Physics-, Social-, and Capability-Based Reasoning for Robotic Manipulation

by

Kenton J. Williams

Submitted to the Department of Mechanical Engineering in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

Signature of Author . . . .
Department of Mechanical Engineering
January 20, 2012

Certified by . . . .
Cynthia Breazeal
Media Arts and Sciences, Associate Professor
Thesis Supervisor

Certified by . . . .
John Leonard
Mechanical Engineering Faculty Reader, Professor
Thesis Supervisor

Accepted by . . . .
David E. Hardt
Mechanical Engineering, Professor
Chair, Committee for Graduate Students

Physics-, Social-, and Capability-Based Reasoning for Robotic Manipulation

by

Kenton J. Williams

Submitted to the Department of Mechanical Engineering
on January 20, 2012, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

Robots that can function in human-centric domains have the potential to help humans with the chores of everyday life. Moreover, dexterous robots with the ability to reason about the maneuvers they execute for manipulation tasks can function more autonomously and intelligently. This thesis outlines the development of a reasoning architecture that uses physics-, social-, and agent capability-based knowledge to generate manipulation strategies that a dexterous robot can implement in the physical world. The reasoning system learns object affordances through a combination of observations from human interactions, explicit rules and constraints imposed on the system, and hardcoded physics-based logic. Observations from humans performing manipulation tasks are also used to develop a unique manipulation repertoire suitable for the robot. The system then uses Bayesian Networks to probabilistically determine the best manipulation strategies for the robot to execute on new objects. The robot leverages this knowledge during experimental trials where manipulation strategies suggested by the reasoning architecture are shown to perform well in new manipulation environments.

Thesis Supervisor: Cynthia Breazeal
Title: Associate Professor

Thesis Supervisor: John Leonard
Title: Mechanical Engineering Faculty Reader, Professor

Acknowledgements

This work was made possible by the MIT Lincoln Laboratory, the National Science Foundation Graduate Research Fellowship Program, the Ford Foundation Pre-Doctoral Fellowship Program, and the MIT Media Lab.

To Dr. Cynthia Brezeal - Thank you so much for welcoming me to the Personal Robots family. Your guidance, support, and patience have been invaluable to me throughout this process.

To Dr. John Leonard - Thank you so much for being my advocate in the mechanical engineering department. I could not have survived these past years without your encouragement and support.

To Polly Guggenheim - Thank you for being you! Giving demos to children and watching their faces light up has kept my enthusiasm for robotics going strong and reminded me that I am here for a much bigger purpose. I have you to thank for that.

To the Robots family - Jin Joo, Nancy, Julian, Kris, Jason, Angela, Siggi, Nick, Adam W., Adam S., Ryan, Mikey, Dan, Matt, Jesse, Sonia, Julie, Peter, Fardad, Philipp, David, Sophie, Natalie, Kailas, John, Jun, Dirk, Ed, Leslie, and Marc. You all have encouraged me, challenged me, and helped me in many ways throughout this process. I am so grateful to you all.

To Nexi, Maddox, Xylo, and Axle - Deciphering and fixing your malfunctions has pushed me to my breaking point many times, and has made me a better roboticist in the process. Thank you!

To my mother - This thesis is a direct result of the many sacrifices you have made for me throughout my life. Your unconditional love and encouragement has molded me into the man that I am today and for that, I am eternally grateful to you. I promise you that the countless sacrifices you made for me will not go in vain. I love you!

To my family- Granny, Auntie, Nicholas, Tammy, Uncle Marty, Elmo, The Blake family, The Pickens family, The Wisham/Mosby family, and my many other relatives. You have been my rock and my motivation throughout

my entire life and I cannot imagine where I would be without you. I love you all!

To Erica - A lot has changed since our days at Jackson, but our friendship has always remained the same. You are my sister and I am so thankful to have you in my life. Who knew that two little nerdy kids from the ghetto would grow up and shut it down. From rags to robots baby!

To Hasani - You are the big brother that I always wanted and I cannot put into words how important you have been to my survival at MIT. I probably would have gone crazy a long time ago if it were not for the good times we have shared laughing and kiki-ing. Thanks for always reminding me that robots break down...and life goes on.

To my friends - Dortch, Denzil, Weslee, Richard, Jairo, Cleveland, Shawn, Andrew, Anslem, Jermaine, Kane, Manny, Monica and my MSRP family. I have kept my sanity thanks to all the laughs and good times we have shared. And a special thank you to Marcelo for your support throughout this entire process and for helping me with my research experiments.

To my fellow Buckeyes - Melody, Gerrick, Ryan, Chris, and Joey. Your love, support, and encouragement have gotten me through many highs and lows, both at MIT and Ohio State. Believe it or not, there is a little bit of each of you in these next few pages.

To my mentors - Aisha Walcott, Legena Henry, Amon Milner, Dr. James McLurkin, Dean Christopher Jones, Dean Blanche Staton, Dean Minnie McGee, Dr. Gregory Washington, Dr. Oliver McGee, Felicia Sawyer, and Ruby Smith. This, and many other accomplishments would not have been possible without your guidance. I am forever grateful for your mentorship.

And last, but never least, to Kingston - Thank you for always giving me two powerful words of advice when I am at my lowest of lows...

Push Yourself!

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background

Robots that can function in human-centric domains have the potential to help humans with the chores of everyday life. Environments such as schools, homes, offices, or hospitals can reveal a wide range of possible applications for robotic assistance. For example, robots with manipulation skills could retrieve objects for elderly persons with limited mobility, manipulate tools to assist doctors during surgery, or even handle hazardous materials in environments that are too dangerous for humans.

However, challenges typically arise for robotic manipulation in human environments because these spaces are often unpredictable, dynamic, and difficult to model. Fortunately, many aspects of these settings can be exploited and used to the robot's advantage. For instance, people often populate human-centric domains, thus robots can occasionally rely on the guidance and assistance from a human partner to complete a task. Also, objects in the environment may exhibit unique characteristics and inherent physical properties that the robot can identify to reduce the cognitive load required to manipulate such objects.

1.2 Motivation

Humans, animals, and dexterous robots are creatures that all have uniquely different techniques for manipulating objects. So a simple task like moving a ball on the floor from point A to point B may involve extremely different maneuvers depending on the creature's action repertoire. For instance, a human may pick up the ball at point A, carry it as he/she walks to the goal, then place it on the ground at point B. A dog might bite the ball, carry it in its mouth, then release the ball at point B. And a robot may push the object with its end-effector as it navigates to point B. In each case, a different maneuver arises because of the creature's own set of manipulation abilities.

The manipulation strategy may also change depending on inherent physical properties of the object, e.g. geometry, size, orientation, etc. These physics-based properties are known as object affordances [1], [2] and can be used by dexterous creatures to determine the action possibilities on an object. For example, if the ball from the above task were now a heavy bowling ball, then the human may roll the ball to the goal instead of carrying it. Or if the ball were now a huge, lightweight beach ball, then the dog from above might tap the ball with its nose to roll it to the goal because the ball is now too large to fit in its mouth.

Social factors also play a role in determining manipulation strategies. For instance, let us now assume that our task is to place a mug into a trashcan. Since the trashcan is simply a receptacle, very little thought is given to how we approach the trashcan or the way that we drop the object into it. However, if we are now giving the mug to a human receiver, then we would most likely approach the human from the front and we might deliver the mug such that the handle is exposed for the human to grasp easily, which may require re-grasping of the object. So our manipulation actions may change to consider human safety and comfort. There has also been work done on understanding non-verbal cues, like eye contact and body pose, which can occur while coordinating object hand-offs with a human partner [3]. Further, shared attention between agents plays a vital role in manipulating an object of mutual interest [4], [5]. In particular, dexterous robots must be aware of postural cues and gaze direction to perform successful object hand-offs to humans or other robots. For example, if the human is occupied or not facing the robot during the time of object hand-off, then the robot must think intelligently about alternative strategies.

Also, educational, manipulative toys are useful for helping babies and toddlers develop manipulation skills. Often times, these toys feature simple objects such as balls, blocks, and cylinders. Through trial and error, children play with these objects repetitively until they are able to utilize concrete strategies for grasping these objects [6]. As their dexterity develops, they are then able to grasp more intricately shaped objects. They can manipulate these new objects by leveraging previously learned grasping techniques [7]. This method of learning to grasp simple objects then applying that knowledge to new objects that have similar physical features can be useful for dexterous robots.

1.3 Significance

If robots are to perform manipulation tasks in human-centric domains, then they must be able to reason about the maneuvers that they implement in these challenging workspaces. Therefore, I propose the research question: *How can a dexterous robot utilize physics-, social-, and capability-based knowledge of its environment for object manipulation?*

The core contribution of this work is the development of a reasoning architecture that allows a dexterous robot to generate successful manipulation strategies using social contexts, the manipulation capabilities of itself and its partner, and knowledge of object affordances in the manipulation environment. This work is unique in that it fuses these three research concepts together, which has yet to be fully explored in robotic manipulation.

This work also provides a foundation for generating strategies used in future interactions from previously solved manipulation tasks. For example, learned physics-based reasoning patterns may be detected from past object interactions and applied to new objects with similar physical properties.

Additionally, the reasoning architecture is versatile in that it can be expanded to include new factors that might influence the manipulation strategies performed in new environments.

1.4 Related Work

1.4.1 Naïve Physics

Schmolze and Davis [8], [9] proposed using basic physical properties of objects to formalize human commonsense knowledge, but these works do not consider the complexities that arise when implementing such an expansive corpus on a humanoid robot. Further, these research approaches are limited because it is difficult to encapsulate a clearly defined body of knowledge that is consistent, comprehensive in scope, and universal among different people. Also, problems arise when trying to distinguish between truly naïve physics and formal physics because it is difficult to discern beliefs about the physical world that are taught by an expert versus those that are formed through exploration.

1.4.2 Object Affordances

The use of object affordances for robotic manipulation was done in [10], [11], and [12]. In [10], affordances were taught to a robotic manipulator through probabilistic relational models. However, the research platform used is a single, whole-arm manipulator so bimanual manipulation maneuvers were not considered. Also, the probabilistic relational models learned by the robot are only used to ascertain object affordances, not for motion planning of manipulation strategies. In [11] and [12], robots learned manipulation skills through imitation and previous exploration. Again, these works use a single, whole-arm manipulator as its research platform, thus, do not consider the complexities of bimanual manipulation. Also, these works differ in that they assume an accurate geometric model of manipulation objects is not available. In my approach, manipulation plans are generated based on fully described geometric models. Further, the robot's motor controllers used in these works rely heavily on tactile feedback from hand and finger sensors, whereas my robot platform uses vision-based sensing for object detection. Moreover, these works do not consider object exchange to human partners, or generalize a full repertoire of manipulation skills for robots.

1.4.3 Human-Robot Coordination

Non-verbal cues, timing, and coordination in human-robot manipulation was explored in [3], [13], and [14]. In [3] and [14], object exchange between partners was investigated with regards to social and physical cues exclusively. However, these works do not consider how physical object properties can influence manipulation maneuvers when handing off objects to humans. In [13], interactive manipulation between a human and humanoid robot was explored where the interaction was represented solely by means of mathematically representing relative configurations between the human's and robot's hands. This approach does not consider the influence of non-verbal cues from human partners during object exchange. It also differs in that its proposed techniques are valid for cooperation tasks that are not properly defined, whereas, my approach assumes that manipulation tasks are be described wholly.

1.4.4 Agent Ability

The consideration of agent capabilities in robotic manipulation was demonstrated in [15] and [16]. The hand and arm capabilities of a robotic manipulator are learned through a representation scheme, which visualizes and inspects directional structures in the robot's own workspace. However, these works do not explore object exchange between human and robot partners, therefore, it does not consider the workspace constraints or capabilities of a human receiver.

CHAPTER 2

Research Platforms

2.1 Codebase

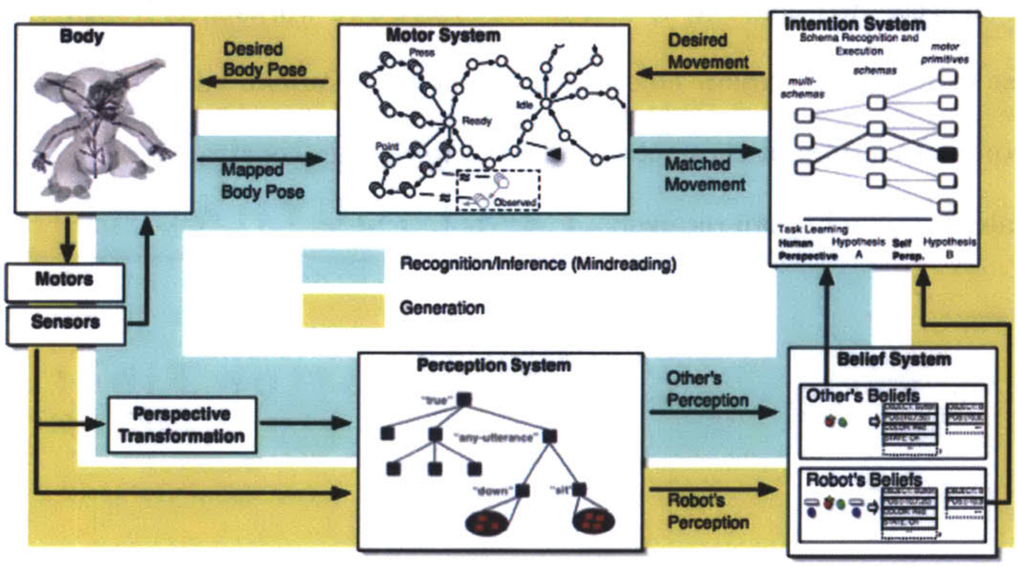


Figure 2-1: R1D1 System Architecture Overview.

The codebase used throughout this research is R1D1, a java-based cognitive architecture used for designing synthetic brains for virtual and physical creatures in complex environments [17]. R1D1 allows creatures to detect information about its workspace through internal and/or external sensors, formulate beliefs about the workspace from that perceptual data, generate task-dependent motor actions, and execute those actions in the workspace. Figure 2-1 illustrates the complete behavior system pipeline.

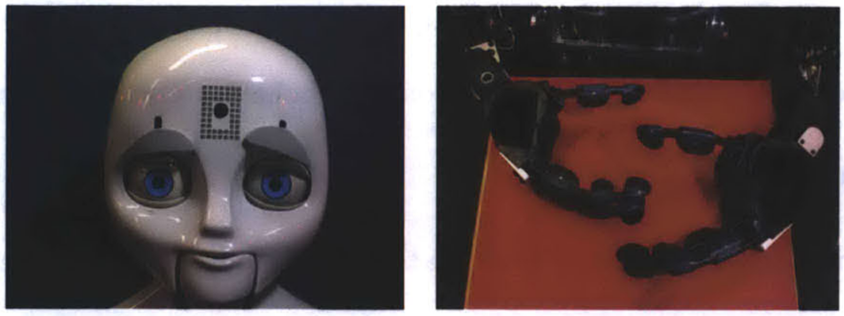
2.2 Physical Robots

The research platforms used are two mobile, dexterous, and social (MDS) robots [18] with varying manipulation, navigation, and social abilities. Figure 2-2 shows Nexi and Xylo, the MDS fleet used throughout this research.

Figure 2-2: MDS robots: Nexi (left), Xylo (right).

2.2.1 Nexi

Nexi is a 47-inch tall humanoid with a mobile base, 4-fingered end effectors with partial range of motion, and a socially expressive face. Figure 2-3 shows a close-up of Nexi's face and end effectors. Nexi is equipped with with two stereo cameras in her eyes, Figure 2-3(a), which are used for vision



(a) (b)

Figure 2-3: (a) Nexi's face, (b) Nexi's hands.

and human face detection. Nexi's end effectors, Figure 2-3(b), feature four-digit hands with opposable thumbs, which are connected to a wrist joint with a wide range of motion. Her fingers are covered with rubber to reduce slippage between object surfaces. Also, her hands are naturally fixed in a U-shaped configuration, which can be challenging when flat-palmed manipulation techniques are preferred.

2.2.2 Xylo

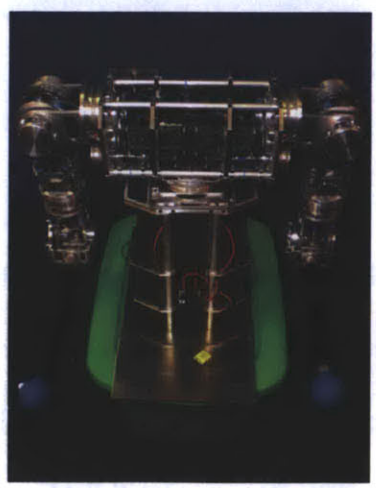


Figure 2-4: Xylo's torso with ball-and-stick end-effectors.

Xylo is 22-inch tall, static, headless humanoid with ball-and-stick end effectors, Figure 2-4. For stability purposes, Xylo must always be placed on top of a supporting structure. Thus, the height of Xylo's end effectors are affected by the height of his support. Xylo's end effectors consist of aluminum rods encased by a thick foam shell with rubber balls attached at the ends. The foam and rubber covering reduce slippage between object surfaces.

2.3 Objects

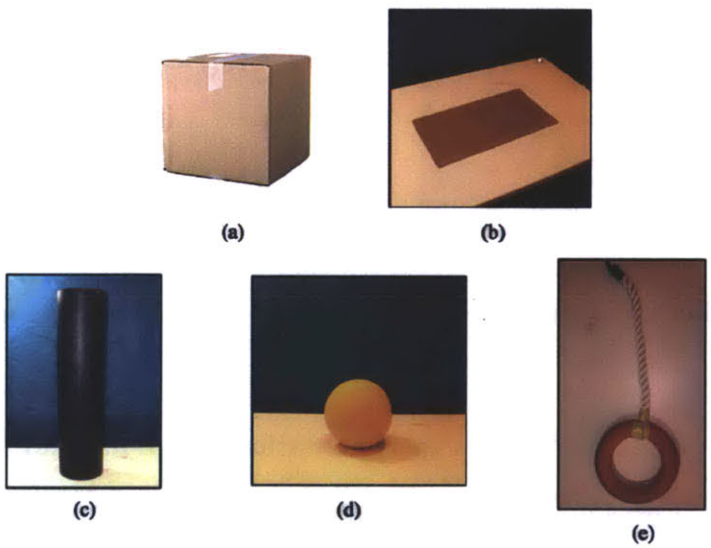


Figure 2-5: Object set for manipulation training. (a) Box, (b) Flat plate, (c) Cylinder, (d) Foam Ball, (e) Ring on rope.

The manipulation object set used to train the robot is shown in Figure 2-5. It consists of a 2-foot length by 1.5-foot width by 1.5-foot height cardboard box, a 1.5-foot by 2.5-foot cardboard plate, a 6-inch diameter by 4-foot long foam cylinder, a 6-inch diameter foam ball, and a 12-inch outer diameter foam ring connected by a flexible rope, Figure 2-5(a-e) respectively. These objects are used to develop Nexi's manipulation skills.

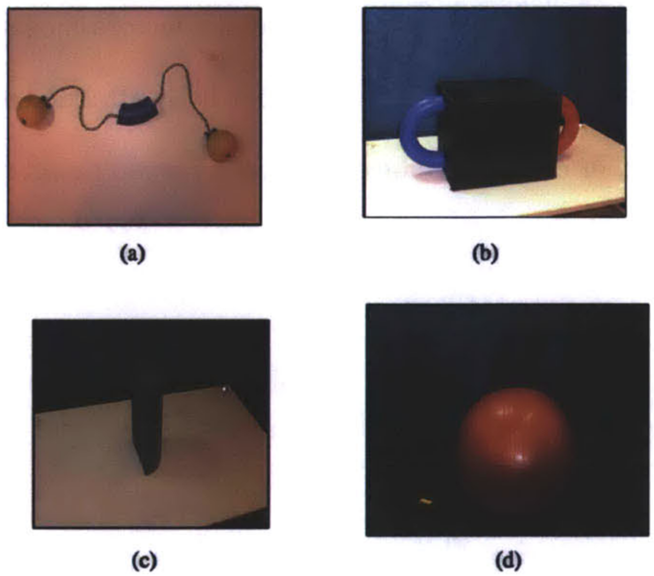


Figure 2-6: Object set for experimentation. (a) Balls on rope, (b) Box with handles, (c) Half cylinder, (d) Inflatable ball.

The object set used for is experimentation shown in Figure 2-6. It consists of two 3-inch diameter foam balls connected by a flexible rope, a 14-inch length by 9-inch width by 10-inch height cardboard box with 2 handles made out of foam rings, a 6-inch diameter by 1.5-foot length foam cylinder sliced in half lengthwise, and a 36-inch diameter inflatable ball, Figure 2-6(a-d) respectively. These objects are used during experimentation only. They are unique because they are either hybrids of the objects in the training set, or they contain similar physical properties as objects in the training set, as shown in Figure 2-7. For example, the half cylinder is a hybrid of the cylinder and the flat plate, Figure 2-7(a), but the inflatable ball is merely a larger version of the foam ball, Figure 2-7(c).

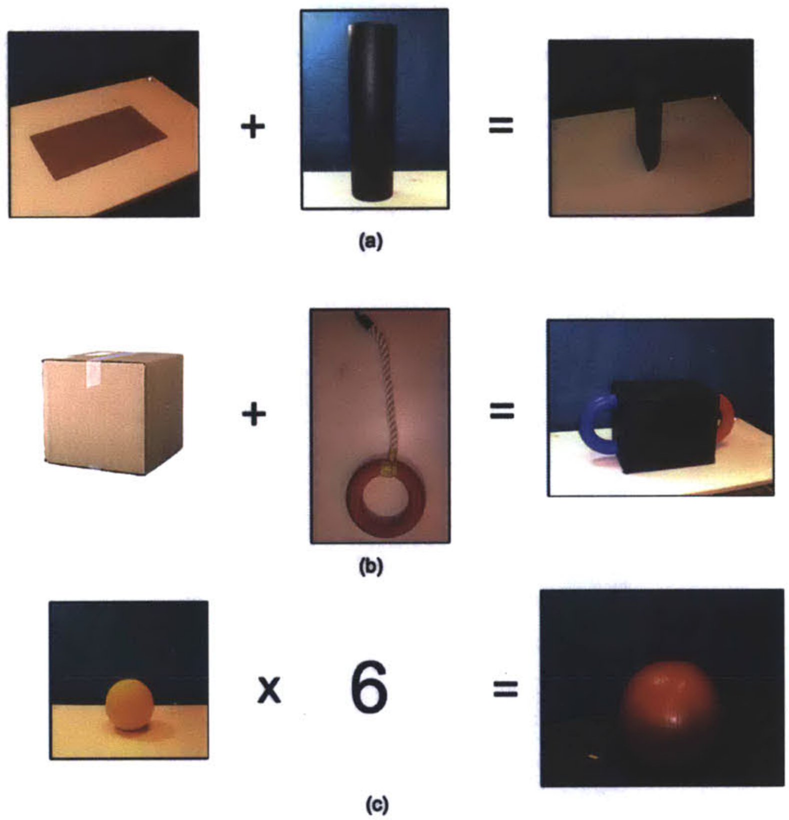


Figure 2-7: Example of hybrid objects. (a) Flat plate and cylinder produces the half cylinder, (b) Box and ring on rope produces the box with handles, (c) Inflatable ball is 6 times the volume of the foam ball.

During training and experimentation, the objects are initially oriented in various configurations. Also, the robot may navigate to these objects using a defined navigation approach direction. Appendix B shows a complete table of the object orientation conventions and navigation approach directions. The orientation and navigation approach directions were defined ad hoc for the purposes of this research.

2.4 Supporting Structures

The supports for the objects are shown Figure 2-8 and include a hanging support for objects with complex geometry, a hanging support for objects with simple geometry, a post, a table, and the floor, Figure 2-8(a-e) respectively.

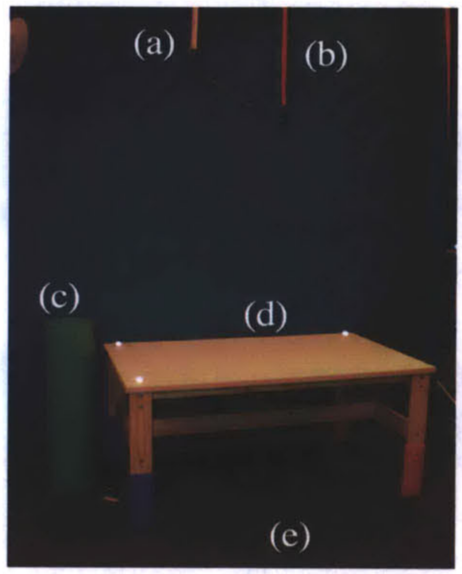


Figure 2-8: Structures used to support objects. (a) Hanging support for large objects, (b) Hanging support for small objects, (c) Post, (d) Table, and (e) The floor.

2.5 Research Environment

The research environment is a 15-foot by 12-foot room surrounded by a Vicon Motion Capture System [19]. Robots, humans, and objects are tagged with small, round, reflective markers that are captured by the Vicon system and their position and orientation data is tracked. As defined for this research, a scene in the environment consists of a single manipulator agent (Nexi), three receiver agents (Xylo, a human, and a receptacle), and objects of assorted shapes, sizes, and orientations that are supported by various structures. A given workspace scene will be any combination of entries selected from each of the five columns illustrated in Table 2.1.

Table 2.1: Possible combinations of a given task environment

MANIPULATOR AGENT	RECEIVER AGENT	OBJECT	OBJECT ORIENTATION	OBJECT SUPPORT
Nexi	Xylo	Flat plate	Upright	Hanging
	Human	Inflatable ball	Flat	Table
	Receptacle	Foam ball	Upside-down	Floor
		Balls on Rope		Post
		Ring on rope		
		Box with handles		
		Cylinder		
		Box		
		Half Cylinder		

The role of the manipulator agent, which can only be Nexi, is to manipulate objects in its environment and deliver them to the receiver agent. The role of the receiver agent, which can be human, Xylo, or a receptacle, is to receive objects from the manipulator agent.

CHAPTER 3

Task Description

3.1 Research Goal

The research goal is to develop a reasoning mechanism that allows a dexterous robot to generate manipulation strategies with considerations to its own manipulation skills, physical properties of objects in its environment, and the capabilities of agents that receives these objects. The robot develops manipulation skills by mimicking human actions and through self-exploration during training sessions. The robot then applies these skills to future manipulation tasks where new objects have similar features as objects previously explored. Sections 4.10 and 5.1 describe the training sessions and

Table 3.1: List of Agent Descriptors

AGENT PROPERTIES		
agent name	is a physical human	is a virtual human
is a virtual robot	is a physical robot	type of end-effectors
number of end-effectors	number of arms	agent is manipulator
agent is receiver	shoulder height from ground in inches	arm length in inches
can manipulate objects	is mobile	agent height in inches
can manipulate magnets	can bend forward	has tactile sensors on end-effectors
has tactile sensors on head	has whole-body tactile sensors	has whole-arm tactile sensors
has a face	has eyes	area of agent palms in square inches

Table 3.2: List of Physical Object Properties

OBJECT PHYSICAL PROPERTIES			
object name	has handles	number of handles	actual volume in cubic inches
volume of bounding box that surrounds object in cubic inches	weight in pounds	has a sharp feature	is a box
has a box feature	is a pyramid	has a pyramid feature	is a 3D polygon
has a 3D polygon feature	is a 2D polygon	has a 2D polygon feature	is a cylinder
has a cylinder feature	is a cone	has a cone feature	is a sphere
has a sphere feature	is a toroid	has a toroid feature	center of mass relative to Vicon centroid in inches
center of mass relative to the ground in inches	is rigid	is flexible	has a cavity
volume of cavity in cubic inches	can conduct electricity	locations of conductive components	active temperature in Celsius
has magnetic features	locations of magnetic features	has reflective features	locations of reflective features
object material composition	object colors	has a flat surface	locations of flat surfaces
is a solid	is a liquid	is a gas	has a light source
is toxic to humans and animals	has a smell	is edible to humans	is alive

experimental research task in detail. The reasoning mechanism takes in a scene as its input, which consists of full descriptions of the manipulator agent, the receiver agent, objects to be manipulated, and any rules or constraints imposed on the system. For instance, a description of the manipulator and receiver agents includes parameters such as: its shoulder height from the ground and arm length to calculate its manipulation range of motion, physical details about its end effectors to determine how it interacts the objects, and a label that classifies the agent as human, robot, or receptacle since these distinct categories will imply different methods for interacting with objects. Table 3.1 shows the list of factors that describe an agent in the environment[1]. Table 3.2 lists the descriptions of an object with regards to unique geometric, spatial, and sensorial properties[2]. For this research, these parameters and descriptors are represented symbolically and hardcoded into the reasoning architecture.

3.2 Hypotheses

3.2.1 Hypothesis I

Hypothesis I involves the system's ability to create a manipulation repertoire for the robot. Nexi's skillset of maneuvers depends on the manipulation techniques used by human test subjects in a pilot study where they are instructed to manipulate various objects. Observations from the pilot study were represented symbolically in the reasoning architecture and these human maneuvers were translated into appropriate movements for the robot. I predict that all of the translated maneuvers will be kinematcially robust for the robot to use in future experimentation. Further, I predict that the repertoire developed for the robot will be sufficient to successfully manipulate all objects used in future experimentation.

3.2.2 Hypothesis II

[1] Every agent descriptor listed in the table may not be used by the reasoning system.

[2] Every object property listed in the table may not be used by the reasoning system.

Hypothesis II outlines the reasoning architecture's ability to generate successful manipulation strategies for the robot to execute in new manipulation tasks. I predict that the system will always select successful strategies from the robot's manipulation repertoire to be used in new manipulation scenarios. Further, I believe that if multiple strategies are possible for a given scene, then the system will accurately rank these maneuvers in order of increasing complexity for the robot while considering imposed system constraints.

3.2.3 Hypothesis III

Hypothesis III highlights how successful the robot is at executing strategies proposed by the reasoning architecture in the physical environment. I hypothesize that the robot will eventually execute proposed manipulation strategies successfully. I believe that the robot may fail at successful execution upon initial attempts due to sensor inaccuracies. However, I predict that recalibration of the robot's motor controllers and vision sensors will prevent these failures on repeated manipulation attempts.

3.2.4 Hypothesis IV

Hypothesis IV focuses on the fluidity between the robot and human receiver agents during object exchange. I predict that the robot will never need to resort to giving a verbal utterance to grab the attention of a human receiver agent. I believe that the system will adequately detect the attentiveness of the human receiver and that occupied humans will pause their actions just in time for successful object exchange. Thus, no verbal statements will be spoken between manipulator and receiver agents. I further predict that humans will experience object exchange in a natural way when receiving objects from a robot partner.

3.3 System Constraints

3.3.1 Workspace Limitations

Many physical descriptors of agents and objects must be declared a priori due to sensor limitations. The Vicon Motion Capture System is useful for tracking positions and orientations of agents and objects, but it does not extract unique object features like handles. Further, human receiver agents are required to interact only within the limits of their own workspace. They are not allowed to enter the robot's workspace or any other domain in the research environment.

3.3.2 Robot Platform Constraints

Nexi's eye cameras are used solely for human face detection, not for feature extraction. Vision-based object recognition is a challenging problem, which will not be addressed in this work. Therefore, many descriptors of objects and agents must be hardcoded as stated in Section 3.1.

Further, Nexi is equipped with current sensors in her fingers, which allows her to detect fluctuations in current draw from her hand motors as she manipulates objects. However, this is a limited tactile sensing ability since it does not accurately detect contact forces and moments. Thus, the Vicon system is heavily relied upon for accurately tracking object positions and orientations.

3.3.3 Reasoning Architecture Constraints

Imposing constraints and rules on the reasoning mechanism can circumvent sensor limitations. A list of system constraints imposed on the

Table 3.3: Constraints Imposed on the Reasoning Architecture

REASONING SYSTEM CONTRAINTS	
Constraint #1	The manipulator agent may translate an object in any direction before, during, or after manipulation.
Constraint #2	The manipulator agent may not rotate an object around any axis located on the object's body before, during, or after manipulation.
Constraint #3	The manipulator agent may not touch the supporting structures in the environment to aid in object manipulation.
Constraint #4	Any object configurations that cause dynamic instability between object and support are not considered for experimentation.
Constraint #5	Any object with at least one handle must be manipulated such that at the time of object delivery to a human receiver agent, at least one handle is available for the receiver agent to grab.

system is listed in Table 3.3. For instance, if objects have features like knobs, grippers, handles, or levers, then humans typically prefer to grasp these features on the object for their own safety and comfort. Therefore, Constraint #5 is imposed, a the rule that states:

> ***Any object with at least one handle, must be manipulated such that at the time of object delivery to a human receiver agent, at least one handle is available for the receiver agent to grab.***

This rule is imposed to consider human safety and comfort during object exchange for objects like the box with handles, Figure 2-6(b). However, this rule may imply that regrasping of the object is necessary. For example, a human manipulator may grasp an object with both handles for its own comfort, but might then reorient the object upon delivery to a human receiver to leave one handle exposed, as shown in Figure 3-1. Thus, object

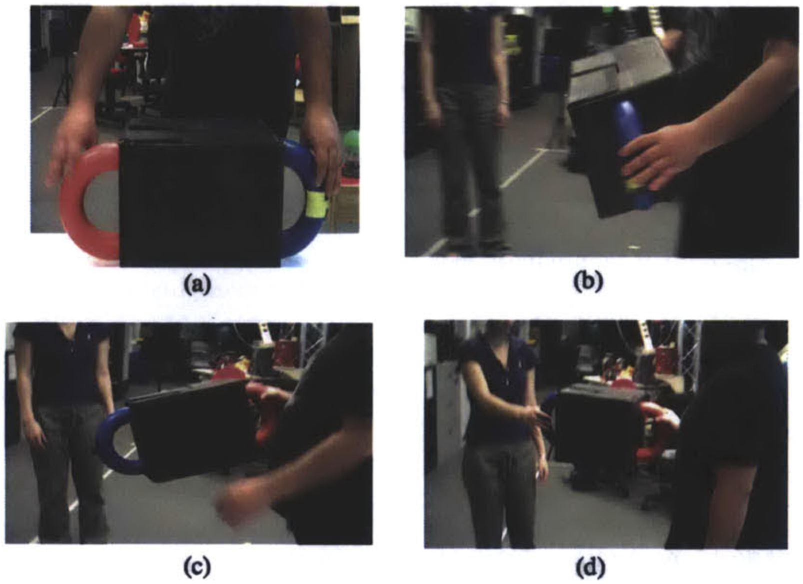


Figure 3-1: Action sequence illustrating object regrasping. (a) Human grabs object with two hands, (b) Human begins to reorient object for partner, (c) Human releases one hand from object, (d) Human delivers object to partner.

reorientation is required. However, object regrasping requires precise coordination and sensing abilities of the dexterous agent. This is a challenging problem in robotic manipulation [20] and will not be addressed here. Therefore, grasping strategies like those in Figure 3-1 are not allowed for the purposes of this research.

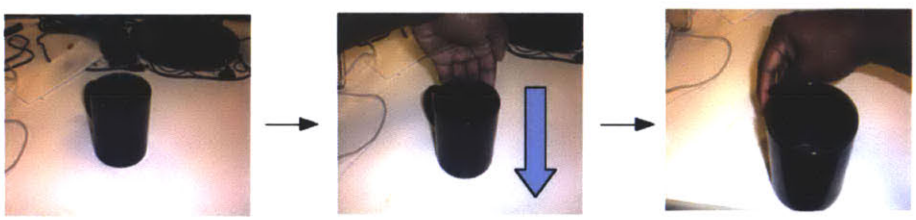
To sidestep regrasping maneuvers, the following constraints are imposed on the reasoning system:

> *The manipulator agent may translate the object in any direction before, during, or after manipulation.*

The manipulator agent may not rotate the object around any axis located on the object's body before, during, or after manipulation.

With these constraints, agents like Nexi, who lack sophisticated tactile sensing, can still deliver objects with graspable features to human receivers given their own unique limitations. Examples of these constraints are

(a) Maneuver not allowed.



(b) Maneuver allowed.

Figure 3-2: Rotation constraints imposed on the reasoning system. (a) Invalid maneuver, (b) Valid maneuver.

illustrated in Figure 3-2. The problem here is that the human manipulator would prefer to grab the mug by the handle for easier grasping. However, the mug is positioned on the table such that the handle is not easily accessible for the human. In Figure 3-2(a), the manipulator rotates the mug around a vertical axis on the object's body so that the handle is more easily accessible. However, this maneuver is a direct conflict of Constraint #2 and will not be allowed. In Figure 3-2(b) on the other hand, the manipulator slides, or translates, the object forward so that the mug is closer to his body. Although the handle is still not in the best configuration for the human to grasp effortlessly, translating the object does help by bringing the object closer within the human's arm and end effector range of motion. Further, this strategy does not conflict with Constraints #1 or #2. Again, maneuvers like those in Figure 3-2(b) may not be realistic for human grasping, but these maneuvers, which are bounded by constraints listed in Table 3.3, are useful for robotic agents with limited tactile sensing.

CHAPTER 4

Methodology

A stated above, a given research workspace consists of a scene which can be any combination of manipulator agent, receiver agent, object, object orientation, and supporting structure, as shown in Table 2.1. For a given scene, the research task is for the manipulator agent to manipulate objects in the environment and deliver them to the receiver agent. The manipulator agent, i.e. Nexi, executes this task by using manipulation strategies generated by the reasoning system.

4.1 System Architecture Overview

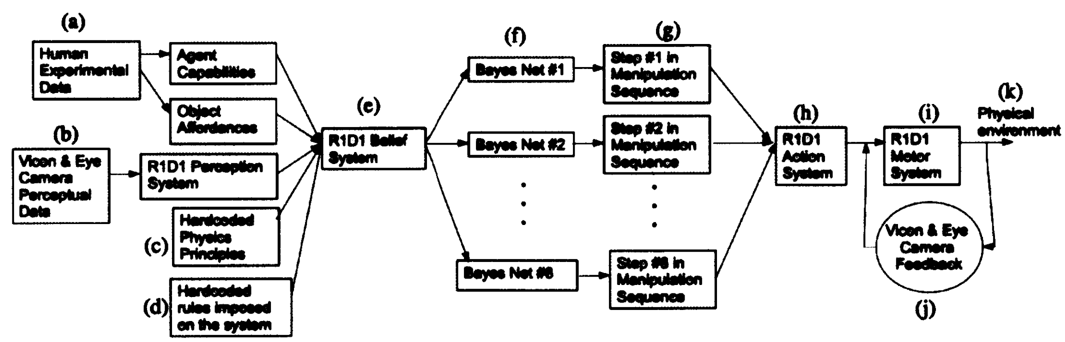


Figure 4-1: Overview of System Architecture.

Figure 4-1 shows the entire research architecture pipeline. Data is collected from videotaped experiments where human test subjects manipulate objects and deliver them to receiver agents. This data is transcribed and used to help determine a manipulation repertoire for Nexi and to aid the robot in learning object affordances [21], Figure 4-1(a). Sensor data from the Vicon system and the robot's eye cameras are interpreted in the R1D1 perception system, Figure 4-1(b). This sensor data is used to determine the location and orientation of items of interest in the environment. Additional physics principles about objects and constraints imposed on the system are hardcoded into R1D1 as shown in Figure 4-1(c) and Figure 4-1(d). The results from the human pilot study, the sensor data, and the hardcoded parameters are then fused in the R1D1 belief system, Figure 4-1(e), where beliefs are generated regarding spatial information about items of interest, drives and motivations of agents, and other task-specific information about the workspace. A series of decision-making tools, called Bayesian Networks [22], [23], then use these beliefs to generate manipulation strategies for the manipulator agent to use in the workspace, Figure 4-1(f) and Figure 4-1(g). The R1D1 action system, Figure 4-1(h), is then responsible for planning and selecting appropriate actions that the robot will use to execute the manipulation strategy. Finally, the R1D1 motor system uses a hybrid of synthesized motions and direct motor control to allow the robot to perform the selected actions in the physical environment, Figure

4-1(i) and Figure 4-1(k). While the robot performs navigation and manipulation skills, it uses feedback from internal and external sensors to detect changes in its environment, Figure 4-1(j). A detailed explanation of the research architecture is discussed in the remainder of this chapter.

4.2 Manipulation Strategies

Figure 4-2: Example situation in which Nexi should deliver the upright cylinder on the floor to a bin.

In order for the reasoning architecture to propose manipulation solutions, a manipulation strategy must be clearly defined. For this work, a full manipulation strategy is defined as an 8-part action sequence that allows a manipulator agent to deliver objects to a receiver agent. This sequence includes: 1) a navigation approach direction to the object, 2) an extending hand action for the object, 3) a pickup maneuver for the object, 4) a retracting hand action for the object, 5) a navigation approach direction to the receiver agent, 6) an extending hand action for delivery to the receiver agent, 7) a releasing maneuver for hand-off to the receiver agent, and 8) a retracting hand action after objects are delivered to the receiver agent. For example, Figure 4-2 shows an example situation in which Nexi's task is to deliver the cylinder, in the upright orientation and supported by the floor, to a receptacle receiver agent. Table 4.1 shows an example solution to

Table 4.1: An 8-step manipulation sequence generated for the situation described in Figure 4-2

MANIPULATION SEQUENCE	STRATEGY IMPLEMENTED
1) Navigation approach direction to object.	Approach object from any direction.
2) Animation playback for hand extension.	Extend forward with both hands.
3) IK grasp maneuver.	Bimanual Smash
4) Animation playback for hand retraction.	Upward Bimanual Retract
5) Navigation approach direction to receiver agent	Approach receiver from any direction.
6) Animation playback for hand extension.	Extend upwards with both hands.
7) IK release maneuver.	Open Fingers
8) Animation playback for hand retraction.	Bimanual Outward Retract

situation as an 8-step manipulation sequence. Conventions and descriptions of grasp maneuvers are outlined in detail in Section 4.3. Figure 4-3 shows the robot performing the proposed manipulation strategy in the physical environment.

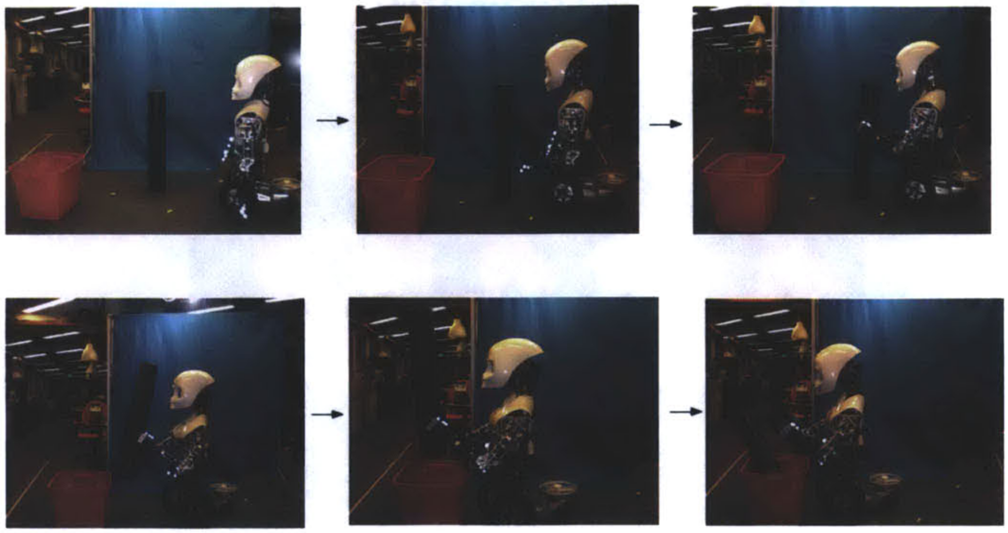


Figure 4-3: Nexi performing the manipulation sequence outlined in Table 4.1.

4.3 Manipulation Repertoire Development

Nexi's manipulation repertoire is established by a combination of exact mimicry of observed human actions and translating certain human actions to suit the kinematics of Nexi's end effectors. First, a pilot study was done where human participants were required to deliver several objects to various receiver agents. The objects used for the pilot study included many items that are not part of the object sets in Figure 2-5 and Figure 32-6, like credit cards, books, wine glasses, etc. These trials were videotaped and the manipulation maneuvers of the human were interpreted and categorized to produce a set of manipulation skills that Nexi could implement. Figure 4-4 shows strategies used by the human participant.

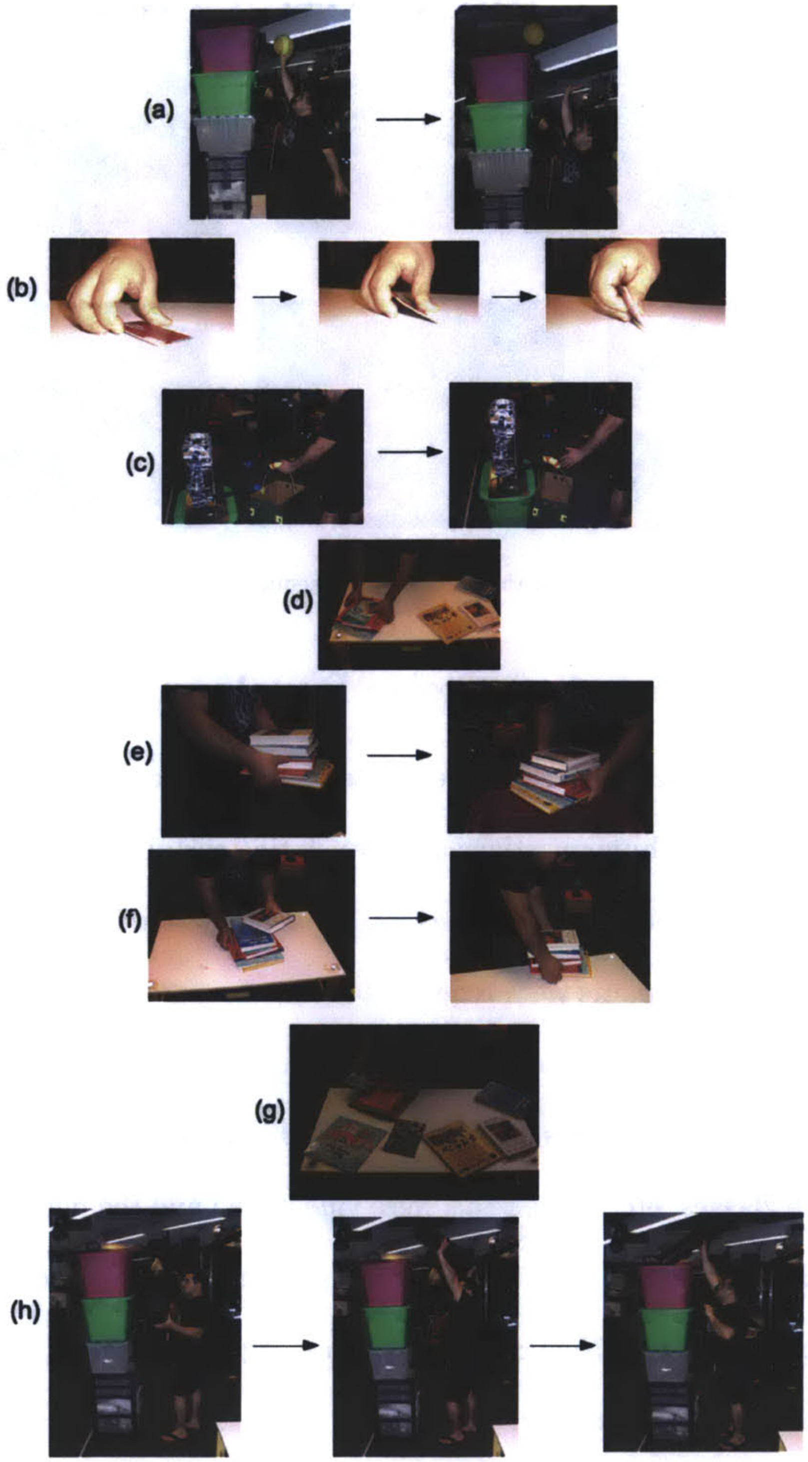


Figure 4-4: Manipulation maneuvers from a human test subject. (a) Toss, (b) Scoop, (c) Hang, (d) Bimanual Power Grab, (e) Bimanual Retract Downward, (f) Push and Collect, (g) Unimanual Power Grab, (h) Overhead Lift.

Clearly, the abilities of a human will be vastly different and more extensive than that of a robot. So the reasoning mechanism finds feasible maneuvers from the human dataset that are kinematically robust for the robotic to use. To do this, I first transcribed observed human actions and fed these descriptors into the reasoning architecture as input. The system then uses pre-programmed logic to determine if maneuvers are suitable for Nexi. It filters these human actions by considering factors listed in Table 3.1 such as tactile sensing, end effector dynamics, and the manipulator agent's arm range of motion. Figure 4-5 shows the pseudo-code for a portion of

```
PSEUDO-CODE
WHILE {grasping}
    IF {action requires whole-body manipulation}
    ELSE IF {action requires whole-arm manipulation}
    ELSE IF {action requires flat, open palmed hands}
                → THEN {action is invalid for robot}

WHILE {navigating}
    IF {action requires approach behind a human}
                → THEN {action is invalid for robot}

WHILE {entire action}
    IF {action requires object re-grasping}
          IF {alternative actions are possible}
                → THEN {use alternative action for robot}
          ELSE IF {alternative actions are not possible}
                → THEN {action is invalid for robot}
```

Figure 4-5: Pseudo-code of the logic used for repertoire determination.

the logic used by the reasoning system. For example, it was observed that the human participant sometimes used whole-arm and whole-body manipulation to complete tasks. Because these maneuvers require whole-body tactile sensing, which Nexi does not have, these maneuvers get eliminated from the robot's manipulation skillset. This is expressed in lines 1-5 in Figure 4-5. The system also filters out any strategies that conflict with the constraints in Table 3.3 as shown in lines 11-16 in Figure 4-5.

Finally, the system generates an inventory of manipulation skills available for Nexi to use in future manipulation tasks. Table 4.2 lists and describes Nexi's manipulation repertoire and Figure 4-6 illustrates the robot performing these actions in the physical environment.

Table 4.2: Manipulation Repertoire for Nexi

MANEUVER NUMBER	MANEUVER NAME	DESCRIPTION
1	Push and Collect	The robot grabs one end of the object, pushes it across a surface, then uses her other hand to grab the other end of the object.
2	Bimanual Smash	The robot performs a bimanual power grasp.
3	Unimanual Power Grab	The robot performs a power grasp with one hand.
4	Upward Bimanual Retract	The robot retracts both hands upward.
5	Downward Bimanual Retract	The robot retracts both hands downward.
6	Upward Unimanual Retract	The robot retracts one hand upward.
7	Downward Unimanual Retract	The robot retracts one hans downward.
8	Open Fingers	The robot opens her fingers.
9	Close Fingers	The robot closes her fingers.
10	Hang	The robot hangs objects onto supports.
11	Bimanual Outward Retract	The robot retracts her hands away from each other to release an object.

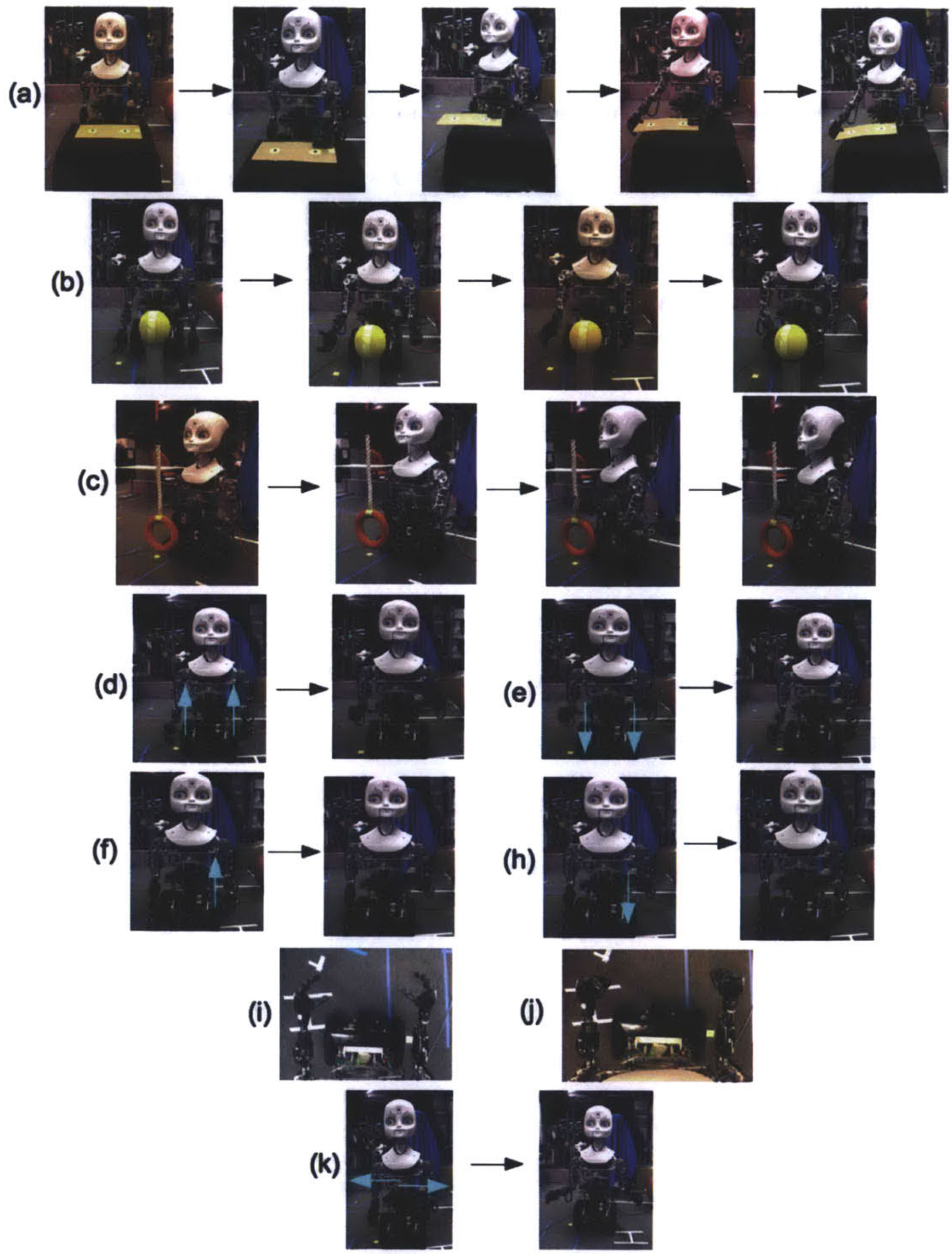


Figure 4-6: The 11 maneuvers in Nexi's manipulation repertoire. (a) Maneuver 1, (b) Maneuver 2, (c) Maneuver 3, (d) Maneuver 4, (e) Maneuver 5, (f) Maneuver 6, (g) Maneuver 7, (h) Maneuver 8, (i) Maneuver 9, (j) Maneuver 10, (k) Maneuver 11.

4.4 Object Affordances

Observations from the pilot data are also used by the reasoning system to learn object affordances [10], [21] and factors in the scene that may affect Nexi's manipulation strategy. These observations are used in conjunction with known physics-based logical rules that are hardcoded into the reasoning mechanism. These rules, which come from [8], [9], [24], [25], and [26], are expressed symbolically in R1D1. Appendix F lists these physics-based rules[3]. For instance, we know from physics that round objects can roll on a surface [25]. So we can state in the reasoning system that if the manipulation object is a round, or has a round feature, then a particular action may cause the object to roll if it supported by a surface. So if the object in question is a cylinder, which has a round feature, then it has the potential to roll.

However, from the pilot data, it was observed that a cylinder could only roll on a surface from a tapping action if it is laying flat. If the cylinder is in the upright orientation, a tapping action would cause it to fall, not roll, as shown in Figure 4-7. So the reasoning system uses a hybrid of hardcoded physics-based reasoning and observations from human object interactions to determine that a manipulator agent can tap a cylinder and cause it to roll only if the cylinder is in the flat orientation and supported by a surface.

[3] Not all rules listed in Table V are used by the reasoning mechanism for strategy generation.

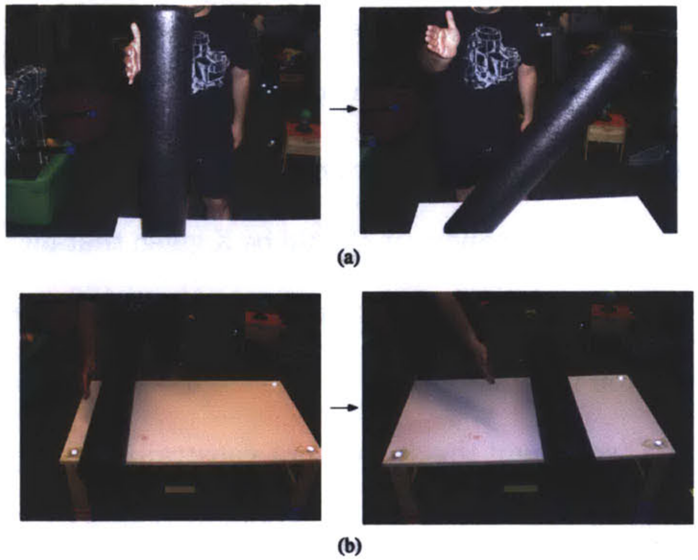


Figure 4-7: Observations from human interactions with a cylinder. (a) Tapping causes the cylinder to fall if it is upright initially, (b) tapping causes the cylinder to roll is it is flat initially.

This process of learning affordances through a hybrid of observations from human interactions and hardcoded physics-based logic is used extensively to build a physics-based reasoning architecture. The robot can then leverage this knowledge corpus for future object interactions.

4.5 Bayesian Networks

Once a physics-based knowledge corpus is established, the next step is to determine a way for the robot to make decisions about which maneuvers to select in their action repertoire for a given scene. Bayesian Networks (BN) are powerful statistical tools that are used extensively throughout robotics for representing beliefs [27]. BNs are probabilistic graphical models that represent a set of variables and their conditional dependencies [22], [23]. For example, BNs are used in the medical field as a way for doctors to diagnose patients. BNs can determine the probability that a patient has condition X given that the patient has symptoms A, B, and C [28], [29]. This logic is used similarly for our reasoning architecture to determine the probability that the appropriate manipulation maneuver should be X given that factors A, B, and C are present in the scene. In the context of this analogy, X represents just one step in the 8-step manipulation sequence discussed in Section 4.2. And factors A, B, and C represent the descriptors, constraints, and rules presented in Table 3.1, Table 3.2, Table 3.3, and Appendix F. Since only one BN is generated for each step in the 8-step manipulation sequence, a total of 8 BNs must be used by the reasoning system in order to determine a full manipulation strategy, as seen in Figure 4-1(f).

For this research, the variables in the BNs are the physics-, social-, and capability-based factors that are present in the manipulation environment. Factors such as, the kinematics and dynamics of the manipulator agent's end effectors, the geometry of the objects, the structures supporting the objects in the scene, the rules and constraints imposed on the system, etc. are all represented symbolically in R1D1 as variables in the BNs.

A single BN is a graph comprised of nodes whose values represent data from the factors discussed above. These nodes are linked through a probabilistic network that can determine how the values of all other nodes affect one node in particular. For this research, the BNs use observations from the pilot study in Section 4.3 and manipulation training sessions discussed in Section 4.10 to perform its analysis. The system finds the probability that a particular manipulation action was used given the values of certain factors in previous manipulation tasks. Thus, the probabilistic nature of the network becomes more statistically significant as more trials are performed and catalogued.

Figure 4-8 shows a segment[4] of a BN used by the reasoning system to

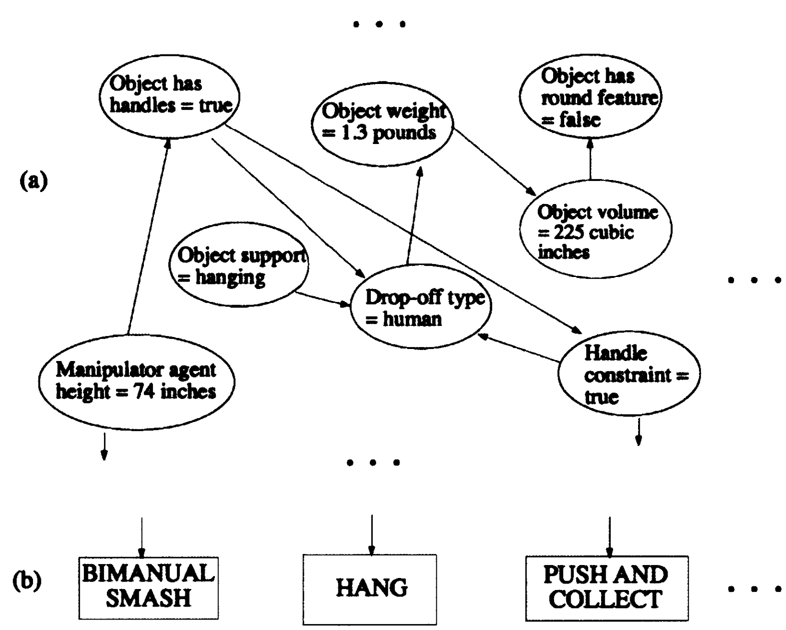


Figure 4-8: A segment of a Bayesian Network used by the reasoning system. (a) Factors that influence manipulation strategies (b) A portion of Nexi's manipulation repertoire.

determine step 3 in a manipulation sequence, i.e. the IK grasp maneuver. Ellipses are used to denote continuations of nodes in the network. Figure 4-8(a) indicates a cluster of nodes that represents the physics-, social-, and capability-based factors and their unique values. Figure 4-8(b) indicates a

---

[4] Due to the expansive nature of the BN graphs, it is difficult to show the network in its entirety.

cluster of nodes that represents the maneuvers available in Nexi's manipulation repertoire. Each factor node can influence other factors nodes, maneuver nodes, or both. The goal is for the reasoning system to determine the best strategy for the robot to execute during a new manipulation task. Therefore, it determines which maneuver yields the highest probability of success when the factor nodes occupy certain values for a given scene. This analysis is based on the probabilities detected from previous interactions. The maneuvers get ranked in order of decreasing success rate and the reasoning system suggests that the robot execute the maneuver with the highest probability of success. The reasoning system does this for each of the 8 BNs to suggest a full 8-step manipulation strategy.

JavaBayes [30] is a software toolkit that allows Java developers to create, modify, and export BNs. It calculates statistical probabilities and expectations, and performs robust analysis on BNs created by the user. For this research, JavaBayes is used to create and analyze Bayesian models and serves as an intermediary between the belief system and action system in R1D1. The reasoning architecture takes in perceptual data from the environment, i.e. the physics-, social-, and capability-based factors discussed above, then generates beliefs about the best possible manipulation strategy for the robot to perform for a given task. R1D1 then translates these beliefs into motor actions that the manipulator agent can perform on objects in the physical world.

4.6 Vision Pipeline

4.6.1 Vicon Object Recognition

As stated previously, position and orientation data is sent to robot via a Vicon motion capture system. The Vicon system consists of nine cameras with pulsating light emitting diodes that can track reflective markers, which are easily attachable to clothing and other materials. The positions of these reflective markers are tracked with high accuracy and often within a few millimeters of error tracking at a rate of 100-120 Hertz [19]. Objects, supports, and both the manipulator and receiver agents in the environment are tagged with these reflective markers that are detected by the Vicon system's cameras. Human agents may wear hats and gloves covered with reflectors to detect their head and hand locations.

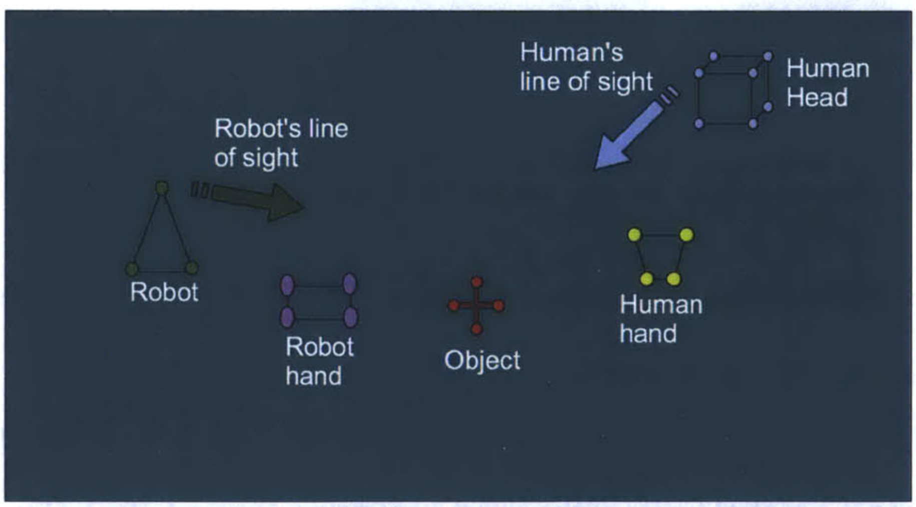


Figure 4-9: Graphic visualization of the Vicon system tracking an object, two agents, agents' hands, and forward vectors representing the agent's line of sight.

Figure 4-9 shows a graphic visualization of the Vicon system tracking objects and agents, which are tagged with reflective markers. The Vicon data is fed into the robot's perception system where it is interpreted, then used to tell the robot where items of interest are located in the workspace, Figure 4-1(b). This data is also used as visual feedback for the robot to execute motor actions in the physical environment, Figure 4-1(j).

In order to extract unique features on objects in the environment, the raw Vicon data undergoes a series of manipulations via 3D vector algebra in the R1D1 perception system. Figure 4-10 shows an example of the feature extraction process for the box with handles. The Vicon system sends the

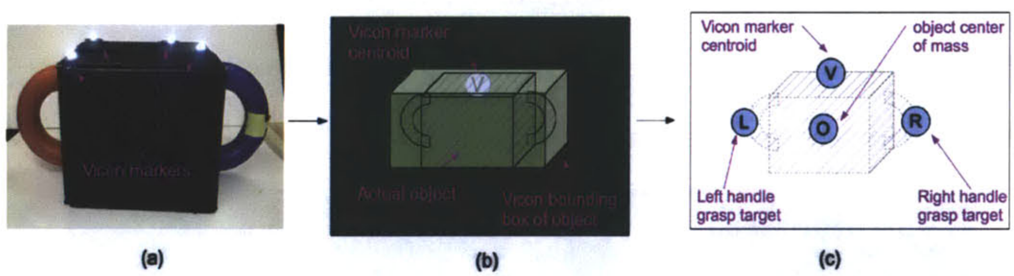


Figure 4-10: Feature extraction using Vicon. (a) Object in the physical environment, (b) Vicon marker centroid and bounding box, (c) Extracted features from Vicon centroid.

position of individual markers as well as the centroid of these markers to the R1D1 perception system. Next, a bounding box which represents the rectangular volume that encapsulates the object gets created by manually recording the spatial limits of the object's geometry, Figure 4-10(b). Once bounding boxes and marker centroids are established, the Vicon marker centroid is manually manipulated in the perception system using 3D transformation matrices to distinguish object features, Figure 4-10(c). This allows the robot to track unique features on the objects like surfaces, edges, and graspable structures like handles, given the position and orientation data of the object.

4.6.2 Human Face Detection & Shared Attention

As stated above, human agents may wear hats and gloves covered with reflectors to detect their head and hand locations, Figure 4-11. This allows

Figure 4-11: Hats and gloves tagged with reflective markers.

the robot to determine the position and orientation of a human's hands during object exchange and a forward vector which is an approximation of the human's line of sight. The line of sight vector is calculated in R1D1 by algebraically manipulating the Vicon marker centroid between reflective dots on the hats. This vector is approximated such that it can be used universally, irrespective of the varying head profiles among different humans.

Nexi is equipped with two Point Grey stereo cameras [31] that run human face detection algorithms similar to those used in [32], [33], [34]. Figure 4-12 shows an example of human face detection as seen

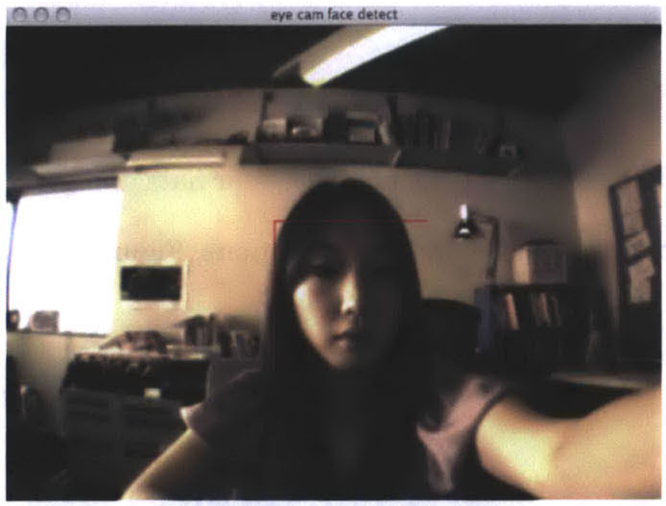


Figure 4-12: Human face detection through Nexi's eye cameras.

through Nexi's eye cameras. When the robot detects a human face, a red half-square is drawn around it to indicate its position. The robot uses these methods to determine a human receiver agent's readiness to receive an object. For example, as the robot approaches a human receiver for object drop-off[5], she looks in the direction of the human's face (as reported from Vicon) to see if the human is ready to receive the object. If the robot detects a human face, then it assumes that the human receiver is engaged for object hand-off and proceeds to deliver the object. When the robot reaches its final navigation goal, it extends its arms towards the human receiver's hands (as reported from Vicon) and waits 2 seconds for the human to grab the object.

[5] Robots are trained to always approach humans from the front.

After 2 seconds has passed, the robot opens its fingers releasing the object. It is assumed that 2 seconds is a sufficient amount of time for the human to grab the object. The robot immediately retracts its arms upon releasing the object.

If the robot does not detect a human face during the approach, then it gives a verbal utterance, like "Here you go!," to get the human's attention before object delivery. It is assumed that this utterance is sufficient to get the human's attention. Once the robot makes this statement, it performs the same object delivery maneuver discussed previously.

4.7 Navigation

Robot navigation is performed using a standard A* (a-star) navigation planner [35], [36] implemented in R1D1. For a given environment, the algorithm treats the receiver agents, the table support, and the post support as obstacles. The navigation target can be the location of the manipulation object or the location of the receiver agent depending on which step in the manipulation strategy the robot is performing. These locations are reported from the Vicon system. While approaching the navigation goal, the robot utilizes two factors that ensure appropriate arrival, a safe navigation approach distance and final orientation vector. The approach distance factor determines the best range between the robot and the receiver agent or objects to allow for safe manipulation. The orientation vector determines how the robot should orient its body to ensure that it is facing the navigation target upon arrival. Both of these factors are determined through trial and error and hardcoded into the navigation planner. The robot is trained to approach objects and receptacles from the direction that is most appropriate for manipulation, but to always approach humans from the front. Figure 4-13 shows a top view of example navigation paths generated in R1D1 when a robot delivers an object to a receptacle, Figure 4-13(a), and to a human, Figure 4-13(b).

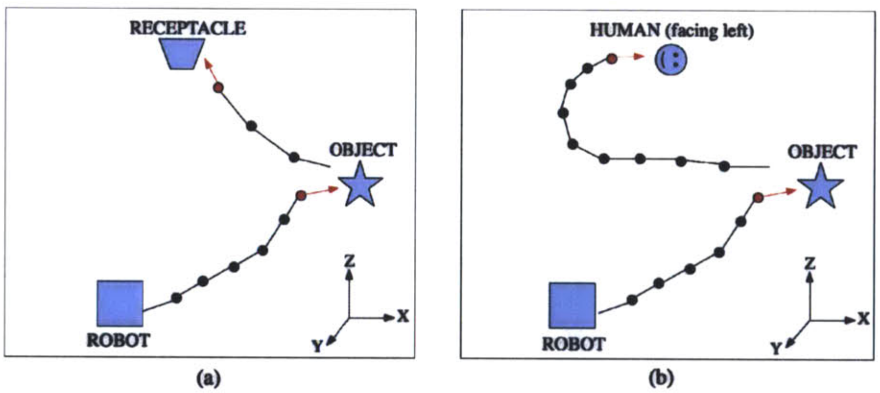


Figure 4-13: Example navigation paths generated in R1D1. (a) Robot delivers object to a receptacle. (b) Robot delivers object to a human.

4.8 Manipulation

4.8.1 Animation Playback

Robotic manipulation is performed through a hybrid of animation playback using Maya software [37] and direct motor control in R1D1, as shown in Figure 4-14. Once a full manipulation repertoire is established, Maya is used to generate animations for virtual creatures that mimic the observed human motor actions. These virtual animations can then be played back on the

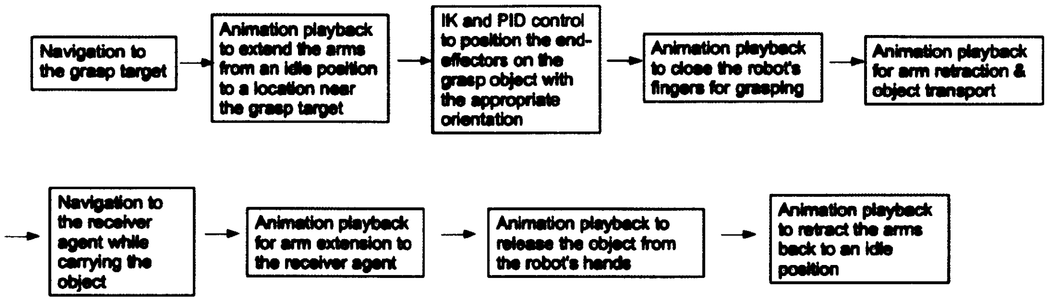


Figure 4-14: Navigation and Manipulation Pipeline.

physical robot [38]. These animations are used for the actions that occur before and after actual object grasping, i.e. arm extension and arm retraction. Arm extension animations position the robot's end-effectors from an idle position to a location near the grasp target for object pick-up or to a location near the receiver agent for object drop-off. Arm retraction animations position the end-effectors from locations on the actual grasp object to locations that allow the robot to safely navigate while carrying objects and to avoid any obstacles in the environment. Arm extension and retraction animations are created with consideration to the robot's kinematics to ensure that the robot's hands and arms do not collide with obstacles in the environment like the table support, the post support, the receptacle, or a human receiver agent. These animations also leverage the safe navigation distance factor discussed above to prevent obstacle collision. Further, items like the table, the post, and the receptacle are carefully selected for the environment so that their sizes and geometry decrease the likelihood of collision with the robots' arms and hands.

4.8.2 Inverse Kinematics & PID Control

Once arm extension animations are played back, Nexi then uses inverse kinematics (IK) [39] and proportional-integral-derivative (PID) control [39], implemented in R1D1, to position her end-effectors to a final position and orientation on the grasp target. Reflective markers are placed in random locations on the objects. The Vicon system finds the centroid of these markers and reports the location and orientation of this centroid to the R1D1 perception system, Figure 4-1(b). However, this reflective marker centroid may not represent the physical object's center of mass or any graspable feature on the object. Therefore, we alter this marker centroid using vector algebra in R1D1 to extract unique graspable features on the objects like corners, handles, and surfaces. This algebraic manipulation process is done for each object and fed to the reasoning system a priori, as discussed in Section 4.6.1. The robot also has reflective markers on her hands. So end-effector centroids are also generated by the Vicon system, which represent the robots' left and right hands. The IK and PID control move the end-effector centroids from their locations at the end of the extension animation to graspable features located on the object in question. The end-effector centroids land on the objects with desired orientations, which are determined through trial and error during manipulation training sessions, as discussed in Section 4.10. Maya animations are then used to close the robots' fingers as well as retract the objects from their supports with the retracting animations discussed above.

4.9 Agent Capabilities

4.9.1 Robot Range of Motion

The reasoning system allows Nexi to determine if objects in the environment can be safely manipulated given the range of motion in her arms. To do this, the system finds a vertical range *H* in which Nexi can safely manipulate the object if it is located within *H*. Figure 4-15(a) shows an arc

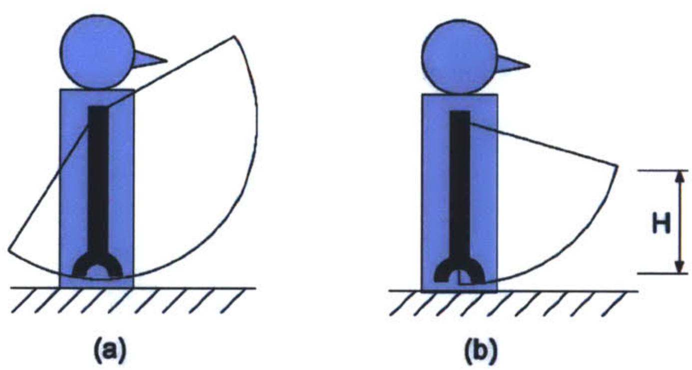


Figure 4-15: (a) Nexi's true arm range of motion, (b) Nexi's safe graspable range of motion.

that represents Nexi's true arm range of motion when the arm is fully extended. The origin of this arc is measured from Nexi's shoulder joint and its length extends to the ends of her end effectors. As seen in Figure 4-15(b), this arc is then reduced by 30% to determine a safe range of motion which prevents current spikes in Nexi's arm motors caused by torque maxima when the arms are fully extended at the arc limits[6]. A safe vertical range H is then found from the distance between the two vertical limits of the safe range of motion arc. H is 3.4 feet and is measured from a vertical distance of 3 inches from the ground. Thus, Nexi can safely manipulate any object located within H.

It is assumed that Nexi will always navigate to target items such that her body is always directly in front of objects and agents in the environment, thus, a horizontal grasp range of motion does not need to be considered. Further, the positions of all receiver agents in the environment are carefully selected such that object exchange happens within Nexi's safe graspable range of motion, so the robot does not need to perform this calculation when approaching receiver agents for object delivery.

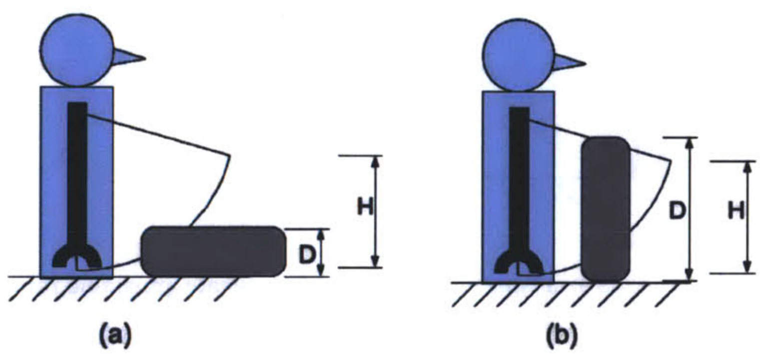


Figure 4-16: (a) Cylinder in the flat orientation outside of Nexi's graspable range of motion, (b) Cylinder in the upright orientation inside Nexi's graspable range of motion.

To further ensure that Nexi can grasp objects, a rule is imposed which states that for any object orientation, at least 31% of the object's vertical

[6] The determination of this criterion is discussed in Section 4.10.2.

dimension D must be located within the safe grasp range H in order for Nexi to be able to manipulate the object[7]. For example, Figure 4-16(b) shows that Nexi is capable of manipulating the cylinder placed on the floor in the upright position since at least 31% of D is within Nexi's safe graspable range of motion H. However, when placed in the flat position on the floor, the cylinder is not within Nexi's safe graspable range of motion no matter how close the robot navigates to the object as shown in Figure 4-16(a).

4.9.2 Grasp Capacity

The reasoning architecture uses a surface area metric to determine if the robot should grasp objects bimanually or unimanually. Figure 4-17

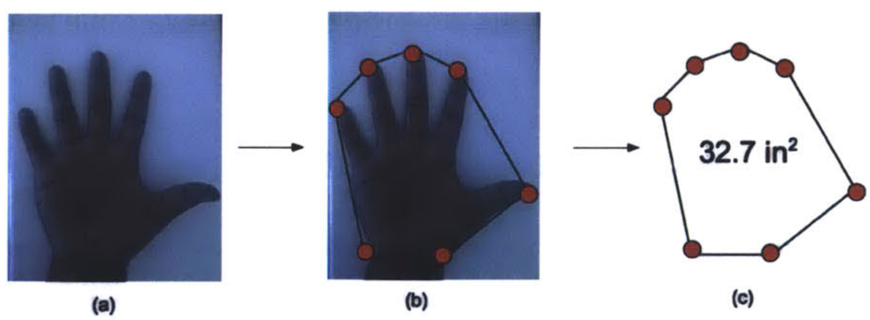


Figure 4-17: (a) Snapshot of palm, (b) Fitting points to the palm extremities, (c) Area calculation of palm polygon.

illustrates how this process is performed. First, an outline of the manipulator agent's end effector is created with the fingers spread open as wide as possible, Figure 4-17(a). Next, a polygon is generated by connecting the extreme points of the end effector, Figure 4-17(b). Finally, the area of this

[7] The determination of this criterion is discussed in Section 4.10.2.

polygon is calculated and represents the manipulator agent's palm surface area P, as seen in Figure 4-17(c). A similar analysis was performed on Nexi's end effectors and it was determined that Nexi has a P value of 16.92 in^2. P is compared with the surface area of each object and if P is at least 84% greater[8] than the surface area of the object, then the system determines that the manipulator agent is capable of grasping the object with only one hand, i.e. unimanually. If P is less than 84% greater than the surface area of the object, then the system determines that the manipulator agent must grasp the object with two hands, i.e. bimanually. The weight of the object would typically contribute to this metric also, however, for the purposes of this research, all objects are extremely lightweight and are chosen to ensure that the robot's arm and hand motors are powerful enough to overcome load torques.

4.9.3 Receiver Agent Capability

Because humans have extremely advanced dexterity and are capable of whole-body maneuvers like reaching, squatting, and bending over, the reasoning system assumes that human receiver agents have an unbounded range of motion for receiving objects. Again, the only criterion employed is that Nexi must manipulate objects with at least one handle such that the handle exposed for the human to grab upon object exchange.

[8] The determination of this criterion is discussed in Section 4.10.2.

The reasoning architecture utilizes two distinct criteria for delivering objects to Xylo. Xylo's arms are held at a fixed distance of 22 inches apart. Thus, the reasoning system determines that if the largest horizontal dimension L on the object is at least 23 inches in length, then the object can

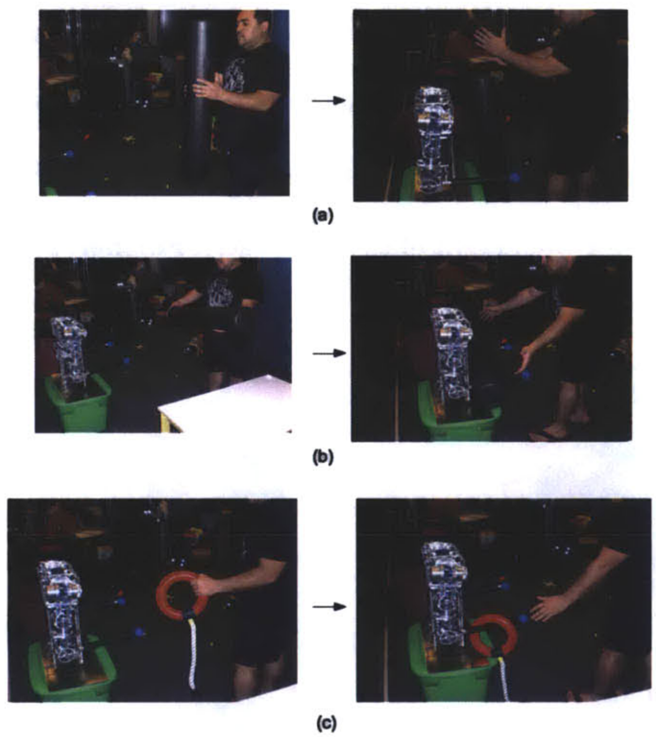


Figure 4-18: (a) Failure to lay the upright cylinder on Xylo, (b) Successfully laying the flat cylinder on Xylo, (c) Successfully hanging the ring on rope on Xylo.

be laid on Xylo's arms, Figure 4-18(b). The system adds an extra inch to the fixed distance of Xylo's arms in order to ensure safe support. Figure 4-18(a) shows an example where the object falls between Xylo's arms because it does not meet the L criterion.

If the object has a ring feature, and may or may not fulfill the *L* criterion, another criterion is applied which states that objects with ring features can be looped around Xylo's end effectors as seen in Figure 4-18(c).

The dimensions of the receptacle are carefully chosen to guarantee that Nexi will be able to clear the receptacle's height upon object delivery. The only criterion utilized by the reasoning system is that Nexi must grasp objects and retract them from their supports in such a way that leaves enough clearance for the height of the receptacle. Further, any arm retracting

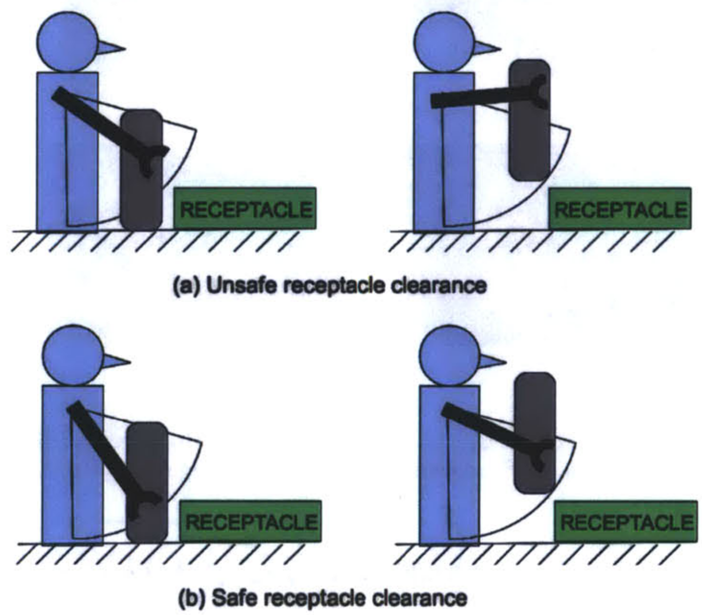


Figure 4-19: (a) Nexi unsafely clears receptacle height (b) Nexi safely clears receptacle height.

action must occur within Nexi's safe graspable range of motion. Figure 4-19 shows examples of both a safe and unsafe clearance of the receptacle's height. For instance, as seen in Figure 4-19(a), grasping the at the top of the cylinder on the floor is not preferred because Nexi must then lift her arms outside of her safe graspable range of motion in order to clear the receptacle's height.

4.10 Manipulation Training

The training object set from Figure 2-5 was used to teach the robot how to manipulate objects. Nexi was placed in the research environment and objects were placed in front of her in various orientations on various supports. Through trial and error, Nexi manipulated the objects and the data was recorded, transcribed, and represented symbolically in the reasoning architecture. Appendix A shows the full list of object combinations used for manipulation training. Appendix Orientations illustrates the conventions used for object orientation. Nexi performed each object combination three times to test the repeatability of the maneuver used.

4.10.1 Learning Affordances

In addition to learning object affordances through the pilot study discussed in Section 4.4, Nexi learns about object affordances through her own exploration. Using every maneuver in her manipulation repertoire, discussed in Section 4.3, she repeatedly tries to manipulate the object set. After each set of three trials, Nexi's attempts were labeled with a success rate S which took the value of 0, 1, or 2. If $S = 0$, then the maneuver used was a failure and should not be used to manipulate the object. If $S = 1$, then the maneuver used could possibly manipulate the object, but has a low probability of success due to potential instabilities between the object and the robot's end effectors. If $S = 2$, then the maneuver used successfully

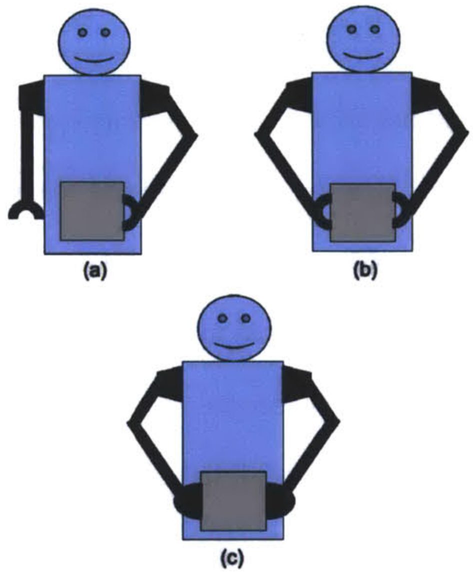


Figure 4-20: (a) Failed maneuver ($S = 0$), (b) Semi-successful maneuver ($S = 1$), (c) Extremely successful maneuver ($S = 2$).

manipulates the object with minimal or no potential instabilities. Figure 4-20 highlights some sample trials where Nexi attempted to manipulate the box. In Figure 4-20(a), she attempts to grasp the box unimanually. There is no way possible that this maneuver will successfully pick up the box, so it is labeled $S = 0$. In Figure 4-20(b), she now attempts to grasp the box bimanually with her fingers opened. It is possible that this maneuver will work, but through repetition, it is learned that the maneuver has a high probability of failure due to the small number of contact points between Nexi's fingers and the box's surface. Thus, it is labeled $S = 1$. In Figure 4-20(c), Nexi attempts to grasp the box bimanually with her fingers closed. This maneuver successfully manipulates the box and creates more contact between her fingers and the object surface, thus reducing potential instabilities. Therefore, it is labeled $S = 2$. Through this method of reinforcement learning [40], Nexi learns about the interaction between her end effectors and object features which can be used for future manipulation attempts on the experimental object set.

4.10.2 Determining Metrics

Additional testing was performed during the training process to determine certain metrics used by the reasoning architecture. For instance, to determine a safe range of motion for Nexi, as illustrated in Figure 4-15, an extreme for the motion arc was found by enforcing that arms could not extend behind the robot. At the other motion arc extreme, the specifications of Nexi's shoulder motors [41] were used to calculate a maximum angle that the arm could traverse while holding a 2-lb load such that it does not exceed the torque and current limits of the motor. Thus, the safe range of motion H is the vertical distance between these two extremes of the motion arc.

To determine the criterion shown in Figure 4-16, the cylinder and ring on rope objects were placed in the flat orientation and supported by an adjustable hanging support. Nexi was required to grasp each object as it was raised and lowered to various heights. The vertical dimension D of each object, 6 in. and 2 in. respectively for the cylinder and ring on rope, was fixed and the height of the object for each successful trial was noted. These heights were then used to find a ratio between D and Nexi's safe range of motion H. It was determined that if at least 31% of D was within H, then successful manipulation was possible.

To determine if the robot should grasp objects unimanually or bimanually, a surface area criterion is used. Once a grasp polygon is determined for Nexi, as demonstrated in Figure 4-17, trials are performed where the robot is required to grasp four different balls of increasing surface area on a table. Balls 1-4 had surface areas of 6.79 in^2, 11.522 in^2, 20.143 in^2, and 25.804 in^2 respectively. Nexi could manipulate balls 1 and 2 with one hand, but needed two hands for balls 3 and 4. Since ball 3 was the cut off between unimanual and bimanual grasping, Nexi's hand polygon area, 16.92 in^2, was divided by the surface area of ball 3, 20.143 in^2, which yielded a value 0.84. Thus, a 84% surface area ratio between the robot's hand and the object of interest was used to determine the use of bimanual or unimanual manipulation.

4.10.3 Ranking Strategies

Once the robot manipulated all objects in the training set, it sometimes discovered that different strategies could be used to successfully manipulate the same object in a given configuration. Therefore, it is necessary to have a metric that ranks the proposed strategies used for future manipulation attempts. The reasoning architecture takes in factors such as the maneuver success rate S, the handle criteria (Criteria #5) from Table 3.3, the capabilities of receiver agents, and maneuvers that require the least arm usage (bimanual vs. unimanual) for the robot to determine a rank number R for each proposed strategy. For instance, if a scene has n possible proposed manipulation strategies, a strategy with rank number $R = 1$ implies that it is the preferred strategy for manipulation. The system then ranks the remaining strategies $R = 2$, 3, 4, etc., in order of decreasing favor, until all n proposed strategies have been ranked. If the reasoning system determines that there are multiple strategies with $R = 1$ within n, then the system selects an $R = 1$ strategy to employ at random. However, this case was never seen in experimental trials.

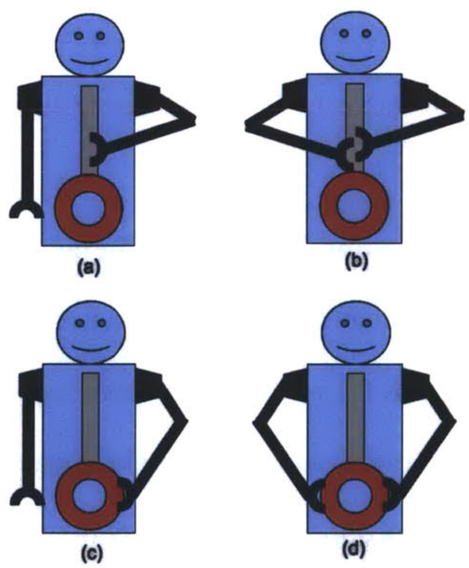


Figure 4-21: (a) Unimanual grasp at rope, (b) Bimanual grasp at rope, (c) Unimanual grasp at handle, (d) Bimanual grasp at handle.

Figure 4-21 shows examples of successful manipulation strategies for the ring on rope supported by a hanging structure. Table 4.3 shows the resulting rankings generated by the reasoning system for each strategy in

Table 4.3: Rankings generated for the ring on rope supported by a hanging structure

	STRATEGY (a)	STRATEGY (b)	STRATEGY (c)	STRATEGY (d)
XYLO	$R = 2$	$R = 3$	$R = 1$	$R = 4$
RECEPTACLE	$R = 1$	$R = 4$	$R = 2$	$R = 3$
HUMAN	$R = 1$	$R = 2$	$R = 3$	Invalid

Figure 4-21 with consideration to each receiver agent.

In general, unimanual manipulation strategies are preferred over bimanual strategies because they require less effort by the robot. Also, Criteria #5 from Table 3.3, which states that handles must always be free for human receivers, always trumps any other factor when determining rank. And because object delivery to the receptacle is fairly easy for Nexi, in these cases, the only significant factor in ranking strategies is the ease of manipulation for the robot.

According to Table 4.3 and Figure 4-21, for object delivery to Xylo, strategy (c) is preferred because it involves unimanual manipulation as opposed to strategies (b) and (d) which involve two hands. Also, strategy (c) involves grasping at the handle, which will make it easier for Nexi to place around Xylo's end effector as opposed to strategy (a) where the object could act as a pendulum and swing forcefully making it more difficult to hook the handle around Xylo's arm.

According to Table 4.3 and Figure 4-21, for object delivery to a human, strategy (a) is preferred because if leaves the entire handle open for the human to grab as opposed to strategy (c). Strategy (a) is also easier for the robot because it is performed unimanually as opposed to strategy (b). And strategy (d) is invalid because it does not leave any part of the handle free to grab, which is a direct conflict of Criteria #5 from Table 3.3.

4.10.4 Error Correction

Error correction was done with the aid of human intervention. During the experimental trials[9], if the robot failed to successfully manipulate an object, upon pick up or delivery, a human participant clicked a button on a

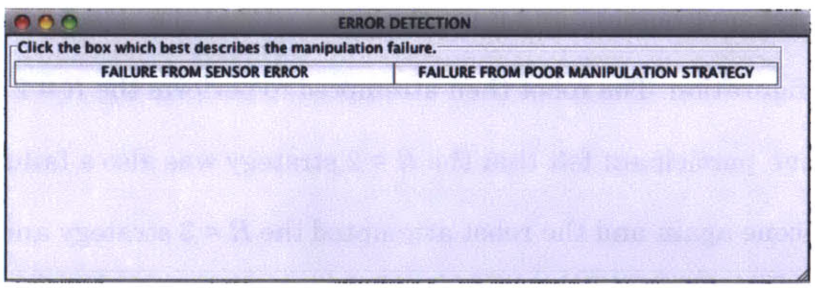


Figure 4-22: Screenshot of the Java dialogue box for error correction.

Java dialogue box within the reasoning system. Figure 4-22 shows a screen shot of the dialogue box. The participant was asked to determine whether they believed that the failure was caused by an incorrect proposed manipulation strategy or from perceptual errors caused by the vision sensors.

If the participant clicked "FAILURE FROM SENSOR ERROR", Nexi is stopped and I then performed a homing routine that calibrates the robot's body motors. This attempts to reduce sensor noise caused by repeated use of the robot. After the recalibration routine is complete, the robot is triggered by

[9] A detailed outline of the experimental setup is discussed in Chapter 5.

the reasoning system to perform the exact same proposed manipulation strategy as before. If the robot still failed to execute the manipulation and the human participant believed sensor error was the cause, then that trial was deemed to be caused by sensor error and the robot performed no alterative strategies.

If the human participant clicked “FAILURE FROM POOR MANIPULATION STRATEGY”, this implied that the proposed strategy ranked with $R = 1$ failed. Thus, the robot was stopped and I entered the experimental environment and reset the scene so that the object was in its initial configuration. The robot then attempted to perform the $R = 2$ strategy. If the human participant felt that the $R = 2$ strategy was also a failure, then I reset the scene again and the robot attempted the $R = 3$ strategy and so forth. Every failed attempt was deemed an unfit manipulation strategy for the scene, including trials where only a single strategy, with $R = 1$, was proposed by the reasoning architecture.

CHAPTER 5

Experimentation

5.1 Experimental Setup

Table 5.1: Possible scene combinations for experimentation

MANIPULATOR AGENT	RECEIVER AGENT	OBJECT	OBJECT ORIENTATION	OBJECT SUPPORT
Nexi	Xylo	Box with Handles	Upright	Hanging
	Human	Inflatable ball	Flat	Table
	Receptacle	Half Cylinder	Upside-down	Floor
		Balls on Rope		Post

A given research workspace for experimentation consists of a scene which can be any combination of manipulator agent, receiver agent, object, object orientation, and support, as shown in Table 5.1. And for a given scene, the research task is for the Nexi to manipulate objects in the environment and deliver them to the receiver agent. All scene combinations from Table 5.1 are considered for experimentation.

5.1.1 Santa's Workshop Task

The experimental environment consists of Nexi, objects in certain configurations supported by various structures, two human participants,

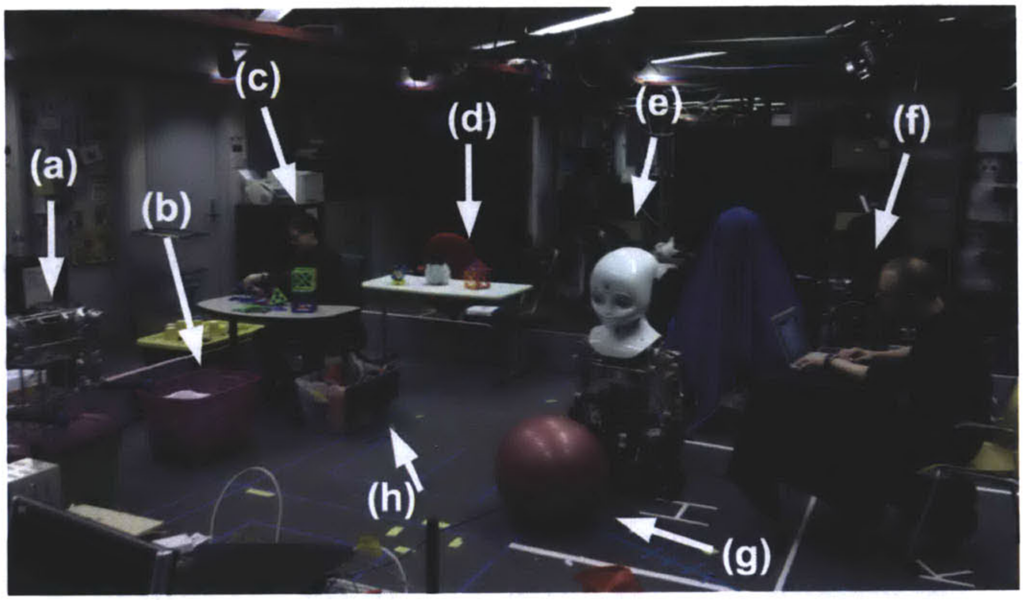


Figure 5-1: Example experimental setup. (a) Xylo as Santa's elf, (b) Receptacle as a Toy for Tots box, (c) Human participant A as Santa's helper, (d) Table for completed toys, (e) Nexi as the robot Santa, (f) Human participant B as Santa's inspector, (g) Manipulation object (Inflatable ball in the flat orientation supported by the floor), (h) Spare toy box.

Xylo, and a receptacle, as shown in Figure 5-1. A mock Santa's Workshop scenario is used to describe the research task where participants are told that the goal is to help Santa Claus build and sort as many toys as possible for distribution to the children of the world. Appendix C shows the prompt given to human participants that describes the Santa's Workshop task in detail. Nexi, Figure 5-1(e), acts as Santa Claus and her role is grasp toys, i.e. the experimental object set, Figure 5-1(g), from their supporting structures and deliver them to either Xylo, a receptacle, or human Participant A to be sorted. Xylo, Figure 5-1(a), acts as one of Santa's elves, and his role during the task is to remain static and receive toys from Nexi. The receptacle, Figure 5-1(b), acts as a Toys for Tots bin and its role is also to remain static and to collect any toys that Nexi may drop inside it. Participant A acts as Santa's helper, Figure 5-1(c), and the helper's role during the task is to build as many toys as possible from of a box of random toy parts, Figure 5-1(h), and place them on a nearby table, Figure 5-1(d), when they are finished building each toy. Participant A is also told that Nexi may deliver toys to them throughout the task while they are constructing. If this occurs, then they must receive the toy from Nexi and place it on the nearby table. Participant B, Figure 5-1(f), acts as Santa's inspector throughout the task. The inspector's role is ensure quality control by monitoring Nexi as she manipulates objects and delivers them to receiver agents. Participant B sits at a station equipped with a computer that runs the error detection interface discussed in Section 4.10.4. They were instructed to watch for failed object manipulation attempts by Nexi and to make judgments about the type of failure that was detected. If Participant B felt that Nexi's failure was caused by a faulty vision sensor and not the attempted manipulation strategy, then they were instructed to click the box on the Java dialogue window that says, "FAILURE FROM SENSOR ERROR", as shown in Figure 4-22. On the other hand, if Participant B felt that Nexi's failure was a result of an insufficient proposed manipulation strategy, then they were told to click the box on the Java dialogue window that says, "FAILURE FROM POOR MANIPULATION STRATEGY." My role as the experimenter was to operate the reasoning system, generate different scene combinations, and to create those scenes in the physical environment by placing objects in their appropriate configuration. As described in Section 4.10.4, I also intervened during manipulation failures. 12 randomly generated scene combinations were used for each experiment.

After the task was complete, both participants were asked to fill out a questionnaire regarding the Santa's Workshop task. Appendix D shows the questionnaire. Analysis of the human questionnaire data is discussed in Chapter 6.

5.1.2 Random Scene Combination Generation

According to Table 5.1, there exists 180 combinations of possible manipulation environments [42]. MATLAB [43] and Microsoft Excel [44] software was used to generate and categorize these combinations. Each scenario was labeled with a unique number to distinguish it from other scenarios. A random number generator in Java was integrated into the reasoning system for testing purposes.

During testing, the reasoning system generates a random number that is within the limits of the number of possible scene combinations. From there, the system matches this random digit with the appropriate scene according to the catalogued Excel spreadsheet of combinations. When the system displays the given scene in the Java console, the experimenter creates the scene in the physical environment with the objects, orientations, and supports. Once the correct scene is established, the experimental task if for the manipulator agent, Nexi, to manipulate the objects in the environment and deliver them to a receiver agent. The experimental task is always the same, only the environment changes between experimental trials. This process of randomly generating a scene, creating that scene in the physical world, and performing the research task is repeated until all 12 scene combinations have been considered per experiment.

5.2 Combination Reduction

Certain factors about the workspace can lead to reductions in the number of possible scene combinations to be considered for experimentation. For example, some scene combinations get eliminated because of symmetry in the object's geometry. Table 5.1 shows that every object has three possible orientations[10], upright, upside-down, and flat as shown in

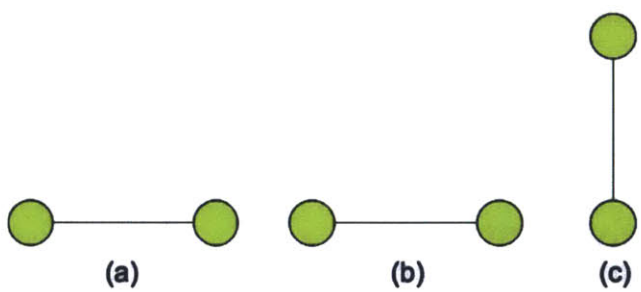


Figure 5-2: Orientations for the balls connected by rope. (a) Flat, (b) Upside-down, (c) Upright.

Appendix B. Figure 5-2 shows the possible orientations for the balls on rope object. Note that the flat orientation, Figure 5-2(a), is the same as the upside-down orientation, Figure 5-2(b). Because these orientations are the same,

[10] Orientation conventions are defined ad hoc.

only the flat and upright orientations are considered during experimentation. This analysis is done for other objects with similar geometry.

Further, scenes that contain impossible or dynamically unstable object and support configurations are eliminated. For instance, the balls on rope object in Figure 5-2 can only exist naturally in the upright configuration, Figure 5-2(c), i.e. if it is attached to a hanging support. However, because it is not a rigid structure it cannot exists naturally in the upright orientation on the floor, the table, or the post. Other scene combinations that contain similar impossible or unstable initial conditions are also eliminated.

As stated previously, there exists 180 combinations of possible manipulation environments. However, after using the combination reduction techniques discussed above, there exists 60 possible scene combinations used for testing. Appendix E shows the table of experimental combinations used. Thus, several human test subjects were recruited until all 60 scene combinations in Appendix E were tested.

5.3 System Evaluation

The reasoning architecture will be evaluated based on the following criteria:

1) *How well does the reasoning architecture develop a manipulation repertoire for Nexi?*

2) *How well does the reasoning architecture generate strategies to successfully accomplish the manipulation task?*

3) *What is Nexi's success rate for executing proposed manipulation maneuvers in the physical environment?*

These criteria will be examined using results from the human pilot study, the robot's manipulation training sessions, observations from experimental trials, and questionnaire responses from human test subjects.

CHAPTER 6

Results

6.1 Participant Demographics

To perform the experiments, human test subjects were recruited from the MIT and greater Boston area. There were 13 people in total, 9 men and 4 women, and the mean age was 32.6 years. Participants were randomly assigned to play the role of Santa's helper or Santa's inspector. It was sometimes difficult to schedule two human test subjects to perform the task during the same time block. For instance during experimental trials, 5 dyads were tested and three participants did not have a partner. Thus, 8 total experiments were performed. In the cases were only one participant was available, the task was split up into two sessions. In the first session, the human test subject played the role of Santa's inspector and the robot only delivered objects to Xylo and the bin, which eliminated the need for a human receiver. In the second session, I played the role of Santa's inspector while the human participant played the role of Santa's helper.

6.2 Example Scenarios

The following four scenarios highlight a sample of data collected from the experimental trials. For each scenario, only rank $R = 1$ manipulation strategies proposed by the reasoning architecture are featured. For simplicity, these scenarios feature the same object, the box with handles. Scenarios A, B, and C highlight successful strategies, however in Scenario D, one step in the manipulation sequence is deemed impossible to execute, thus the entire strategy cannot be performed. The robot does not perform any manipulation strategy that has at least one impossible step in its sequence.

6.2.1 Scenario A: Box with handles, flat, to the receptacle

Table 6.1: Proposed manipulation strategy for Scenario A

MANIPULATION SEQUENCE	STRATEGY IMPLEMENTED ($R = 1$)
1) Navigation approach direction to object.	Approach object from the front.
2) Animation playback for hand extension.	Extend forward with both hands.
3) IK grasp maneuver.	Bimanual Smash
4) Animation playback for hand retraction.	Upward Bimanual Retract
5) Navigation approach direction to receiver agent	Approach receiver from any direction.
6) Animation playback for hand extension.	Extend outwards with both hands.
7) IK release maneuver.	Open Fingers
8) Animation playback for hand retraction.	Bimanual Outward Retract

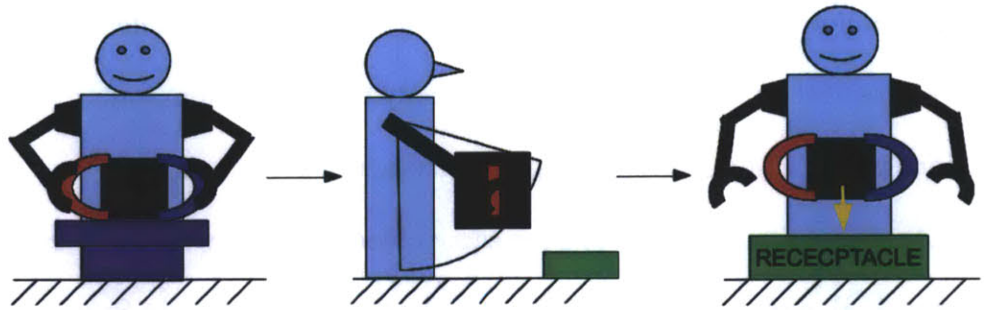


Figure 6-1: Manipulation sequence for Scenario A.

Scenario A consists of the box with handles in the flat orientation being supported by the table and the receptacle as the receiver agent. Table 6.1 and Figure 6-1 illustrate the $R = 1$ manipulation strategy proposed by the system. In step #1, the robot is suggested to approach the object from the front in order to successfully perform the Bimanual Smash maneuver on the objects handles in step #3. Because the receiver agent is the receptacle, step #5 suggests that the navigation approach direction does not matter, thus, Nexi can approach the receiver from any direction. The robot then dumps the objects inside the receptacle as shown in steps #6 - #8.

6.2.2 Scenario B: Box with handles, upright, to Xylo

Scenario B consists of the box with handles hanging in the upright orientation and Xylo as the receiver agent. Table 6.2 and Figure 6-2 illustrate the $R = 1$ manipulation strategy proposed by the system. In step #1, the robot is suggested to approach the object from the front in order to successfully perform the Bimanual maneuver in step #3. The robot performs this maneuver in order to leave one handle on the object free so that

Table 6.2: Proposed manipulation strategy for Scenario B

MANIPULATION SEQUENCE	STRATEGY IMPLEMENTED ($R = 1$)
1) Navigation approach direction to object.	Approach object from the front.
2) Animation playback for hand extension.	Extend forward with both hands.
3) IK grasp maneuver.	Bimanual Smash
4) Animation playback for hand retraction.	Downward Bimanual Retract
5) Navigation approach direction to receiver agent	Approach receiver from the front.
6) Animation playback for hand extension.	Extend outwards with both hands.
7) IK release maneuver.	Hang
8) Animation playback for hand retraction.	Bimanual Outward Retract

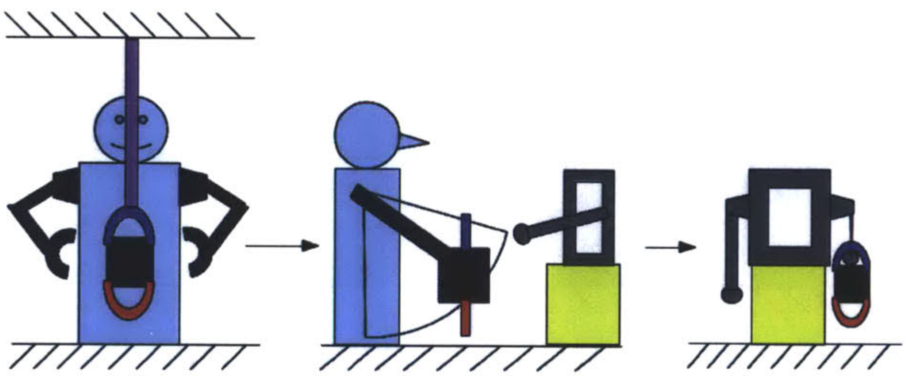


Figure 6-2: Manipulation sequence for Scenario B.

it can be hung on one of Xylo's end effectors as presented in step #7. Because Xylo is the receiver agent, step #5 suggests that Nexi approach Xylo from the front.

6.2.3 Scenario C: Box with handles, flat, to a human

Table 6.3: Proposed manipulation strategy for Scenario C

MANIPULATION SEQUENCE	STRATEGY IMPLEMENTED ($R = 1$)
1) Navigation approach direction to object.	Approach object from the side.
2) Animation playback for hand extension.	Extend forward with both hands.
3) IK grasp maneuver.	Bimanual Smash
4) Animation playback for hand retraction.	Upward Bimanual Retract
5) Navigation approach direction to receiver agent	Approach receiver from the front.
6) Animation playback for hand extension.	Extend upwards with both hands.
7) IK release maneuver.	Open Fingers
8) Animation playback for hand retraction.	Bimanual Outward Retract

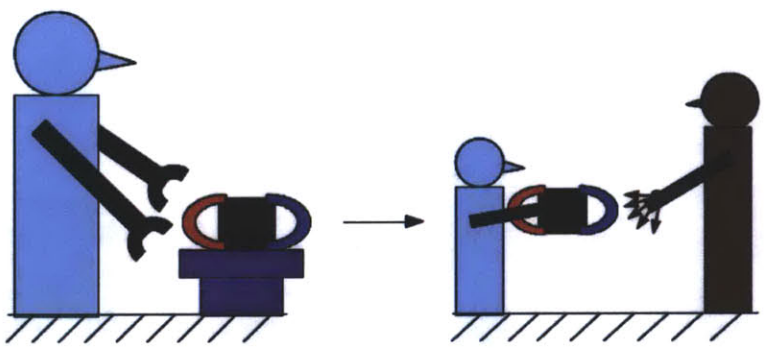


Figure 6-3: Manipulation sequence for Scenario C.

Scenario C consists of the box with handles in the flat orientation being supported by the table and a human as the receiver agent. Table 6.3 and Figure 6-3 illustrate the $R = 1$ manipulation strategy proposed by the system. In step #1, the robot is suggested to approach the object from the side in order to successfully perform the Bimanual Smash maneuver on the sides of the object, which do not have handles, as shown in step #3. The robot performs this maneuver in order to leave the handles on the object exposed so that the human receiver can grasp the object easily, as outlined in steps #6 - #7. This complies with Constraint #5 in Table 3.3. Because the receiver agent is a human, step #5 suggests that the robot approach from the front.

6.2.4 Scenario D: Box with handles, hanging flat, to Xylo

Table 6.4: Proposed manipulation strategy for Scenario D

MANIPULATION SEQUENCE	STRATEGY IMPLEMENTED ($R = 1$)
1) Navigation approach direction to object.	Approach object from any direction.
2) Animation playback for hand extension.	Extend upward with both hands.
3) IK grasp maneuver.	Bimanual Smash
4) Animation playback for hand retraction.	Downward Bimanual Retract
5) Navigation approach direction to receiver agent	Approach receiver from the front.
6) Animation playback for hand extension.	Extend upwards with both hands.
7) IK release maneuver.	IMPOSSIBLE
8) Animation playback for hand retraction.	Bimanual Outward Retract

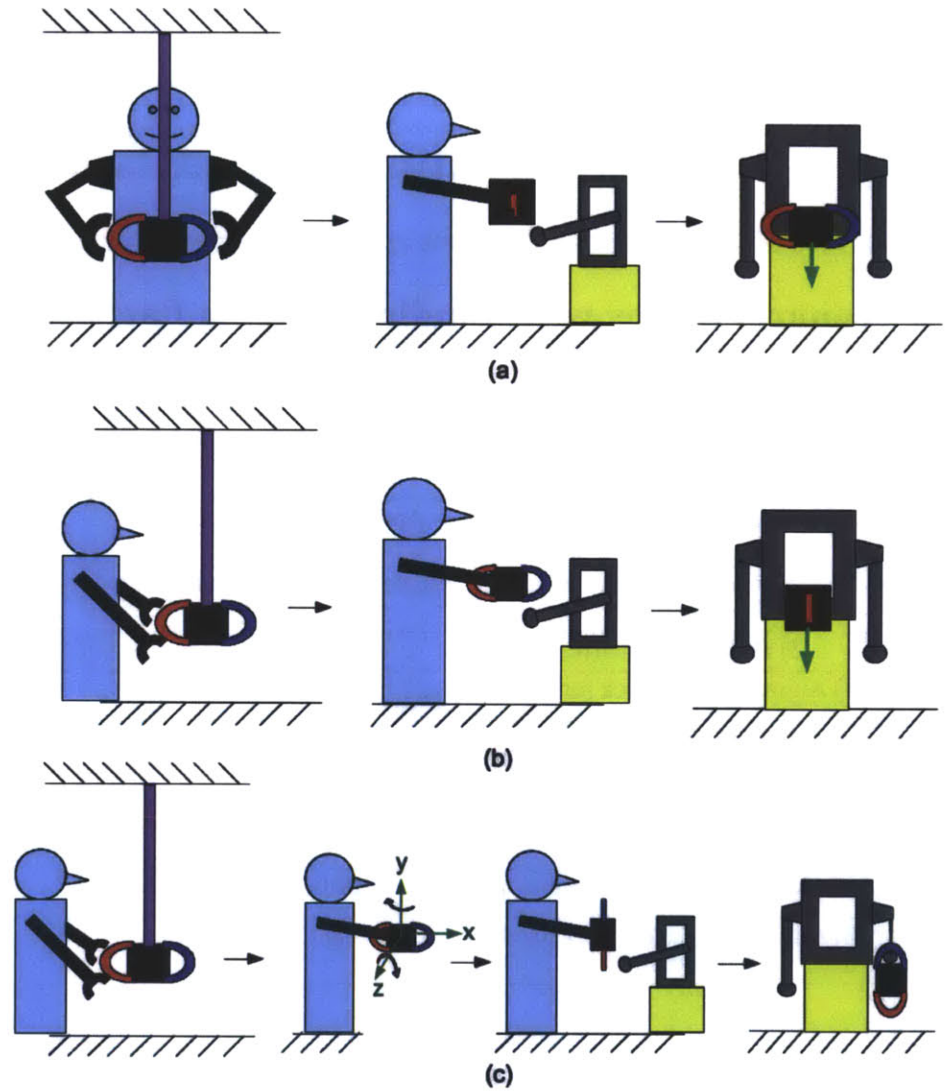


Figure 6-4: Invalid maneuvers. (a) Invalid maneuver since the object will fall, (b) Another invalid maneuver since the object will fall, (c) Invalid maneuver due to regrasping.

Scenario D consists of the box with handles in the flat orientation being supported by a hanging support and Xylo as the receiver agent. Table 6.4 and Figure 6-4(a) illustrate the $R = 1$ manipulation proposed by the system. Since the object is hanging, step #1 suggests that the navigation approach direction does not matter, thus, Nexi is told to approach the object from any direction. In step #3, the robot can perform the Bimanual Smash maneuver on any two opposing sides of the object, e.g. either the sides with handles or the sides without handles. This all depends on the initial navigation approach direction. In step #5, the robot is advised to approach Xylo from the front, but #7 shows that it is impossible to successfully deliver the object to Xylo given its initial configuration. Using criteria discussed in Section 4.9.3, the system reasons that the object would fall if Nexi tries to lay it on Xylo because the horizontal distance between Xylo's arms is larger than the width of the object. This is also shown in Figure 6-4(b), which illustrates the $R = 2$ proposed strategy. Further, as seen in Figure 6-4(c), attempting to hang the object on Xylo's end effectors would require an object regrasp, which is a direct violation of Constraint #2 from Table 3.3. Thus, the robot cannot execute any strategies and Scenario D is deemed impossible to complete by the reasoning architecture.

6.3 Experimental Results

This Section outlines the results obtained from experimental trials. 82 total trials were generated among the 8 experiments performed. Table 6.5 describes the results from the 13 questionnaires collected from human test subjects. Point values were assigned to each response on the questionnaire; 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree. The minimum, maximum, mean, and standard deviation values of all responses were tabulated for each question.

Table 6.5: Analysis of questionnaire responses

	Question	Sample Size	Min	Max	Mean	SD
1	The robot grabbed every toy successfully each time.	13	1	4	2	1.043907845
2	If you were playing the role of Santa, you would have grabbed the toys the exact same way as the robot.	13	2	5	4	1.126601424
3	The robot's failures were caused mainly because she used an incorrect way of grabbing the toys.	13	1	3	2	0.800640769
4	There were better ways for the robot to grab the toys that she did not use.	13	1	5	2	1.3634421
5	The robot delivered toys to Xylo successfully.	13	4	5	5	0.506369684
6	The robot failed to grab the toys most of the time.	13	1	5	3	1.450022104
7	It was difficult to keep track of the errors the robot made.	13	1	1	1	0
8	You could have done a better job of grabbing the toys than the robot.	13	1	5	3	1.739436985
9	The robot delivered toys to the Toys for Tots bin successfully.	13	4	5	5	0.43852901
10	The robot's failures were caused mainly because her hands could not fit around the shape of the toys.	13	1	4	2	1.091928428
11	The robot delivered toys to the human helper successfully.	13	4	5	5	0.43852901
12	There would be more toys sorted if you were playing Santa instead of the robot.	13	4	5	5	0.277350098
13	The robot's failures were caused mainly because her hands were slightly off when she tried to grab the toys.	13	3	5	5	0.630425172
14	If you were playing the role of Santa, more toys would have been sorted successfully.	13	5	5	5	0
15	The robot interrupted you many times when you were trying to build toys.	13	1	2	1	0.277350098
16	The robot should have dropped the toys on the floor next to you when she approached your post instead of handing them off to you.	13	3	5	4	0.59914469
17	It was difficult to take toys from the robot.	13	1	3	1	0.650443636
18	You could not take toys from the robot many times because you were busy building toys.	13	1	3	1	0.650443636
19	You had to take objects from the robot in ways that did not feel natural.	13	1	2	1	0.277350098
20	You took objects from the robot in natural way.	13	3	5	5	0.554700196
21	Receiving toys would have been easier if the robot were a human.	13	1	3	2	0.660225292
22	The robot delivered toys to you easily.	13	3	5	5	0.776250026
23	The robot sometimes had to speak in order to get your attention when delivering toys.	13	1	1	1	0

The data from the following four figures comes from observations during the 82 experimental trials. Figure 6-5 shows the overall success rate for Nexi executing manipulation strategies in the physical environment, irrespective of whether the failure was caused by sensor error or a poor manipulation strategy. Figure 6-6 shows a comparison of the type of failure modes experienced by the robot, either sensor error failure or faulty strategy failure. Figure 6-7 highlights the robot's rate of success when grasping objects from their supports. Figure 6-8 showcases the robot's success rate for delivering objects to receiver agents.

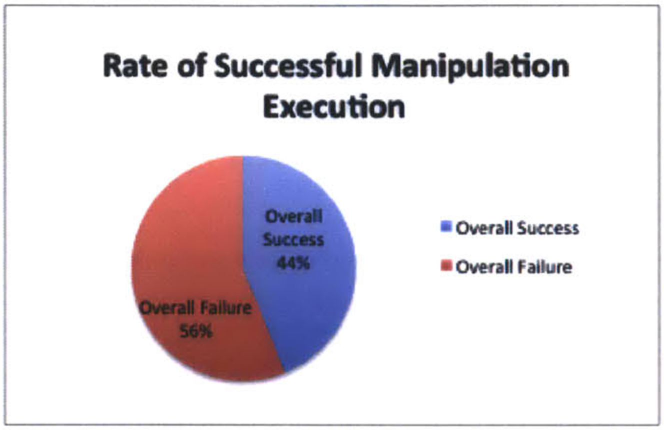


Figure 6-5: Overall success rate for manipulation execution.

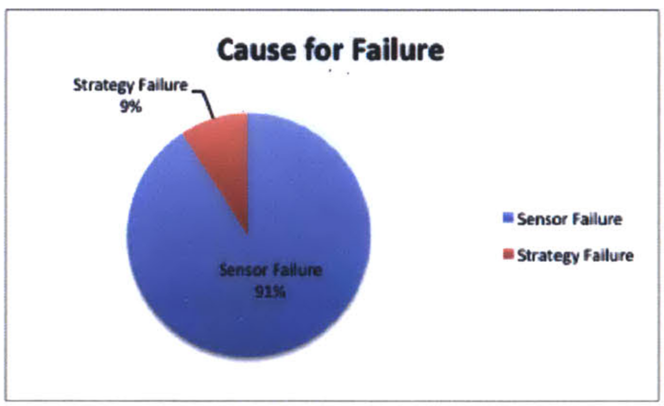


Figure 6-6: Comparison of failure modes.

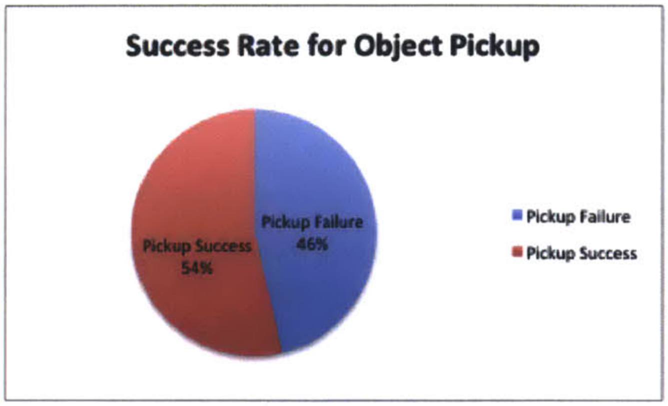


Figure 6-7: Success rate for object pickup.

Figure 6-8: Success rate for object delivery.

6.3.1 Repertoire Development

As stated in Section 5.3, the first benchmark for evaluating the reasoning system is: *How well does the reasoning architecture develop a manipulation repertoire for Nexi?* Figure 6-9 compares the average responses to questions regarding the effectiveness of the robot's manipulation repertoire at solving the manipulation task (questions Q2 an Q4). Many people felt that

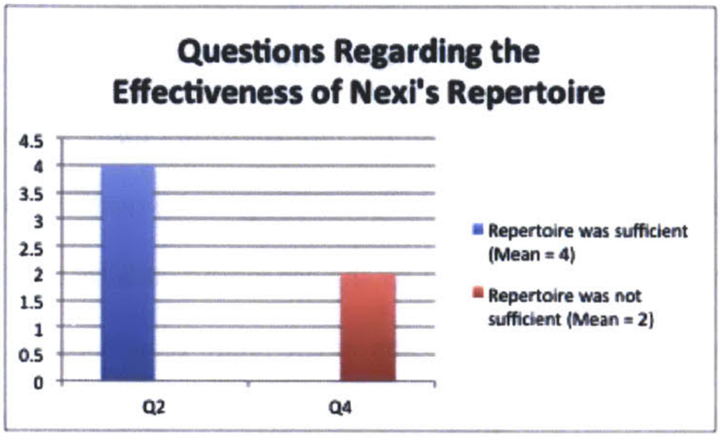


Figure 6-9: Mean responses to questions regarding the effectiveness of the robot's manipulation skillset.

the robot's skillset was sufficient (mean response = 4) for accomplishing the given task. Only a few participants felt that Nexi's repertoire was not sufficient (mean response = 2). This shows that the general trend was that Nexi came equipped with a satisfactory collection of dexterous maneuvers to solve the research task, which validates Hypothesis I.

6.3.2 Strategy Generation

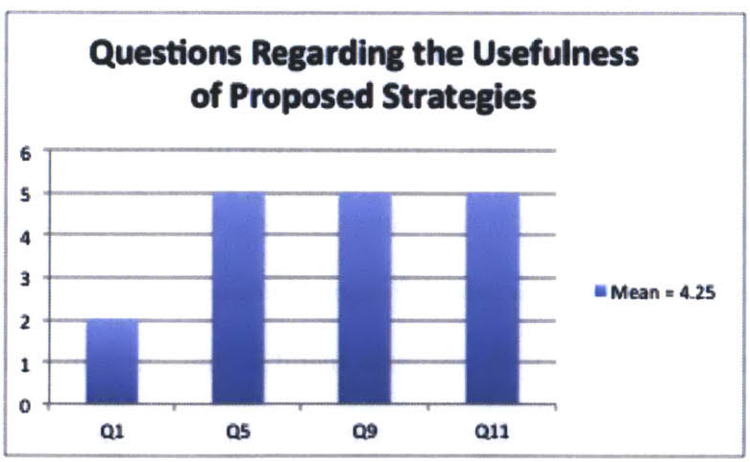


Figure 6-10: Mean responses to questions regarding the usefulness of proposed manipulation strategies.

As mentioned in Section 5.3, the second benchmark for evaluating the reasoning system is: *How well does the reasoning architecture generate strategies to successfully accomplish the manipulation task?* Figure 6-10 shows the mean values for questions pertaining to the usability of manipulation strategies generated by the reasoning architecture (questions Q1, Q5, Q9, and Q11). Figure 6-11 shows the mean responses to questions regarding the uselessness of proposed strategies (questions Q3, Q6, and Q16).

Figure 6-11: Mean responses to questions regarding the inapplicability of proposed manipulation strategies.

Overall, people felt that proposed strategies were useful (Mean = 4.25) to robot whereas fewer people felt that these strategies were useless (Mean = 3). Although there is only a moderate difference between the mean values in Figure 6-10 and Figure 6-11, the general trend is that participants believed that strategies performed by the robot were useful at solving the manipulation tasks. These results reasonably validate Hypothesis II.

According to participants who played the role of Santa's helper in the experimental task, receiving objects from the robot was relatively easy. Figure 6-12 represents the mean replies to questions concerning the difficulty of receiving objects from the robot (Q16, Q17, Q18, Q15, Q19, and Q23). Figure 6-13 denotes the mean responses to questions that consider the easiness of object exchange with the robot (Q11, Q20, and Q22).

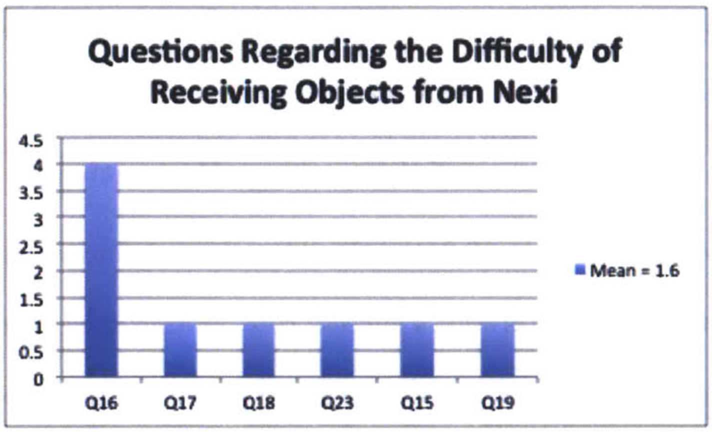


Figure 6-12: Mean responses to questions regarding the difficulty of receiving objects from the robot.

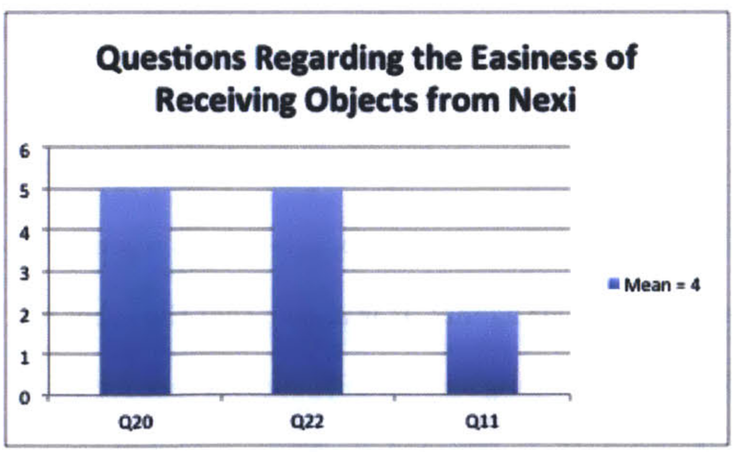


Figure 6-13: Mean responses to questions regarding the ease of receiving objects from the robot.

People largely felt that is was not difficult to exchange objects with the robot (Mean = 4) as opposed to only of few people who did (Mean = 1.6). Despite a few outlier responses, this shows a significant trend that participants felt comfortable receiving objects from Nexi and did so in a natural and fluid manner. Further, it was noted that robot did not need to give any verbal utterances to catch the attention of the human upon object delivery. All participants stopped what they were doing when Nexi approached their station. These results validate Hypothesis IV.

6.3.3 Successful Strategy Execution

As indicated in Section 5.3, the third benchmark for evaluating the reasoning system is: *What is Nexi's success rate for executing proposed manipulation maneuvers in the physical environment?* Figure 6-5 shows that robot was only 44% successful at performing manipulation. However, this low success rate encompasses all failures caused by sensor errors and inadequate strategies, as well as failures during both object grasping and object delivery attempts.

Figure 6-7 and Figure 6-8 denote the success rates for when the robot attempted to grasp objects from their supports versus when the robot delivered objects to receiver agents. It is shown that the robot had a 54% success rate for object grasping and an 89% success rate for object delivery. This suggests that the robot struggled much more to manipulate objects, but was incredibly effective at delivering objects to the appropriate receivers.

Finally, Figure 6-6 compares the failure modes that were caused by sensing errors versus those caused by inadequate strategies suggested by the reasoning scheme. The data suggests that 91% of failures were caused by perceptual errors in the robot's vision sensors whereas only 9% of failures were caused by a lacking manipulation strategy. Further, Figure 6-14 summarizes the mean responses from human participants which considers

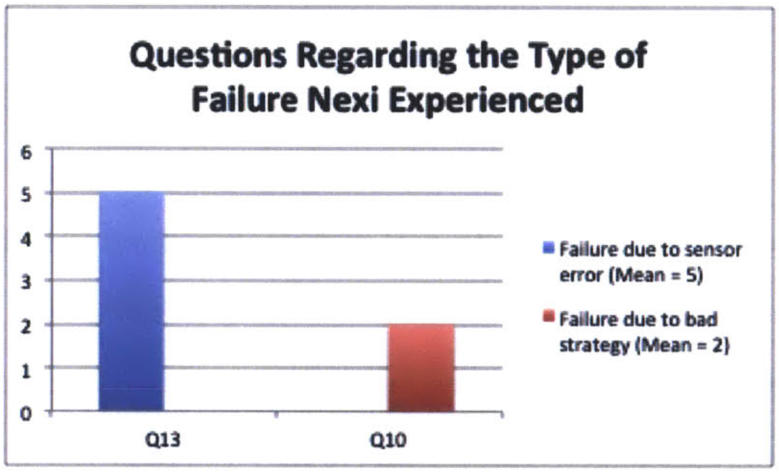


Figure 6-14: Comparison of the responses to the robot's failure modes.

questions regarding their interpretation of the failure modes experienced by the robot, i.e. whether they were from sensor error (Q13) or from a poor strategy (Q10). The overall trend is that people felt that the robot failed to perform manipulation maneuvers due to sensor errors (Mean = 5) rather than inadequate strategies suggested by the reasoning architecture (Mean = 2), which reasonably validates Hypothesis III.

6.4 Discussion

The results presented above provide evidence that the reasoning architecture developed in this work helps the robot solve new manipulation tasks by leveraging a corpus of physics-, social-, and capability-based knowledge.

Results from Section 6.3.1 indicate that the reasoning system excels at developing a manipulation repertoire for Nexi to use during future manipulation tasks. This suggests that system accurately interprets and transforms observed human actions into robust maneuvers, given the unique kinematics of the robot. Thus, Hypothesis I is strongly validated.

Results from Section 6.3.2 reveal that people generally felt that the robot's attempted strategies could have successfully performed the manipulation tasks. These outcomes moderately validate Hypothesis II. However, only a small difference is found in the mean responses of participants who felt this way (Mean = 4.25) and the responses of those who did not (Mean = 3). I believe this small gap is a result of the broad nature of certain questions that are relevant to this topic. For instance, Q1 states, "The robot grabbed every toy successfully each time." The mean response to this statement was 2, meaning that people strongly disagreed with this statement. Although this question is relevant to investigation of manipulation strategy generation, it does not explicitly explain the cause of manipulation successes. Thus, this brought down the average to 4.25 for questions regarding successful strategy generation. Additionally, Q16 states, "The robot should have dropped the toys on the floor next to you when she approached your post instead of handing them off to you." The mean response to this statement was 4, meaning that people strongly agreed. Again, this is valid for examining strategy generation, but the robot was never trained to drop objects on the floor when delivering them to a human. Thus, the average for questions concerning poor strategy generation increased to 3. Therefore, there was a smaller gap between the two average values (Mean = 4.25 and Mean = 3).

Results from 6.3.2 also express that the manipulation strategies spawned by the reasoning system are exceptionally useful for object hand-offs to human partners. Overall, participants experienced natural, fluid object exchanges with the robot. This is indicated by the mean responses from human test subjects who experienced positive interactions with Nexi (Mean = 4), versus the mean responses of those who did not (Mean = 1.6). Additionally, the robot did not need to rely on a verbal utterance to grab its partner's attention during any of the experimental trials. These results successfully validate Hypothesis IV.

Hypothesis III can be reasonably validated from outcomes in Section 6.3.3. The data indicates that although the robot had a low overall success rate for accomplishing manipulation tasks (44% success rate), the majority of failures were caused by perceptual errors in the robot's vision sensors (91% sensor failures) versus insufficient manipulation strategies generated by the reasoning scheme (9% strategy failures). The data also suggests that the robot was exceedingly effective at delivering objects to receivers agents (89% delivery success), but struggled much more to grasp objects from their supporting structures (11% grasp success). This could be because it is much easier for the robot to dump an object into a static receptacle or give an object to a human partner who may take a more dominant role in the exchange process, thus requiring less dexterity from the robot.

CHAPTER 7

Conclusions and Future Work

7.1 Contributions

The research question proposed was: *How can a dexterous robot utilize physics-, social-, and capability-based knowledge of its environment for object manipulation?* The approach to this problem was the development of a reasoning architecture that uses object affordances, social contexts, and agent capabilities to determine probabilistic factors that are used in a Bayesian network system that selects appropriate manipulation maneuvers in a given workspace. This contribution is unique in that it fuses these three research concepts together, which has yet to be fully explored in robotic manipulation.

7.1.1 Physics-Based Reasoning

Physics-based reasoning was showcased by the reasoning system's ability to determine a manipulation repertoire for Nexi as well as the robot's ability to apply previously learned information to new manipulation tasks. The robot learned about object affordances through a combination of self-exploration during manipulation training and knowledge observed from a pilot study where a human test subject manipulated various objects. This knowledge was represented symbolically and used as input for a system of Bayesian networks in the reasoning architecture. The system then used outcomes from the Bayesian analysis to propose manipulation strategies for the robot to execute when manipulating new objects whose affordances are hybrids of previously manipulated objects previously. According to results of experimental trials, it was found that the system successfully developed a manipulation repertoire for Nexi and the robot performed reasonably well in new manipulation environments using her knowledge corpus.

7.1.2 Social-Based Reasoning

Social-based reasoning was demonstrated through Nexi's ability to deliver objects to a human receiver while considering factors that arise when performing object exchange with a human. Nexi considered grasping constraints imposed by the reasoning system, face detection from vision sensors, and the attentiveness of the human participant to coordinate successful object hand-offs. Results from questionnaires presented to human participants in experimental trials showed that the robot performed exceptionally well when delivering objects to them. Moreover, participants felt that they experienced natural, fluid interactions with Nexi.

7.1.3 Capability-Based Reasoning

Capability based reasoning was showcased by the system's ability to generate successful manipulation strategies given constraints in the environment and kinematic limitations of the robot. Metrics determined through manipulation training sessions were used by the Bayesian network system to determine valid maneuvers for the robot to execute for new manipulation tasks. Results from experimental trials showed that the system reasonably generated effective maneuvers for the robot to perform. However, the robot did not successfully execute these strategies in the physical environment as frequently due to sensor inaccuracies.

7.2 Research Limitations

7.2.1 Physics Simulators

One weakness of this work is that the physical properties of objects, like volume, weight, and unique geometric features, must be hardcoded. Future work could include the integration of the robots on-board cameras for object feature extraction. The use of a physics simulator [45], [46] or a commonsense reasoning architecture [47] could also potentially be integrated into the reasoning architecture. This could simplify the process of learning object affordances during manipulation training.

7.2.2 Tactile Sensing

Nexi lacks sophisticated tactile sensing in her palms and fingers. As opposed to a capacitive sensing scheme, Nexi uses the current draw in her hand motors to approximate the tactile forces. However, this is not an accurate way to determine the forces and moments experienced by Nexi's hands during object manipulation. To sidestep these limitations, rules and constraints were imposed on the system so that Nexi could grasp objects given her unique manipulation repertoire. With enhanced tactile sensors, maneuvers like object re-grasping and re-orientation could be considered.

As stated in Section 4.7, navigation waypoints and graspable points of interest on objects are determined through trial and error in order to train the robots. Location and orientation inaccuracies of points of interest in the environment are remedied by iteratively tweaking navigation and manipulation parameters in R1D1. However, feedback from additional sensors could correct perceptual errors caused by noisy Vicon and robot eye cameras in real-time during task execution.

7.2.3 Limited Repertoire Development

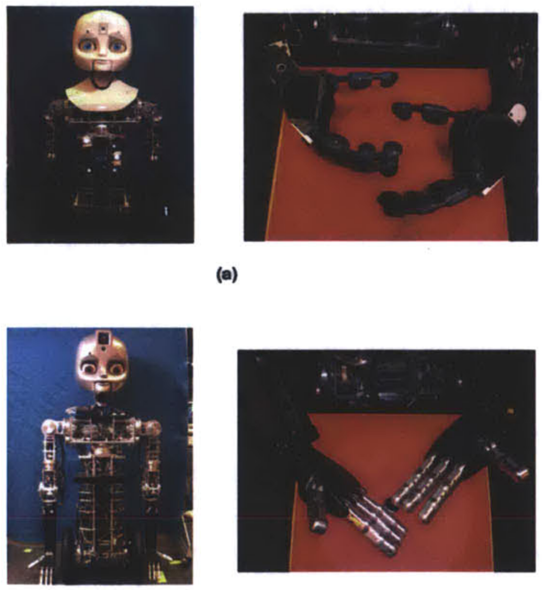
Nexi's manipulation skillset is limited to the maneuvers that it saw a human perform during a pilot study. For this work, only two human test subjects were used for the pilot trials. Clearly, manipulation maneuvers will vary between different people, thus, if more human subjects were tested, then the robot's own manipulation repertoire could expand.

7.3 Future Research Endeavors

7.3.1 Spatial Reasoning

For this work, objects and receiver agents are located in open spaces so that the robot can easily manipulate objects without bypassing obstructions in the environment. However, another layer of reasoning that could be integrated into the architecture is spatial reasoning, as demonstrated in [48] and [49]. The spatial reasoning logic could determine maneuvers that the robot could perform to circumvent obstacles when objects and agents are partially or fully occluded by supporting structures or other items in the environment.

7.3.2 Robotic Receiver Agents



(a)

(b)

Figure 7-1: MDS platforms Nexi and Maddox. (a) Nexi and her end effectors, (b) Maddox and his end effectors.

Our research group's 2nd generation MDS platform, called Maddox, could also be used for this work. Figure 7-1(b) shows Maddox and his end effectors. This work does not consider Nexi or Maddox as receiver agents because of complexity. However, it is important in robotic manipulation to explore the unique dynamics and coordination difficulties that arise when dexterous robots act as agents to receive objects from other agents, particularly other robots, as demonstrated in [50]. Also, as displayed in Figure 7-1, Maddox's hands are capable of positioning themselves in fully open configurations, as opposed to Nexi whose hands are naturally fixed in a U-shaped configuration. Thus, Maddox has the potential to perform more complex dexterous maneuvers than Nexi. Further, if Maddox acted as a manipulator agent, then an entirely new manipulation repertoire would be established which could be vastly different than that of Nexi's.

7.4 Broader Impacts

As stated before, robots that can perform everyday manipulation tasks in human-centric environments have the potential to help humans. Robots must be able to reason about the maneuvers that they implement despite the challenging environments that they are expected to operate in.

7.4.1 Cultural Implications

Societal norms can sometimes have a major impact on the manipulation strategies that people choose to implement. These cultural

Figure 7-2: Japanese custom of receiving and delivering objects bimanually.

routines can trump any physics- or logic-based reasoning used for object manipulation in social settings. For instance, when giving out a business card, most people from the United States of America would deliver the card to the receiver using one hand. Logically, this makes sense because the card is small and lightweight. However, in Japan it is customary to give out and accept business cards with two hands [51]. Generally speaking in Japanese culture, delivering and receiving objects viewed as gifts bimanually is a sign of respect [52], as seen in Figure 7-2. Robots that interact in culturally diverse environments must be aware of such traditions.

Figure 7-3: Women from Tanzania carrying large objects.

In many African, Caribbean, and Latin American countries [53], people carrying large objects using whole-body manipulation with their head as a support is common, Figure 7-3. However, many people from other parts of the world [54] are more likely to perform whole-body manipulations using their chest, stomach, shoulders, or sides as supports, Figure 7-4. In the context of

Figure 7-4: Child from the U.S.A. carrying large objects.

human-robot teaming for object manipulation, knowledge of these cultural differences may alter the way the robot offers support to its human partner. This information is not necessarily useful to the robot when considering its own manipulation strategies because whole-body manipulation with its head as a support may cause major instabilities or damage to the robot. However, this information becomes important for the robot when assisting a human with a whole-body manipulation. For example, if a robot in Tanzania is delivering an object to a human who intends to perform a whole-body manipulation, then the robot knows that there is a high probability that the human will support the object with their head. Therefore, the robot may extend its arms at a higher elevation to get closer to the region of interest for object exchange. On the contrary, a robot in the United States assisting with the same task may extend its arms at a much lower elevation because it is aware that the human is more likely to support the object with their torso instead.

Figure 7-5: Waiter serving food in a formal dining setting.

In formal dining situations, like those in Figure 7-5, many servers are trained to serve food from either the left or right side of the diner and to collect items from the opposite side of the diner [55]. The appropriate side for which the waiter should serve and collect items differs according the country's rules of etiquette [56]. Again, from a purely logical standpoint, a navigation approach to the left or to the right of the diner is sufficient to deliver the items, however the rules of dining protocol supersede those of logic. If robots are to act as servers in formal dining situations, then they must be aware of these cultural practices.

7.4.2 Future Scenarios

Imagine an environment where a dexterous robot assists a human during a cooking task. As the robot delivers ingredients and tools to the human chef, it must consider the physical properties of the objects to determine its best strategies to manipulate these items. Further, in the social context, the robot should consider the chef's safety and comfort during object hand-off, e.g. it should manipulate objects like a knife such that the handle is available for the chef to grab, not the sharp end. Further, the robot assistant must also consider the cook's attentiveness and readiness to receive these items. If the cook is occupied and cannot receive the object at the time of delivery, then the robot must reason about strategies to either attract the cook's attention or perform alternative actions until the cook is ready to receive the objects.

Or imagine a disaster scenario where robots are working alongside other robots and humans to rescue victims and find potential hazards, like bombs and toxic materials. Dexterous robots must consider their own manipulation abilities and those of others in these complex domains. For instance, robots may be better equipped to handle dangerous, brute force tasks like hazard removal while humans handle more delicate manipulation tasks like taking human vitals in the hot zone.

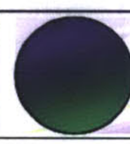
In both of these future scenarios, object affordances, social contexts, and agent abilities all play a key role in determining the manipulation strategies implemented by robots.

Appendices

Appendix A: Manipulation Training Scene Combinations

	OBJECT	SUPPORT	ORIENTATION	NAVIGATION APPROACH
1	cylinder	floor	flat	front
2	cylinder	floor	flat	side
3	cylinder	floor	upright	front
4	cylinder	table	flat	front
5	cylinder	table	flat	side
6	cylinder	table	upright	front
7	cylinder	hang	flat	front
8	cylinder	hang	flat	side
9	cylinder	hang	upright	front
10	ring on rope	floor	flat	side
11	ring on rope	floor	flat	front
12	ring on rope	table	flat	side
13	ring on rope	table	flat	front
14	ring on rope	hang	flat	side
15	ring on rope	hang	flat	front
16	ring on rope	hang	upright	front
17	ring on rope	hang	upright	side
18	flat plate	floor	flat	front
19	flat plate	floor	flat	side
20	flat plate	table	flat	front
21	flat plate	table	flat	side
22	flat plate	hang	flat	front
23	flat plate	hang	flat	side
24	flat plate	hang	upright	front
25	flat plate	hang	upright	side
26	flat plate	post	flat	front
27	flat plate	post	flat	side
28	box	floor	flat	front
29	box	table	flat	front
30	box	hang	flat	front
31	ball	floor	flat	front
32	ball	table	flat	front
33	ball	hang	flat	front
34	ball	post	flat	front

Appendix B: Object Orientation and Navigation Conventions

	ORIENTATION			NAVIGATION APPROACH	
OBJECT	**FLAT**	**UPRIGHT**	**UPSIDE -DOWN**	**FRONT**	**SIDE**
Cylinder			**Same as Flat**		
Half Cylinder					
Ring on Rope			**Same as Flat**		
Balls on Rope			**Same as Flat**		
Box		**Same as Flat**	**Same as Flat**		**Same as Front**
Box with Handles			**Same as Flat**		
Flat Plate			**Same as Flat**		
Foam Ball		**Same as Flat**	**Same as Flat**		**Same as Front**
Inflatable Ball		**Same as Flat**	**Same as Flat**		**Same as Front**

Appendix C: Santa's Workshop Task Prompt

In this task, you will be working in Santa's Workshop. The goal is to build and sort as many toys as possible for the children of the world. A robot Santa will repeatedly grab toys from a holding station and deliver them to a nearby sorting station. The sorting station will contain 3 posts which has, 1) one of Santa's elves named Xylo, 2) a Toys for Tots box, and 3) Santa's helper. The robot Santa will decide which of these three posts he delivers toys to. You will play the role of either Santa's helper or Santa's inspector.

Instructions for Santa's Helper

Your role as Santa's helper is to build as many toys as possible using the spare parts in the toy box next to your post. There are no limitations regarding the type of toys you can build. Once you are finished building a toy, you must place it on the table beside you and start building the next toy. Throughout the task, the robot Santa may come over to your post and deliver toys to you. If this happens, then you must take the toy from the robot and place it on the table beside you. After this exchange, you can continue building toys.

Instructions for Santa's Inspector

Your role as the inspector is to ensure quality control of the toy sorting process. You will sit at a station equipped with a computer and watch as the robot Santa repeatedly takes toys from the holding station over to the sorting station. If the robot fails to successfully pick up a toy or deliver it to the sorting station, then you must track this error. If you feel that the way the robot attempted to grab the toy was ok, but it failed because its hands were not close enough to the toy, then you must click the box on the computer screen that says, "FAILURE FROM SENSOR ERROR". However, if you feel that the failure was because there is no way possible for the robot to successfully grab the toy using with the grasping technique that it tried, then you must click the box on the computer screen that says, "FAILURE FROM POOR MANIPULATION STRATEGY." It is important to watch the robot carefully so that you can correctly identify the type of error.

Appendix D: Santa's Workshop Task Questionnaire

<u>Complete this section **only** if you played the role of Santa's helper</u>

	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE
If you were playing the role of Santa, more toys would have been sorted successfully.					
The robot interrupted you many times when you were trying to build toys.					
The robot should have dropped the toys on the floor next to you when she approached your post instead of handing them off to you.					
It was difficult to take toys from the robot.					
You could not take toys from the robot many times because you were busy building toys.					
You had to take objects from the robot in ways that did not feel natural.					
You took objects from the robot in natural way.					
Receiving toys would have been easier if the robot were a human.					
The robot delivered toys to you easily.					
The robot sometimes had to speak in order to get your attention when delivering toys.					

Complete this section **only** if you played the role of Santa's inspector

	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE
The robot grabbed every toy successfully each time.					
If you were playing the role of Santa, you would have grabbed the toys the exact same way as the robot.					
The robot's failures were caused mainly because she used an incorrect way of grabbing the toys.					
There were better ways for the robot to grab the toys that she did not use.					
The robot delivered toys to Xylo successfully.					
The robot failed to grab the toys most of the time.					
It was difficult to keep track of the errors the robot made.					
You could have done a better job of grabbing the toys than the robot.					
The robot delivered toys to the Toys for Tots bin successfully.					
The robot's failures were caused mainly because her hands could not fit around the shape of the toys.					
The robot delivered toys to the human helper successfully.					
There would be more toys sorted if you were playing Santa instead of the robot.					
The robot's failures were caused mainly because her hands were slightly off when she tried to grab the toys.					

Appendix E: Scene Combinations Used for Experimentation

	OBJECT	SUPPORT	ORIENTATION	RECEIVER
1	1/2 cylinder	floor	flat	bin
2	1/2 cylinder	floor	flat	human
3	1/2 cylinder	floor	flat	xylo
4	1/2 cylinder	floor	upright	bin
5	1/2 cylinder	floor	upright	human
6	1/2 cylinder	floor	upright	xylo
7	1/2 cylinder	floor	upside-down	bin
8	1/2 cylinder	floor	upside-down	human
9	1/2 cylinder	floor	upside-down	xylo
10	1/2 cylinder	table	flat	bin
11	1/2 cylinder	table	flat	human
12	1/2 cylinder	table	flat	xylo
13	1/2 cylinder	table	upright	bin
14	1/2 cylinder	table	upright	human
15	1/2 cylinder	table	upright	xylo
16	1/2 cylinder	table	upside-down	bin
17	1/2 cylinder	table	upside-down	human
18	1/2 cylinder	table	upside-down	xylo
19	1/2 cylinder	hang	flat	bin
20	1/2 cylinder	hang	flat	human
21	1/2 cylinder	hang	flat	xylo
22	1/2 cylinder	hang	upright	bin
23	1/2 cylinder	hang	upright	human
24	1/2 cylinder	hang	upright	xylo
25	1/2 cylinder	hang	upside-down	bin
26	1/2 cylinder	hang	upside-down	human
27	1/2 cylinder	hang	upside-down	xylo
28	balls on rope	floor	flat	bin
29	balls on rope	floor	flat	human
30	balls on rope	floor	flat	xylo
31	balls on rope	table	flat	bin
32	balls on rope	table	flat	human
33	balls on rope	table	flat	xylo
34	balls on rope	hang	flat	bin
35	balls on rope	hang	flat	human
36	balls on rope	hang	flat	xylo
37	balls on rope	hang	upright	bin

38	balls on rope	hang	upright	human
39	balls on rope	hang	upright	xylo
40	box w/ handles	floor	flat	bin
41	box w/ handles	floor	flat	human
42	box w/ handles	floor	flat	xylo
43	box w/ handles	table	flat	bin
44	box w/ handles	table	flat	human
45	box w/ handles	table	flat	xylo
46	box w/ handles	hang	flat	bin
47	box w/ handles	hang	flat	human
48	box w/ handles	hang	flat	xylo
49	box w/ handles	hang	upright	bin
50	box w/ handles	hang	upright	human
51	box w/ handles	hang	upright	xylo
52	inflatable ball	floor	flat	bin
53	inflatable ball	floor	flat	human
54	inflatable ball	floor	flat	xylo
55	inflatable ball	table	flat	bin
56	inflatable ball	table	flat	human
57	inflatable ball	table	flat	xylo
58	inflatable ball	hang	flat	bin
59	inflatable ball	hang	flat	human
60	inflatable ball	hang	flat	xylo

Appendix F: Physics-Based Logical Rules

PHYSICS-BASED LOGICAL RULES	
Rule #1	Liquids assume the shape of the container that houses them.
Rule #2	If an object has a cavity, it may be used as a container to hold other objects.
Rule #3	If an object is not supported, it will fall downwards under the influence of gravity.
Rule #4	A pendulum may experience oscillatory motion if it is supported by its pivot point.
Rule #5	Objects attached by a flexible member are constrained to move in space by the motion limits of the member.
Rule #6	A translational force may cause a flat surface to slide on top of another flat surface if the force overcomes the static friction between the surfaces.
Rule #7	An object has 6 degrees of freedom (translations and rotations about the x, y, and z axes) unless it is constrained by a support.
Rule #8	Humans usually prefer to grab objects by their handles, grips, or levers.
Rule #9	Round objects can roll if they are supported by a surface.
Rule #10	If the sum of forces and moments experienced by a body is equilibrium, the body will be at rest.
Rule #11	An object A may fit into the cavity of object B if object A's volume is smaller than object B's volume.
Rule #12	A body's center of mass is the average location of all the mass on that body.
Rule #13	A magnet is a material that produces a magnetic field.
Rule #14	Opposite magnetic poles attract each other.
Rule #15	Similar magnetic poles repel each other.
Rule #16	Many metals attract magnets.
Rule #17	Humans cannot manipulate extremely hot objects with their bare hands
Rule #18	Humans cannot manipulate extremely cold objects with their bare hands
Rule #19	If the sum of forces and moments experienced by a body is not equilibrium, the body will move.
Rule #20	Gases assume the shape of the container that houses them.
Rule #21	Solids have a definite size and shape.
Rule #22	Solids do not assume the shape of the container that houses them.
Rule #23	A triangle has 3 sides and 3 corners.
Rule #24	A square has 4 sides and 4 corners.
Rule #25	A rectangle has 4 sides and 4 corners.
Rule #26	A cone may remain at rest if it is supported by a surface on its flat side.
Rule #27	A pyramid may remain at rest if it is supported by a surface on its flat side.

Rule #28	A body is dynamically unstable if its center of gravity is not directly above the center of its support polygon with respect to the supporting surface.
Rule #29	A body that does not have a round feature will not roll about an axis attached to itself if it experienced by a force on a surface.
Rule #30	A body may rotate about a pivot point on the supporting surface if it is experienced by a force hat causes a positive moment with respect to its mass center.
Rule #31	Many metals conduct electricity.
Rule #32	Foam is a deformable material.

Bibliography

[1] J. J. Gibson, *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin, 1979.

[2] J. J. Gibson, *The Theory of Affordances*, In Perceiving, Acting, and Knowing, 1977.

[3] M. K. Lee, J. Forlizzi, S. Kiesler, M. Cakmak, and S. Srinivasa, "Predictability or Adaptivity? Designing Robot Handoffs Modeled from Trained Dogs and People, *International Conference on Human Robot Interaction,* March 2011.

[4] D. Droeschel, J. Stückler, D. Holz, and S. Behnke, "Towards Joint Attention for a Domestic Service Robot – Person Awareness and Gesture Recognition using Time-of-Flight Cameras," *International Conference on Robotics and Automation,* May, 2011.

[5] L. Kopp and P. Gärddenfors, "Attention as a Minimal Criterion of Intentionality in Robots," *Cognitive Studies,* vol. 89, Lund University.

[6] A. Needham, "How Infants Grasp Two Adjacent Objects: Effects of Perceived Display Composition on Infants' Actions," *Developmental Science,* vol. 2, number 2: 219-233, 1999.

[7] E. Oxtop, N. S. Bradley, and M. A. Arbib, "Infant Grasp Learning: A Computational Model," *Experimental Brain Research,* vol. 158: 480-503, 2004.

[8] J. Schmolze, *"Physics for Robots", AAAI Magazine,* March 1986.

[9] E. Davis, "The Naïve Physics Perplex", *AAAI Magazine,* vol. 19, number 4, December 1998.

[10] B. Moldovan, M. van Otterlo, P. López, J. Santos-Victor, and L. de Raedt, "Statistical Relational Learning of Object Affordances for Robotic Manipulation," *International Conference on Inductive Language Programming,* June 2011.

[11] M. Lopes, F. Melo, and L. Montesano, "Affordance-based imitation learning in robots," *International Conference on Intelligent Robots and Systems*, 2007.

[12] D. Katz, Y. Pyuro, O. Brock, "Learning to Manipulate Articulated Objects in Unstructured Environments Using a Grounded Relational Representation," *Proceedings of Robotics: Science and Systems* IV, pp 254-261. 2008.

[13] B. Adorno, A. Lanari Bó, P. Fraisse, and P. Poignet, "Towards a Cooperative Framework for Interactive Manipulation Involving a Human and a Humanoid," *The International Conference on Robotics and Automation,* June 2011.

[14] A. Edsinger and C. Kemp, "Human-Robot Interaction for Cooperative Manipulation: Handing Objects to One Another," *Robotic and Human Interactive Communication,* August 2007.

[15] F. Zacharias, C. Borst, and G. Hirzinger, "Capturing Robot Workspace Structure: Representing Robot Capabilities," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, USA, Oct 29 - Nov 2, 2007.

[16] H. Liu, Exploring Human Hand Capabilities Into Embedded Multifingered Object Manipulation, *IEEE Transactions on Industrial Informatics*, vol. 7, number 3, August 2011.

[17] D. Isla, R. Burke, M. Downie, and B. Blumberg, "A Layered Brain Architecture for Synthetic Creatures," *The International Joint Conference on Artificial Intelligence,* July 2001.

[18] C. Breazeal, M. Siegel, M. Berlin, J. Gray, R. Grupen, P. Deegan, J. Weber, K. Narendran, and J. McBean, "Mobile, Dexterous, Social Robots for Mobile Manipulation and Human-Robot Interaction," *Special Interest Group on Computer Graphics and Interactive Techniques,* August 2008.

[19] "Motion Capture Systems From Vicon," 2004, http://www.vicon.com/

[20] L. Chang, S. Srinivasa, and N. Pollard, "Planning pre-grasp manipulation for transport tasks," *The International Conference on Robotics and Automation,* May 2010.

[21] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory-motor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24: 15-26, 2008.

[22] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence*, vol. 29, number 3: 241-288, September 1986.

[23] T. Bayes, "An Essay towards solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London*, vol. 53: 370-418, 1763.

[24] E. Davis, *Representations of Commonsense Knowledge,* San Mateo, California: Morgan Kaufmann Publishers, Inc., 1990.

[25] R. Resnick, D. Halliday, and K. Krane, *Physics, Volume I,* 5th Edition, New York, New York: John Wiley & Sons Inc., 2002.

[26] "American Society for Testing and Materials (ASTM) International-Standards Worldwide," 2010, http://www.astm.org/

[27] J. Pearl, "Probabilistic Reasoning in Intelligent Systems," *Morgan Kaufmann*, San Francisco, CA, 1988.

[28] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data," *Journal of Computational Biology*, vol. 7, number (3/4): 601-620, August 2000.

[29] J. Uebersax, "Genetic Counseling and Cancer Risk Modeling: An Application of Bayes Nets," *Ravenpack International,* Marbella, Spain, 2004.

[30] F. Cozman, "JavaBayes 0.346 – User Manual and Download," 1998, http://www.cs.cmu.edu/~javabayes/Home/

[31] Point Grey Research, "Stereo Vision Camera", 2010, http://www.ptgrey.com/

[32] S. Krotosky, S. Cheng, and M. Trivedi, "Face Detection and Head Tracking using Stero and Thermal Infrared Cameras for "Smart" Airbags: A Comparative Analysis," *IEEE Intelligent Transportation Systems Conference,* October 2004.

[33] J. Kruman and G. Kirk, "Video Occupant Detection for Airbag Deployment. Applications of Computer Vision," In *WACV,* 1998.

[34] M. Devy, A. Giralt, and A. Marin-Hernandez, "Detection and Classification of Passenger Seat Occupancy Using Stereovision," *IEEE Intelligent Vehicles Symposium,* 2000.

[35] P. Hart, N. Nilsson, B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," In *SIGART Newsletter,* vol. 37: 28-29, 1972.

[36] S. LaValle, *Planning Algorithms.* Cambridge: Cambridge University Press, 2006.

[37] "Maya – 3D Animation, Visual Effects & Compositing Software – Autodesk," 1999, http://usa.autodesk.com/maya/

[38] C. Breazeal, A. Brooks, J. Gray, M. Hancher, J. McBean, D. Stiehl, and J. Strickon. Interactive robot theater. *Communications of the Association for Computing Machinery,* 46:76–85, 2003.

[39] E. Umex-Eronini, *System Dynamics and Controls,* CL-Engineering, 1998.

[40] M. N. Nicolescu and M. J. Mataric, "Learning and Interacting in Human-Robot Domains," In *IEEE Transactions on Systems, Man and Cybernetics,* vol. 31, issue 5: 419-430, 2001.

[41] MICROMO- The Faulhaber Group, "Miniature Drive Systems Catalogue", 2011-2012, http://www.faulhaber.com/

[42] MathIsFun, "Combinations and Permutations", 2011, http://www.mathsisfun.com/combinatorics/combinations-permutations.html

[43] Mathworks©, "MATLAB- The Language of Technical Computing", 2001, http://www.mathworks.com/products/matlab/index.html

[44] Microsoft© Office, "Microsoft Excel-Spreadsheet-Office.com", http://office.microsoft.com/en-us/excel/

[45] E. Coumans, et. al., "Bullet- Game Physics Simulation", 2011, http://bulletphysics.org/

[46] J. Jerez and A. Suero, "Newton Game Dynamics", 2010, http://newtondynamics.com/

[47] P. Singh, T. Lin, E. Mueller, G. Lim, T. Perkins, and W. Li Zhu, "Open Mind Common Sense: Knowledge Acquisition from the General Public," *Lecture Notes in Computer Science,* vol. 2519: 1223-1237, 2002.

[48] E. Akin Sisbot, R. Ros, and R. Alami, "Situation Assessment for Human-Robot Interactive Object Manipulation," *IEEE RO-MAN*, Atlanta, GA, USA, 2011.

[49] H. Jang, H. Moradi, S. Hong, S. Lee, and J. Han, "Spatial Reasoning for Real-Time Robotic Manipulation," *IROS*, 2006.

[50] A. Agah and K. Tanie, "Human Interaction with a Service Robot: Mobile Manipulator Handing Over an Object to a Human," In *IEEE International Conference on Robotics and Automation*, 1997.

[51] "Japanese Customs and Traditions", 2011, http://news.toky.com/page/2/

[52] "Japanese Etiquette Mistakes", 2010, http://raffatourism.blogspot.com/2010/06/top-10-japanese-etiquette-mistakes.html

[53] S. Sigrid, "Life's Great Mysteries", 2010, http://shanghaisigrid.typepad.com/blog/lifes-great-mysteries/

[54] K. Knight, "Benefits Christmas", 2011, http://www.dailymail.co.uk/news/article-1340653/Christmas-benefits-Eloise-Littles-spent-3k-taxpayers-money-presents.html

[55] Food & Beverage Underground, "Increase Your Waiter Tips by Keeping Your Hands Full", 2011, http://www.foodandbeverageunderground.com/waiter-tips.html

[56] S.E. Smith, "What is Proper Serving Etiquette?", 2011, http://www.wisegeek.com/what-is-proper-serving-etiquette.htm