higuchi et al. (2012) add an element of their own with their theory that the central nervous system attempts to maintain a constant safety margin between the edges of the aperture and the agent. In their experiment participants walk through apertures of various sizes but their overall body width is modified by having them hold a bar a various lengths (under the assumption that the participants and the bar produce a bar-plus-participant system). Their idea is that with bar lengths that exaggerate the width of the body, if A/S ratio is the sole factor, then there should be an exaggerated rotation; whereas with a consistent safety margin, mathematically unaffected by exaggerated body width, rotation degree should reduce with increased body width.

Of the three papers described above, only Higuchi et al. (2012) have an explicit processing element, proposing that the central nervous system achieves the safety margin at the time of crossing. If we break up the aperture crossing task into three time segments: planning, rotation, and rotation termination; then we might plausibly place Higuchi et al.'s theory in the rotation termination phase. Although not explicitly stated in their respective theories, it seems reasonable to suggest that A/S ratio (Fath & Fajen, 2011; Warren & Whang, 1987), head-sway, and stride length (Fath & Fajen, 2011), are factors in a planning phase. By working the different processes together it is possible to come up with a high-level processing description in which A/S ratio, head-sway, and stride length make up a planning phase, shoulder rotation occurs, ended by the central nervous system, which Higuchi et al.

(2012) suggest might be done by taking the difference between the aperture width and body width until it reaches some constant value. This processing description is high-level and clearly biased toward the tail end, with Higuchi et al. providing the most detail. Beyond that, the cognitive processing description is rather vague leaving open many details and questions.

5.2 Contributions

The previous section is a description of the state of both theory (more abstractly), and cognitive processing description as it exists in previous literature. This section reviews the cognitive processing description, proposed as an alternative to the above, in order to briefly outline the contributions in this thesis. A list of contributions and brief reviews will be provided below.

(C1) An alternate affordance theory that aperture passage is a geometric affordance.

The geometric affordance theory proposes that aperture passage results from a desire to move through space with intended actions (such as walking), which leads to a comparison between geometric properties of the environment and geometric properties embedded in body schemas (either current or recalled). If there is a body schema that satisfies the geometric requirements and the actions requirements, the action is deemed feasible. The theory proposes that the geometric comparison is used at different stages of aperture passage from an initial judgment phase, where passage is judged plausible or implausible, to a planning stage, where the motor plan to achieve the appropriate body schema enacted; and finally during action to monitor if intermediate body postures satisfy the geometric requirements.

Strictly speaking, the theory of geometric affordances does not fit in with previous affordance theories as it proposes the use of representations and processes. However, it does address some difficulties in previous affordance theories. It is unclear, for example, how the vision system becomes tuned to the size of an agent's body.

Gibson's explanation is largely evolutionary-based, proposing that we have our evolutionary history in a particular environment and our actions capabilities are a complimentary to the conditions in that environment. What is not clear is how an ontogenetic trait (such as shoulder width) can be accounted for such that different members (large vs small) perform differently. For example, it is not clear how an evolution-based explanation can account for individual differences, which are evident in the rotation studies.

The second major contribution provided is:

(C2) A computational-level processing description of the overall aperture passage process.

This, of course, must be broken down into a number of sub-parts, each of which are a contribution. As described in the previous section, apart from Higuchi et al. (2012), there is very little discussion in the literature about the kind of processing involved in aperture passage. The following list enumerates a number of details not yet described in the previous literature.

(Q1) How an agent identifies an aperture,

(Q2) The role of A/S ratio, head-sway, and stride length in planning (if any);

(Q3) and whether that planning involves:

(a) a specification of when to start rotation,

(b) a specification of kinemantic properties of the rotation (e.g. rotation speed),

i. and whether speed is explicitly a factor in that calculation;

(Q4) Whether the rotation strategy used while walking through an aperture without carrying a bar is the same as the strategy used while carrying a bar.

I implemented a model of the judgment phase in order to address (Q1). The passability judgment is the process by which an agent identifies an aperture as passable *before* attempting to walk through it. Although this is clearly an important process in the overall aperture passage process, none of the known literature discuss how the process occurs. One of the weaknesses of the direct perception approach as described in the aperture-passage literature is that it provides no way to account for how the agent knows it would be able to pass through an aperture once the shoulders are rotated.

In order to account for this rather large oversight, the theory of geometric affordances proposes that representations of past postures is stored in memory in the form of body schemas, and that these body schemas can be used to to make comparisons between the geometric properties of the aperture and the geometric properties of the body in the previously achieved schema. Theory proposing body schemas is not new and discussion of the supporting literature is presented in section 2.3.1. In order to provide a detailed processing description of body schema:

(C3) a body schema system that stores postures at biomechanical constraints that can then be later retrieved was implemented.

All models described in Chapter 4 use a body schema memory system developed for the purposes of this thesis. The models automatically encode body postures before the experiment starts and retrieves body postures that are later used as goal states for the motor system. That each of the models successfully use the body schema system provides a functional evaluation of the body schema system.

Body schema is how the theory of geometric affordances gets around the difficulty of ontogenetic traits related to the A/S ratio, as discussed above. Although there is no theory of their development presented in the thesis, the theory of geometric affordances supports a theory that proposes that the representational content of body schema is inferred from proprioceptive and visual processes as supported by evidence (Schwoebel & Coslett, 2005). Other relevant aperture passage studies also imply that body schema can be refined over time (Wagman & Taylor, 2005; Higuchi et al., 2006).

Q2 remains largely un-addessed. At a theory level it is assumed that head-sway and stride-length might effect perceptual processes but largely the theory and model are a level of abstraction above the influence of those factors. Although there are particular algorithms for visual perception, they are largely non-theoretical. The models does use an analog to the A/S ratio, however. Both Model 1 and Model 2 multiply the agent's width by a parameter ($RadiusMultiplier$), when judging whether rotation is required.

The model assumes, however, that the output of visual processes is a representation from which geometric information can be extracted and used to make a geometric comparison against body schemas to determine aperture passability. In order implement geometric comparison a new vision module was introduced to Python ACT-R.

Because the vision module is an attempt to respect the ideas at the core of affordances, it is designed to work without semantic labels, therefore purposed differently than the vision system in ACT-R/E (Trafton & Harrison, 2011). Without semantic labels, detectors that detect openings (such as doorways) and obstacle detectors had to be built. It is the intention that future work will enhance the number and quality of detectors. Accompanying the vision module is a 3D camera system developed for this project, specific to the MORSE robotics simulator.

I developed a novel 3D camera for the purposes of this project. The camera is placed on the 3D simulated agent (simulated robot) and provides as output a retinotopic scene description that is enhanced with egocentric depth information. This vision system does not use standard computer vision techniques, instead it exploits the structured 3D information, exploiting features such as object IDs to quicken processing when possible.

Two other key advances of ACT-R 3D over ACT-R/E (Trafton & Harrison, 2011) is a middleware that provides simulation-time synchrony between MORSE and Python ACT-R and an ACT-R controlled motor system implemented as a production system. One of the main measure used in psychology and one of the main measures predicted by ACT-R is reaction time. To support accurate timing, the middleware not only handles communication between Python ACT-R and the MORSE robotic simulator, it also supports simulation-time synchrony between the two systems as described in 3.4.4.

While ACT-R/E (Trafton & Harrison, 2011) introduced a complex motor system for control of a robot, the motor system developed for the purposes of this thesis has simulation-cycle control of motor movements. Instead of handing-off control of the motor system to robot controller, the motor control system updates the limbs of the 3D agent model at a time resolution of $10ms$. The kinematic properties of the agent are controlled via a motor-control production system in Python ACT-R. Currently the only motor control available is the control designed for shoulder rotation. An interesting future work would be to explore motor learning through production compilation and reinforcement learning.

5.3 Model and Hypotheses

All the practical components were required components in order to test hypotheses related to both the theory (C1) and explore different variants of a computational, information processing model (C2). The theory presupposes that geometric information is either encoded or can be derived from perceptual representations of the environment and that geometric information is either encoded or can be derived from body schema. The processing model makes the same assumptions and implements a geometric comparison mechanism in the form of a top-down filter on visual information based on body schema height, depth, and width information:

5.3.1 Conclusions

One of the main weaknesses of previous aperture-passage literature, as pointed out in (Q1) is that they offer no theoretical account about how an agent can judge the passabilty of an aperture before passing through it. Although researchers since Warren and Whang (1987) generally maintain that judgment of rotation is based on the A/S ratio, there is no account about how apertures are rejected outright.

Regarding the later stages of the rotation task, Higuchi et al. (Higuchi et al., 2012) propose that the central nervous system attempts to maintain a constant safety margin at the time of aperture crossing. In their account, the frontal width of the shoulders is calculated and subtracted from the width of the aperture.

5.3.2 Behavioural Conclusions

The theory of geometric affordances proposes that aperture-passage judgment is based on a geometric comparison between the aperture and the current or a stored body schema. If a body schema is recalled from memory, that body schema is used as the goal-state of the motor system. The theory presupposes the existence of an abstract geometric comparison mechanism. A model implementing such a mechanism in the form of a visual filter is presented in section 3.4 and tested in Chapter 4

The first hypothesis tested in this thesis is that:

($H1$) a geometric comparison mechanism is sufficient to account for both the judgment phase and the rotation phase.

Regardless of the fit to the data, at a functional level, it can be concluded that a geometric comparison mechanism that compares the geometric properties of the environment against the geometric properties of a body schema is sufficient as a mechanism for aperture-passage judgment because an agent is able to select an appropriate posture which guides the appropriate action. Similarly, if the agent does not retrieve a body schema, then the agent avoids the action, avoiding attempting to walk pass through apertures that are not physically passable (given the action constraints). Importantly, action selection is done without any semantic labels and is unique to each individual agent.

Consistent with the literature, it is also hypothesized that:

($H2$) The input to the geometric comparison mechanism differs between the judgment phase and the rotation phase. Specifically, that the geometric representation derived from the body schema is multiplied by a factor in-and-around the range reported in Warren and Whang (1987) during the judgment phase; and that a constant is added to the body schema during the rotation phase as described by Higuchi et al. (2012).

It can be concluded that the geometric comparison mechanism receives qualitatively different input during action planning than during active rotation. This conclusion is supported in the differences between Model 1 and Model 2. As explained in 4.1.2, Model 1 uses a *VisionMultiplier* parameter which multiplies the agent's width by a factor of 1.139. Without a factor of roughly 1.14, the agent will not rotate for an aperture of $50cm$, which is inconsistent with human performance. Although the fit to the human rotation data is reasonable in Experiment 1, the model shows a large over rotation in Experiment 2, suggesting that the factor 1.139 is too large. Model 2, however, also uses a factor during the judgment phase but instead of multiplying the agent width by a factor it adds a constant ($3cm * 2 = 6cm$) to the agent's width to judge rotation termination (in accordance with Higuchi et al. (2012)). Model 2 has a decent fit to the rotation data in Experiment 1 and also a fairly decent fit to the rotation data, both in terms of absolute rotation angle and spatial margin for the first two bar conditions (control and 1.5 times) in Experiment 2.

The support of both ($H1$) and ($H2$) help support the theory that representations of the environment have a geometric element and that the geometric properties can be compared to the geometry of body schema, inline with findings that body schema have geometric components (Cardinali, Brozzoli, & Farnè, 2009; Cardinali, Frassinetti, et al., 2009). Importantly, for a theory of affordances, although representational, this theory is pre-semantic. The agent in Model 1 and Model 2 has no semantic knowledge of the environment and the environment is not semantically tagged in any way. The aperture passage affordance is realized through an agreement between the desired motor goal (to walk) and the geometric match between a body schema (present or recalled) and the geometry of an opening.

A geometric comparison mechanism also provides a more parsimonious account of aperture-passage. Although the input to the mechanism differ at different stages, the same mechanism is responsible for making the judgment. The differing input is, at least at a high-level, consistent with Milner and Goodale's (2011; 1992) Two Visual Streams Hypothesis that proposes action planning is done in the ventral stream and is less accurate than the visual information used in the dorsal stream in the process of carrying out action.

A common extension to Warren and Whang's (1987) original experiments is to have participants pass through an aperture in some agent-plus-object system, one version of which is carrying a bar through an aperture (e.g. Higuchi et al., 2012) and is modeled in Experiment 2 (see Chapter 4). Past research have both implicitly and explicitly assumed behaviour is qualitatively the same when passing through an aperture without carrying an object and when carrying an object (Higuchi et al., 2012, 2006; Wagman & Malek, 2007; Wagman & Taylor, 2005). Although the tasks are similar, for a information processing perspective, it seems unsubstantiated that the tasks are cognitively the same. Hypothesis three tests whether:

($H3$) the information processing steps are the same in situations where rotation does not include a bar as in situations where the agent is carrying a bar.

This hypothesis is largely unsupported although there is still testing to be done before a strong conclusion can be made either way. The hypothesis was tested by maintaining a constant model between Experiment 1 and Experiment 2. Both Model 1 and Model 2 assume that the length of the bar can be perceived in the manner described by Carello and Turvey (2004). Although Model 1 has a poor fit to rotation data in Experiment 2, Model 2 has a reasonable fit in both experiments (at bar lengths *control* and *1.5 times* in Experiment 2). We can, therefore, conclude that it least plausible that the cognitive processing, under the constraints of the experiment, when carrying and not carrying a bar are the same. Some uncertainty about the conclusion comes from the fact that, as of yet, the model has not been evaluated with regards to the kinematic properties (see 3.5.4).

Although ($H3$) is mostly unsupported, it does seem that in the *2.5 times* condition, the processes controlling rotation are different. Higuchi et al. (2012) propose that a bar of that length violates the assumption of agent-plus-object system. It is possible that a more conservative strategy is used to monitor rotation, with a slower rotation rate. Without kinematic data, however, this is purely speculative.

5.4 Discussion and Applications

The theory of geometric affordances extends beyond aperture passage involving shoulder rotation. Theoretically, GA can be used to explain a number of postures changes such as ducking or crawling behaviour. Importantly, as an affordance theory, it supports simulations that involve unlabeled environments, with semantically naive agents. As demonstrated in this thesis, GA supports autonomous action for simulated agents. Such autonomy is environment agnostic: an agent should be able to act in novel environments as well as dynamic environments, with minimal training (though, requiring a database of postures). Unlike previous affordance theories (e.g. Chemero & Turvey, 2007; Stoffregen, 2003; Gibson, 1986), the theory of geometric affordances is not meant to be an all-encompassing theory of action. As exhibited in the the models with the use of a goal module and in the embedding in a symbolic architecture such as ACT-R, the theory is perfectly compatible with the use of symbolic approaches in many scenarios.

An early source of inspiration and an obvious application is synthetic agents to model emergency response such as police, fire, and close-quarters-combat (CQC) simulation. Models such as Shootwell (Cox & Fu, 2005), MOUTbots (Wray, Laird, Nuxoll, Stokes, & Kerfoot, 2005), ICARUS (Choi, Konik, Nejati, Park, & Langley, 2007), SNAP (Ting & Zhou, 2009), and the early ACT-R MOUT (Military Operations in Urban Terrain) models (Best & Lebiere, 2003; ?, ?); model navigation, decision making, and planning for military or rescue operations. In all of these approaches, agents are endowed with semantic knowledge of their environments. However, in modern simulators it is often possible to include dynamic environments: environments that take munitions damage. Approaches that attempt to pre-process environments to embed semantic labels (e.g. Ouellet, Somers, & Davies, 2013) as well do not provide any advantages when dealing with dynamic environments. Autonomous agents, however, are intended to make action decisions on their own. Agents able to exploit geometric affordances could make actions decisions in environments where little semantic information exists. For example, a hole in the wall, large enough to afford passage, could be exploited by an GA capable agent despite the fact the hole may not be labeled as a passageway.

5.5 Future Work

One of the main advantages of using a computational approach in this domain is that identifying the constraints also help identify what measures would be informative in research involving human participants. As discussed in 3.5.4, one of the main assumptions is an assumption about the kinematic profile of the rotation. Although there has been a number of studies on aperture-passage involving shoulder rotation, the only researchers to report kinematic details are Fath and Fajen (2011) and even then, the details are limited.

A study of aperture passage with shoulder rotation where measures of rotation rate are conducted would be an immediate benefit to this research. A fairly straightforward study to conduct could be modeled on Warren and Whang's original research (Warren & Whang, 1987). A three dimension motion analysis system such as the one described in the work by Higuchi, Seya, and Imanaka (Higuchi et al., 2012), would be ideal for capturing the rotation rate. Another important measure includes the distance at which participants begin to rotate, as well as their speed. While model 2 suggests rotation initiation and rotation termination strategies in agreement with a combination of Warren and Whang (1987) and Higuchi et. al (2012), the precise timing and kinematic properties could easily have a major impact on the model results.

Extending the research further could include other models of aperture passage. For example: aperture passage with respect to height (Stefanucci & Geuss, 2010). A ducking posture would be fairly straightforward to model in ACT-R 3D and two models combined might present a challenging test of whether active geometric comparison during aperture crossing can account for both results. It is possible for example, that while the production system loop that controls to motor module in Model 2 is sufficient for shoulder-rotation, it could be that either a well-planned ducking motion or a more precise ducking motion control. While that in-and-of itself would not invalidate Model 2, it would present parsimonious results that GA would be expected to explain.

Another interesting possibility would be to introduce a multitasking condition. Model 3 is a brief exploration of direct visual-motor control of rotation, that suggests that routing visual information through the central production system is a more plausible approach (though more exploration is possible). Despite those results, the two visual streams hypothesis (Goodale, 2011) suggests a different stream of processing when vision is used to inform action. Modeling this in ACT-R 3D is a little less straightforward as it would require implementing a multitasking feature, likely similar to the implementation of Salvucci and Taatgen (Salvucci & Taatgen, 2008).

References

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*(6), 679–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7552179

Anderson, J. R. (2007). *How Can The Human Mind Occur In The Physical Universe?* New York, NY: Oxford University Press.

Best, B., & Lebiere, C. (2003). Spatial plans, communication, and teamwork in synthetic MOUT agents. In *Proceedings of behavior representation in modeling and simulation conference (brims).*

Byrne, M. D. (2001). ACT-R/PM and menu selection: applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, *55*(1), 41–84. doi: 10.1006/ijhc.2001.0469

Byrne, M. D. (2013). How Many Times Should a Stochastic Model Be Run? An Approach Based on Confidence Intervals. *Proceedings of the 12th International Conference on Cognitive Modeling, ICCM 2013*, 445–450.

Card, S. K., Moran, T. P., & Newell, A. (1980, jul). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, *23*(7), 396–410. doi: 10.1145/358886.358895

Cardinali, L., Brozzoli, C., & Farnè, A. (2009). Peripersonal space and body schema: Two labels for the same concept? *Brain Topography*, *21*(3-4), 252–260. doi: 10.1007/s10548-009-0092-7

Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., & Farnè, A. (2009). Tool-use induces morphological updating of the body schema. *Current Biology*, *19*(13), 1157. doi: 10.1016/j.cub.2009.06.048

Carello, C., Grosofsky, A., Reichel, F. D., Solomon, H. Y., & Turvey, M.

(1989). Visually Perceiving What is Reachable. *Ecological Psychology*, *1*(1), 27–54. doi: 10.1207/s15326969eco0101_3

Carello, C., & Turvey, M. (2004). Physics and Psychology of the Muscle Sense. *Current Directions in Psychological Science*, *13*(1), 25–28. doi: 10.1111/j.0963-7214.2004.01301007.x

Chang, C.-H., Wade, M. G., & Stoffregen, T. A. (2009). Perceiving affordances for aperture passage in an environment-person-person system. *Journal of Motor Behavior*, *41*(6), 495–500. doi: 10.3200/35-08-095

Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology*, *15*(2), 181–195. doi: 10.1207/S15326969ECO1502_5

Chemero, A., & Turvey, M. T. (2007). Gibsonian Affordances for Roboticists. *Adaptive Behavior*, *15*(4), 473–480. doi: 10.1177/1059712307085098

Choi, D., Konik, T., Nejati, N., Park, C., & Langley, P. (2007). A believable agent for first-person shooter games. In *Proceedings of the third annual artificial intelligence and interactive digital entertainment conference* (pp. 71–73).

Coslett, H. B., Buxbaum, L. J., & Schwoebel, J. (2008). Accurate reaching after active but not passive movements of the hand: Evidence for forward modeling. *Behavioural Neurology*, *19*(3), 117–125. doi: 10.1155/2008/972542

Cox, C., & Fu, D. (2005). AI for automated combatants in a training application. In *Proceedings of the second australasian conference on interactive entertainment* (pp. 57–64). Creativity & Cognition Studios Press. Retrieved from http://dl.acm.org/citation.cfm?id=1109189

Decety, J. (1996). The neurophysiological basis of motor imagery. *Behavioural Brain Research*, *77*(1-2), 45–52. doi: 10.1016/0166-4328(95)

00225-1

Echeverria, G., Lassabe, N., Degroote, A., & Lemaignan, S. (2011). Modular open robots simulation engine: MORSE. In *2011 ieee international conference on robotics and automation* (pp. 46–51). IEEE. doi: 10.1109/ICRA.2011.5980252

Echeverria, G., Lemaignan, S., Lacroix, S., Karg, M., Koch, P., Lesire, C., & Stinckwich, S. (2012). Simulating Complex Robotic Scenarios with MORSE. In *Simpar* (pp. 197–208).

Fajen, B. R., Riley, M. A., & Turvey, M. T. (2009). Information, affordances, and the control of action in sport. *International Journal of Sport Psychology*, *40*(1), 79–107.

Fajen, B. R., Warren, W. H., Temizer, S., & Pack, L. (2003). A Dynamic Model of Visually-Guided Steering, Obstacle Avoidance, and Route Selection. *International Journal of Computer Vision*, *54*, 13–34.

Fath, A. J., & Fajen, B. R. (2011). Static and dynamic visual information about the size and passability of an aperture. *Perception*, *40*(8), 887–904. doi: 10.1068/p6917

Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, NJ: Erlb.

Glover, S., & Dixon, P. (2002). Dynamic effects of the Ebbinghaus illusion in grasping: support for a planning/control model of action. *Perception & psychophysics*, *64*(2), 266–278. doi: 10.3758/BF03195791

Goodale, M. A. (2011). Transforming vision into action. *Vision research*, *51*(13), 1567–87. doi: 10.1016/j.visres.2010.07.027

Goodale, M. A., & Milner, A. D. (1992, jan). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1374953

Graziano, M. S., Cooke, D. F., & Taylor, C. S. (2000). Coding the location of the arm by sight. *Science (New York, N.Y.)*, *290*(5497), 1782–1786. doi: 10.1126/science.290.5497.1782

Haazebroek, P., van Dantzig, S., & Hommel, B. (2011, nov). A computational model of perception and action for cognitive robotics. *Cognitive processing*, *12*(4), 355–65. doi: 10.1007/s10339-011-0408-x

Harrison, A. M., & Schunn, C. (2003). ACT-R/S: Look Ma, no "cognitive map"! In *Proceedings of the fifth international conference on cognitive modeling* (pp. 129–134). Bamberg, Germany: Universitats-Verlag Bamberg. Retrieved from http://act-r.psy.cmu.edu/papers/466/harrison{_}and{_}schunn{_}03.pdf

Harrison, A. M., & Schunn, C. D. (2002). ACT-R / S : A Computational and Neurologically Inspired Model of Spatial Reasoning Three Visiospatial Systems. In W. D. Gray & C. D. Shunn (Eds.), *Proceedings of the twenty fourth annual meeting of the cognitive science society* (p. 1008). Fairfax, VA: Lawrence Erlbaum Associates Ltd.

Harrison, A. M., & Trafton, J. G. (2010). Cognition for action: an architectural account for grounded interaction. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society (CogSci 2010)*, 200–205. Retrieved from http://cwl.cogsci.rpi.edu/cogsci10/cogsci10{_}proceedings/papers/0033/index.html

Hess, R., & Modjtahedzadeh, A. (1990). A control theoretic model of driver steering behavior. *IEEE Control Systems Magazine*, *10*(5), 3–8. doi: 10.1109/37.60415

Higuchi, T., Cinelli, M. E., Greig, M. a., & Patla, A. E. (2006). Locomotion through apertures when wider space for locomotion is necessary: adaptation to artificially altered bodily states. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *175*(1), 50–9. doi: 10.1007/s00221-006-0525-4

Higuchi, T., Murai, G., Kijima, A., Seya, Y., Wagman, J. B., & Imanaka, K. (2011). Athletic experience influences shoulder rotations when running through apertures. *Human Movement Science*, *30*(3), 534–549. doi: 10.1016/j.humov.2010.08.003

Higuchi, T., Seya, Y., & Imanaka, K. (2012). Rule for Scaling Shoulder Rotation Angles while Walking through Apertures. *PLoS ONE*, *7*(10), 1–8. doi: 10.1371/journal.pone.0048123

Higuchi, T., Takada, H., Matsuura, Y., & Imanaka, K. (2004). Visual estimation of spatial requirements for locomotion in novice wheelchair users. *Journal of experimental psychology. Applied*, *10*(1), 55–66. doi: 10.1037/1076-898X.10.1.55

Hu, Y., Eagleson, R., & Goodale, M. A. (1999). The effects of delay on the kinematics of grasping. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *126*(1), 109–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10333011

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., . . . Grodd, W. (1999, sep). Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An fMRI Study. *Journal of Cognitive Neuroscience*, *11*(5), 491–501. doi: 10.1162/089892999563553

Michaels, C. F., & Carello, C. (1981). *Direct Perception* (J. J. Jenkins, W. Mischel, & W. W. Hartup, Eds.). Englewood Cliffs, NJ: Prentice-

Hall.

Milner, A., & Goodale, M. (2008, jan). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. doi: 10.1016/j.neuropsychologia.2007.10.005

Ouellet, S., Somers, S., & Davies, J. (2013). High-level representation of 3D models of buildings. In R. West & T. Stewart (Eds.), *Proceedings of the 12th international conference on cognitive modeling.* Ottawa: Carleton University.

Parsons, L. M. (1987). Imagined spatial transformation of one's body. *Journal of experimental psychology. General*, *116*(2), 172–191. doi: 10.1037/0096-3445.116.2.172

Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of experimental psychology. Human perception and performance*, *20*(4), 709–730. doi: 10.1037/0096-1523.20.4.709

Pomerleau, D. A., & Jochem, T. (1996). *Rapidly adapting machine vision for automated vehicle steering* (Vol. 11). doi: 10.1109/64.491277

Sahin, E., Cakmak, M., Dogar, M. R., Ugur, E., & Ucoluk, G. (2007). To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control. *Adaptive Behavior*, *15*(4), 447–472. doi: 10.1177/1059712307084689

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human factors*, *48*(2), 362–80.

Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. *Perception*, *33*(10), 1233–1248. doi: 10.1068/p5343

Salvucci, D. D., Markley, D., Zuber, M., & Brumby, D. D. P. (2007).

iPod Distraction : Effects of Portable Music-Player Use on Driver Performance. *CHI 2007 Proceedings: Task & Attention*, 243–250. doi: http://doi.acm.org/10.1145/1240624.1240665

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review*, *115*(1), 101–130. doi: 10.1037/0033-295X.115.1.101

Schwoebel, J., & Coslett, H. B. (2005). Evidence for Multiple, Distinct Representations of the Human Body. *Cognitive Neuroscience*, *17*(4), 543–553.

Schwoebel, J., Coslett, H. B., & Buxbaum, L. J. (2001). Compensatory coding of body part location in autotopagnosia: Evidence for extrinsic egocentric coding. *Cognitive neuropsychology*, *18*(4), 363–381. doi: 10.1080/02643290126218

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701–703. doi: 10.1126/science.171.3972.701

Shneier, M., Chang, T., Hong, T., Shackleford, W., Bostelman, R., & Albus, J. S. (2008, jan). Learning traversability models for autonomous mobile vehicles. *Autonomous Robots*, *24*(1), 69–86. doi: 10.1007/s10514-007-9063-6

Sirigu, A., Duhamel, J.-R., Cohen, L., Pillon, B., Dubois, B., & Agid, Y. (1996, sep). The Mental Representation of Hand Movements After Parietal Cortex Damage. *Science*, *273*(5281), 1564–1568. doi: 10.1126/science.273.5281.1564

Somers, S., & West, R. (2013). Steering Control in a Flight Simulator Using ACT-R. In R. L. West & T. C. Stewart (Eds.), *Proceedings* *of the 12th international conference on cognitive modelling.* Ottawa: Carleton University. Retrieved from http://iccm-conference.org/2013-proceedings/187/0187-paper.pdf

Stefanucci, J. K., & Geuss, M. N. (2010, jul). Duck! Scaling the height of a horizontal barrier to body height. *Attention, Perception & Psychophysics*, *72*(5), 1338–1349. doi: 10.3758/APP.72.5.1338

Stewart, T., & West, R. (2010, jan). Testing for Equivalence: A Methodology for Computational Cognitive Modelling. *Journal of Artificial General Intelligence*, *2*(2), 69–87. doi: 10.2478/v10229-011-0010-8

Stewart, T. C., & West, R. L. (2005). Python ACT-R: A New Implementation and a New Syntax. In *12th annual act-r workshop.*

Stoffregen, T. A. (2000, feb). Affordances and Events. *Ecological Psychology*, *12*(1), 1–28. doi: 10.1207/S15326969ECO1201_1

Stoffregen, T. A. (2003). Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, *15*(2), 115–134. doi: 10.1207/S15326969ECO1502_2

Ting, S.-P., & Zhou, S. (2009). Dealing with dynamic changes in time critical decisionmaking for MOUT simulations. *Computer Animation and Virtual Worlds*(May), 427–436. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cav.291/abstract doi: 10.1002/cav

Trafton, J. G., & Harrison, A. M. (2011). Embodied Spatial Cognition. *Topics in Cognitive Science*, *3*(4), 686–706. doi: 10.1111/j.1756-8765.2011.01158.x

Trafton, J. G., Harrison, A. M., Fransen, B. R., & Bugajska, M. D. (2009). An embodied model of infant gaze-following. In *International conference of cognitive modeling.* Manchester, United Kingdom.

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello II, F. P., Khemlani, S. S., & Schultz, A. C. (2012). ACT-R/E: An embodied cognitive architecture for Human-Robot Interaction. *Journal of Human-Robot Interaction*, *1*(1), 78–95.

Trafton, J. G., Schultz, A. C., Perznowski, D., Bugajska, M. D., Adams, W., Cassimatis, N. L., & Brock, D. P. (2006). Children and robots learning to play hide and seek. *Proceeding of the 1st ACM SIGCHISIGART conference on Humanrobot interaction HRI 06*, 242. doi: 10.1145/1121241.1121283

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of experimental psychology. Human perception and performance*, *24*(3), 830–46.

Tucker, M., & Ellis, R. (2001, dec). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*(6), 769–800. doi: 10.1080/13506280042000144

Turvey, M. T. (1992). Affordances and Prospective Control: An Outline of the Ontology. *Ecological Psychology*, *4*(3), 173–187. doi: 10.1207/s15326969eco0403_3

Ugur, E., & Sahin, E. (2010, jul). Traversability: A Case Study for Learning and Perceiving Affordances in Robots. *Adaptive Behavior*, *18*(3-4), 258–284. doi: 10.1177/1059712310370625

Ungerleider, L., & Mishkin, M. (1982). Two Cortical Visual System. In D. J. Ingle, M. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: MIT Press.

van Doorn, H., van der Kamp, J., & Savelsbergh, G. J. P. (2007). Grasping the Müller-Lyer illusion: the contributions of vision for perception in action. *Neuropsychologia*, *45*(8), 1939–47. doi: 10.1016/j.neuropsychologia.2006.11.008

Wagman, J. B., & Malek, E. A. (2007). Perception of Whether an Object Can Be Carried Through an Aperture Depends on Anticipated Speed. *Experimental Psychology (formerly "Zeitschrift für Experimentelle Psychologie")*, *54*(1), 54–61. doi: 10.1027/1618-3169.54.1.54

Wagman, J. B., & Taylor, K. R. (2005). Perceiving Affordances for Aperture Crossing for the Person-Plus-Object System. *Ecological Psychology*, *17*(2), 105–130. Retrieved from `http://www.tandfonline.com/doi/abs/10.1207/s15326969eco1702{_}3` doi: 10.1207/s15326969eco1702_3

Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of experimental psychology. Human perception and performance*, *10*(5), 683–703. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/6238127`

Warren, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of experimental psychology. Human perception and performance*, *13*(3), 371–83. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/2958586`

West, R. L., & Emond, B. (2005). The Environment as Theory : An Example Using the ACT-R / SOS Environment. *Proceedings of the Sixth International Conference on Cognitive Modeling*, 398–399.

West, R. L., & Nagy, G. (2007). Using GOMS for Modeling Routine Tasks Within Complex Sociotechnical Systems : Connecting Macrocognitive Models to Microcognition. *Joural of Cognitive Engineering and Decision Making*, *1*(2), 186–211. doi: 10.1518/155534307X232848.

West, R. L., & Pronovost, S. (2009). Modeling SGOMS in ACT-R : Linking Macro- and Microcognition. *Journal of Cognitive Engineering and Decision Making*, *3*(2), 194–207. doi: 10.1518/155534309X441853.

West, R. L., & Somers, S. (2011). Scaling up from Micro Cognition to Macro Cognition : Using SGOMS to build Macro Cognitive Models of Sociotechnical Work in ACT-R. In *Proceedings of the 33rd annual cognitive science society* (pp. 1788–1793). Retrieved from `http://palm.mindmodeling.org/cogsci2011/papers/0397/paper0397.pdf`

Westwood, D. A., Heath, M., & Roy, E. A. (2000). The effect of a pictorial illusion on closed-loop and open-loop prehension. *Experimental Brain Research*, *134*(4), 456–463. doi: 10.1007/s002210000489

Wray, R. E., Laird, J. E., Nuxoll, A., Stokes, D., & Kerfoot, A. (2005). Synthetic Adversaries for Urban Combat Training. *AI Magazine*, *26*(3), 82–92.