Identifying Perceptually Indistinguishable Objects: Is That the Same One You Saw Before?

John F. Santore and Stuart C. Shapiro
Department of Computer Science and Engineering
and Center for Cognitive Science
201 Bell Hall
University at Buffalo, The State University of New York
Buffalo, N.Y. 14260-2000
{jsantore|shapiro}@cse.buffalo.edu

Abstract

We are investigating a simulated cognitive robot that, when it sees an object perceptually indistinguishable from one it has seen before, will use reasoning to decide if they are two different objects or the same object perceived twice. We are currently conducting experiments with human subjects to determine what strategies they use to perform this task and how well they perform it.

Identifying Perceptually Indistinguishable Objects

We are investigating how an artificial agent can, by reasoning, identify perceptually indistinguishable objects. Two objects are perceptually indistinguishable to an agent if the agent cannot find any difference in their appearance by using its sensors. Thus one agent may find two objects perceptually indistinguishable but another may find the same two objects perceptually distinguishable.

By identifying perceptually indistinguishable objects we mean the following: when an agent finds an object that is perceptually indistinguishable from one it has encountered before, the agent identifies the object if it successfully decides if the object is the same one it encountered previously, or if it is a new object. If the object has been encountered before, and the agent has encountered more than one such object before, the agent should also know which one it is currently encountering.

People (human agents) often encounter objects that are perceptually indistinguishable from objects that they have seen before. Sometimes this object is, in fact, the object they have seen before and sometimes it is a new object. To identify these objects we need to use background knowledge and contextual cues. Humans regularly accomplish this task in everyday situations. If a person has a copy of the latest Harry Potter book in their bookcase and, upon visiting a friend, they see the latest Harry Potter book in the friend's bookcase, the person intuitively knows that there are two books. The person might exclaim "I have the same book at home." If you have a pruned tree in your yard, and see one that is perceptually indistinguishable to you as you drive to work, you will intuitively know that this is a different tree.

However, people also make mistakes in identifying such objects. Many people have have picked up someone else's book and walked away thinking it was their own copy of the book. People have also been surprised to find themselves talking to the identical twin of the the person that they thought they were talking to.

We hypothesize that several properties of an object will be useful in identifying it. Some cues will very likely lessen the importance of other cues when the two conflict.

We think the object's location is very important. An object in place X that appears to be just like the object that was previously in place X is likely to be the same object.

The mobility of an object is also likely to be important. Some objects are essentially immobile, like trees, some can be moved, like books, and some move on their own, like people. We hypothesize that the less mobile the object is, the more location can be used as a reliable cue to identify an object.

We hypothesize that some kind of temporal knowledge is useful for reasoning about the identity of perceptually indistinguishable objects. An object that an agent is continuously perceiving will logically always be the same object(Pollock, 1974). Generally the longer it has been since an agent last perceived an object, the less certain the agent can be about the identity of a perceptually indistinguishable object that the agent later encounters. If an agent sees an object destroyed, it can assume that an object encountered later is not the same one, even if it is perceptually indistinguishable from the first object.

It seems important to know how common objects of a particular type are. People are usually unique, so it is not unreasonable to assume that a person who is perceptually indistinguishable from one seen before, is the same person. Identical twins are of course the exception to this general rule and can lead people to fail to successfully identify them. Stamps, in contrast to people, are very common. If one takes a stamp out

of a drawer, puts the stamp on a letter, and mails the letter, the next day when one takes a stamp out of the drawer, it intuitively seems to be a different stamp.

Cognitive robots must have a way of associating, or connecting, the robot's concepts with objects in the world. Symbol anchoring is the process of creating and maintaining in time these connections between mental symbols and real world objects(Coradeschi and Saffiotti, 2001). Coradeschi and Saffioti also note that, in a cognitive robot, the connection "must be dynamic, since the same symbol must be connected to new percepts when the same object is re-acquired."(Coradeschi and Saffiotti, 2001)

For a cognitive robot, identifying perceptually indistinguishable objects is a special case of the general problem of symbol anchoring. When an agent encounters two perceptually indistinguishable objects, the same perceptual "sense data" must be connected to different symbols. For instance, two copies of the latest Harry Potter book will provide a robot identical sense data, but they are different objects, so the robot needs different mental symbols for them. This is the complement of the problem of an agent's receiving different sense data from the same object. When an agent looks at the right side of a Pepsi vending machine and sees only the right side and front of the machine, the agent will get different sense data than if the agent is looking from the left side of the machine and sees the left side and front of the machine. In this paper we are only concerned with the problem of identifying an object that has the same sense data as a previously encountered object.

Since people often identify perceptually indistinguishable objects so effortlessly, we would like to give our robot the same strategies that people use. We want to know what cues humans use when they try to identify perceptually indistinguishable objects. We are currently conducting a series of experiments with human subjects to learn how people identify perceptually indistinguishable objects. We will use the subjects' actions, and their self-reported reasons for those actions, to identify what background knowledge people use to identify perceptually indistinguishable objects. We want to know what strategies they use in different situations. We are also interested to see which strategies are more likely to fail. We can then give our robot those strategies that seemed to be most successful.

Our Simulated Cognitive Robot

We are developing a simulated cognitive robot named Cassie, to whom we will give the ability to identify perceptually indistinguishable objects. Cassie currently uses vision to perceive objects in the world. She will use background knowledge and reasoning to identify objects that she finds perceptually indistinguishable. The goal is to give Cassie sufficient background knowledge and identification strategies to do as well at this task as a person can.

Cassie is the generic name given to cognitive agents that are based on the GLAIR robotic architecture (Henry Hexmoor, 1993; Hexmoor and Shapiro, 1997). The simulated robot discussed in this paper is the newest version of Cassie. For a description of previous hardware and software versions of Cassie see (Shapiro, 1998).

GLAIR (Grounded Layered Architecture with Integrated Reasoning) is a three layered robot architecture for cognitive robots and intelligent autonomous agents. GLAIR allows the replacement of the lower layers while keeping the upper layer constant. This allows Cassie's "mind" to be moved to another "body".

The KL (Knowledge Level) is the top level of the GLAIR architecture. The KL provides the "conscious reasoning" for the system. This high level reasoning is implemented using the SNePS(Shapiro and Rapaport, 1992; Shapiro and the SNePS Implementation Group, 1999) knowledge representation and reasoning system. Atomic Symbols in the KL are terms of the SNePS logic(Shapiro, 2000). Symbol structures are functional terms in the same logic(Shapiro, 2000; Shapiro, 1993). All terms denote mental entities rather than objects in the world.

The PML (Perceptuo-Motor Level) is the middle layer of the architecture. At this layer, routine behaviors, including the primitive acts of the KL, are represented and carried out. To continue our anthropomorphic analogy, the PML is where unconscious skills and behaviors reside.

The SAL (Sensory Actuator Level) is the lowest level in the GLAIR architecture. The actual sensors and effectors of Cassie's robotic body reside at this level. The SAL is the level of the very primitive actions that control the sensors and effectors.

The GLAIR architecture anchors Cassie's intensional KL terms to objects in the world(Shapiro and Ismail, 2001). GLAIR is a solution to the problem of symbol anchoring described by Coradeschi and Saffiotti. Cassie's KL concepts of real world entities are aligned with high level processed sensory data from the PML. The PML in turn is responsible for producing processed sense data from the low level raw sensory perceptions of the SAL.

Crystal Space

Crystal Space is the environment that our version of Cassie exists in. Crystal Space is an open source 3D graphics and gaming engine(Jorrit Tyberghein, 2002). The Crystal Space graphics engine provides a visual interface similar to that of id Software's Doom and Quake Games (id Software,). Crystal Space is designed as a modular set of tools for creating graphical applications. It is written in C++ and runs on a wide variety of platforms.

The Crystal Space project consists of several independent modules so users only need to use the features they want. The graphics engine itself provides rendering of an arbitrary three dimensional virtual environment with moving 3D sprites. The Crystal Space engine is capable of rendering a scene from both the

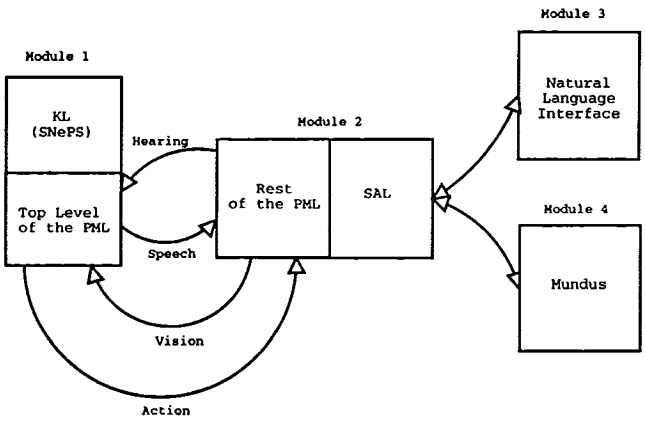

Figure 1: Architecture of the Crystal Space version of Cassie

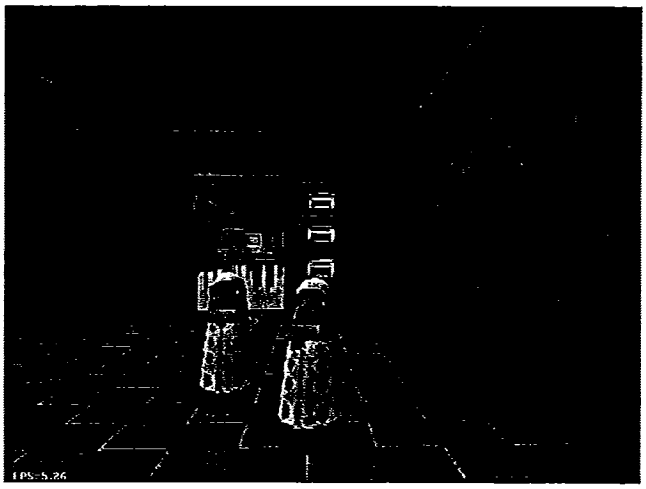

Figure 2: Cassie's view of the world showing two perceptually indistinguishable robots, one of whom she is following. A file cabinet stands against the wall, and a computer room is visible through the door.

first and third person perspective. Other Crystal Space modules provide believable physics and collision detection. We are using these modules to build a three dimensional virtual world that our simulated robot will interact with.

Cassie in a Crystal Space Environment

We are developing this version of Cassie using the Crystal Space tools. The interaction between Cassie and the Crystal Space environment is encapsulated in four modules as shown in figure 1. The modules communicate through standard socket connections. Each connection represents a specific functional connection between the two modules.

The first module implements the KL and some parts of the PML. This module is implemented entirely in Common Lisp. SNePS runs in this module, along with the ATN(Shapiro, 1989) that Cassie uses to understand a fragment of English.

The second module implements the remaining parts of the PML and the SAL. It is written in C++ using the Crystal Space tools. This module regulates the connections between all of the modules.

The third module provides the natural language interface to Cassie. Currently this is typed natural language interaction. Later we intend to use spoken interactions using speech to text technology.

The forth module, the mundus, implements the world itself and Cassie's interaction with the world. The mundus uses the Crystal Space graphics engine to render what Cassie sees. The simulation renders a first person perspective of the world because it renders exactly what Cassie sees at any given time. Figure 2 shows an example rendering of one such scene. The mundus also receives the actions of Cassie's effectors and processes them.

There are four one way connections between the KL/PML module and the SAL/PML module. The first two connections represent the two KL sensory modalities that our robot has, vision and hearing (for natural language input). The other two represent the two acting modalities that our robot is capable of, speech and physical actions in the world. There is a single connection between the SAL and the natural language input and output module which handles all natural language interaction. There is also a single connection between the SAL module and the Mundus module. This two way connection provides vision information to the SAL module for processing and communicates Cassie's low level actions to the mundus.

Our current working version of Cassie will respond to simple directional commands to move around in the world. By the time of the workshop, we expect to have a version capable of more advanced commands.

Cassie has three sensory modalities at the SAL/PML level which we shall refer to as vision, hearing and bump detection. Bump detection is only used in the service of movement, to provide feedback about collisions; no bump information is passed up to the KL level. The hearing modality is entirely devoted to natural language interaction. Cassie uses vision to perceive objects in the world.

We are not concerned, in this paper, with the processing of sensor data into sense data so we will not discuss vision in the SAL level. We will concentrate on visual perception at the PML level.

We represent visual information at the PML level as a two dimensional feature vector. The dimensions of the feature vector are shape and material. Some of the possible values for shape are generic, such as "box shaped", and some are more specific. Any object with a flat horizontal surface supported by four vertical pillars has the shape value "table shaped", for example. Materials are the visual appearance of an object's texture. Materials can also be generic or specific. The material "wooden" is a generic material, while "Harry Potter front cover"

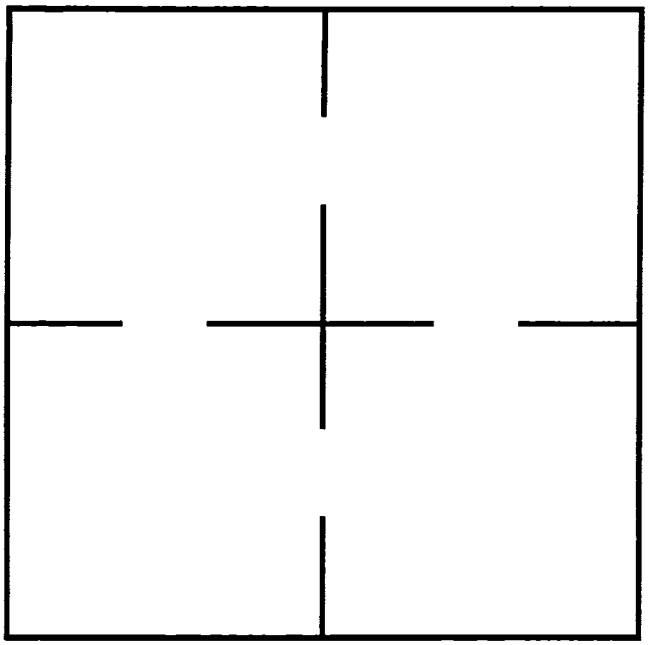

Figure 3: Floor plan of the four room worlds

is a specific material. Objects can have only a single shape, but may have more than one material.

Cassie finds an object to be perceptually indistinguishable from an object she has seen before if the objects have the same shape, and share all the same values for their materials. If she sees an object with a shape value of "table shaped" and a single material value of "wooden" then she can identify this as a wooden table. If she sees an object with a shape value of "box shaped" and material values of "Harry Potter front cover", "Harry Potter book spine" and "book pages" lying on the table shaped object, she can identify the object as a Harry Potter book. If Cassie goes into another room and sees an object with shape value "box shaped" and the three material values "Harry Potter front cover", "Harry Potter book spine" and "book pages", the new object will be perceptually indistinguishable from the first. Cassie will have to rely on her reasoning to decide if it is the same object, or new one.

The simulated worlds

The simulated worlds we are using in the Crystal Space environment are based on two floor plans. Both worlds are closed suites of rooms in a building. The floor plan of the first world is a simple square, subdivided into four equal sized, interconnected square rooms. Figure 3 shows this floor plan. The other world is a model of part of an academic building, with 8 rooms connected by three corridors. Figure 4 shows this larger floor plan. The screenshot shown in figure 2 shows part of this second suite of rooms. Using these two floor plans, we create different test worlds by using different materials for the walls, floors, and ceilings of the

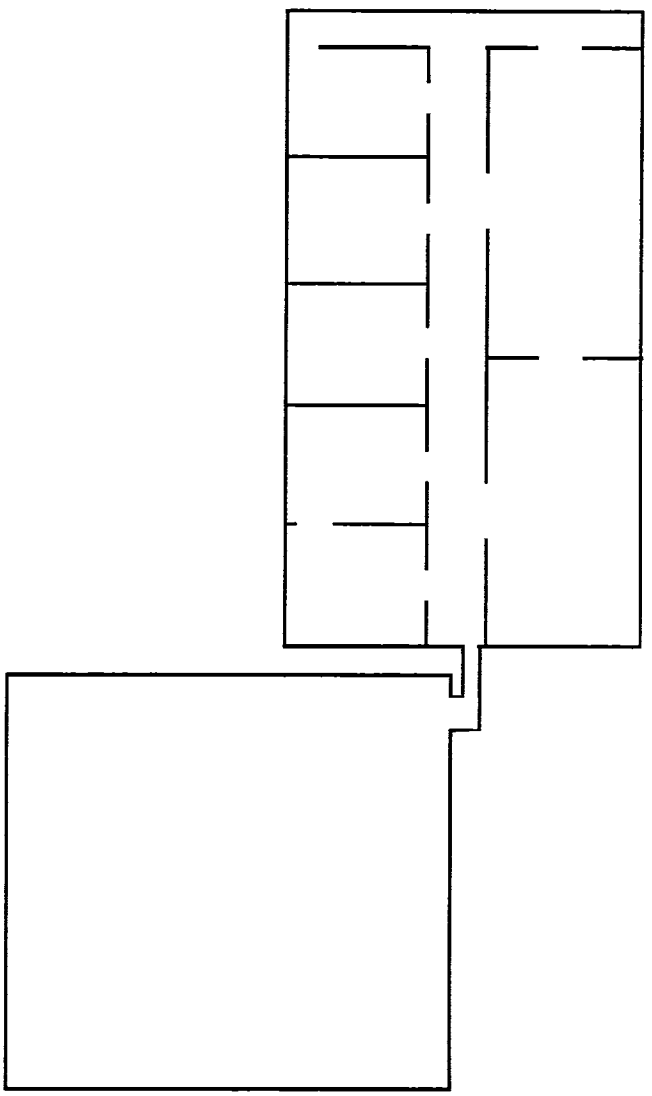

Figure 4: Floor plan of the large worlds

Object	Quantity
Table	13
Chair	25
Monitor	10
Keyboard	10
Computer	10
FileCabinet	1
PepsiMachine	1
Stove	1
WhiteBoard	1
BulletinBoard	2
Poster	2
Robot	5-6
Machine	1
Book	1
Car	1
Person	0-5
Bottle	1
Glass	2

Table 1: List of objects and how many of each there are in the larger simulated world.

rooms, and by placing different objects in the of rooms of the world. Some worlds built using the smaller floor plan have chairs, tables, glasses and bottles while others have only tables and robots. The worlds we've built using the floor plan of the larger suite contain all of the objects listed in table 1. All of the worlds built using the larger suite's floor plan include a computer room, a lab, two class rooms, a lounge and a parking garage. These rooms are filled with appropriate objects.

Experiments with Human Subjects

We have designed a set of experiments to elicit the strategies people use to identify perceptually indistinguishable objects. These experiments are also designed to gauge how well people can identify perceptually indistinguishable objects; we will compare Cassie's performance with human performance. Human performance is a measurable benchmark of what is reasonable to expect of Cassie.

We describe the experiments, along with some preliminary results, below. We will able to present more complete preliminary results of these experiments at the workshop.

Materials and Apparatus

For the experiments with human subjects, we are using the same environment that we are using for our current version of Cassie. The program that the subjects use is functionally the same as the "mundus" module from figure 1. Using this program subjects interact with the exact same virtual worlds that Cassie will interact with. Subjects use their eyes to see the same first person view of the world that Cassie sees through the socket connection. Subjects use keyboard navigation to move themselves around the world where Cassie sends action requests through the socket connection. Subjects have the same movement limitations that Cassie has.

Design and Procedure

The experiments are protocol analysis(Newell and Simon, 1972; Ericsson and Simon, 1984) experiments. In the protocol analysis style, subjects are asked to explain their thought processes as they participate in the experiment. In our experiments, subjects are asked to verbally describe their actions and explain why they are performing those actions as they participate in the experiment. Subjects speak into a headphone-mounted microphone which records their verbal reports on cassette tapes. The subjects are a mix of paid and unpaid adult volunteers with varying experience playing 3D games. The subjects' success or failure in the task is also recorded. For some tasks, the time subjects take is recorded.

Subjects are not aware of the layout of the suite of rooms when they begin a task. Each subject works on two tasks, one with the floor plan from Figure 3 and one with the floor plan from Figure 4.

We are currently using the following tasks:

1. Counting stationary objects: The subject must count the number of glasses in the suite of four rooms. The glasses are all perceptually indistinguishable. There are two variations of this experiment. In variation one, the four rooms look different. In the second variation, two of the rooms are perceptually identical, and the other two rooms are also perceptually identical. The subjects are timed and end the experiment when they believe they know the correct number of glasses.
2. Counting mobile objects: The subject must count the number of robots in the same small suite of rooms as the first variation of task one. The robots move randomly and can change rooms. The robots move at a constant rate of approximately half the maximum possible speed of the subject. There are two variations to this experiment. In the first variation, all robots are perceptually indistinguishable. In the second variation, there are two groups of robots; members of the same group are perceptually indistinguishable from one another. The subjects are timed and end the experiment when they believe they know the correct number of robots.
3. Following a robot: The subject is to follow a robot tour guide through the larger suite of rooms. There are several distractor robots wandering in the suite. The distractors are perceptually indistinguishable from the robot that the subject is following. The experimenter ends the experiment when either the subject has followed the robot through its compete path, or the subject starts following one of the distractor robots. Figure 2 shows a screenshot of this task; in the screenshot, a distractor robot has wandered near the robot tour guide.

4. Following a person: The subject is to follow a person who is the tour guide through the larger suite of rooms. There are several distractor people in the suite. The distractors are perceptually distinguishable from the person the subject is following. The experimenter ends the experiment when, either the subject has followed the person through his compete path, or the subject starts following one of the distractor people. Since people usually have a unique appearance, we hypothesize that our subjects will behave differently than in the "Following a robot" task described above.

Preliminary results

In this section we will describe some preliminary results from our human subjects experiments. Sixteen subjects have participated in the experiment so far.

Obviously with so few subjects we cannot yet do very much quantitative analysis. However, there are two trends emerging that have been surprising. We predicted that counting glasses would take less time than counting robots, since counting unmoving glasses seemed like a task that people do more easily than counting moving robots. We've had six subjects in the two variations of the glass counting experiment so far and ten subjects in the two variations of robot counting experiment. The glass counters take on average a minute more than the robot counters. The robot counters take an average of two minutes and 56 seconds to finish the task, the glass counters take an average of three minutes 52 seconds.

We expected the robot counting task to be the most difficult for subjects. The subjects have to (at least tentatively) identify all of the perceptually indistinguishable robots in order to accurately count them. However, so far the robot following task has been the most difficult. Of the 13 subjects tested, only 54% have successfully followed the robot to the end of its entire path. In contrast, five of the seven (71%) subjects in the robot counting task have successfully counted all of the robots.

We use the protocol data collected from these experiments to get insight into what strategies people use to identify perceptually indistinguishable objects. The subjects have already used most of the strategies that we hypothesized were useful.

Since subjects do not know the layout of the suite of rooms, they begin the task by familiarizing themselves with it. In all of the counting tasks, the tasks for which the subjects defined the end time of the experiment, subjects entered each room at least twice.

Subjects often used the location of an object to help them identify the object. Subjects used the location of the glasses almost exclusively when counting glasses.

When counting moving robots, subjects reported using (and appeared from their actions to use) the location of the robots, the robots' observed speed, and the time since the subject last saw a perceptually indistinguishable robot. Subjects report noticing that they can move more quickly than the robots. Subjects try to move fast enough to make a complete final circuit of the rooms before the robots in the room they start from have the chance to move to another room. This almost certainly accounts for the robot counting tasks taking less time than than the glass counting tasks, where the subjects feel no such pressure to move quickly.

Subjects in the robot following task use all of the strategies that the subjects in the counting tasks did. They also use two that we did not predict. When they loose track of the robot that they are supposed to be following (the "focus robot"), some subjects resort to a random guess. Subjects who used this "strategy" account for most of the those who fail to successfully complete this task.

Most of the subjects who succeeded in the robot following task used some sort of plan recognition while following the focus robot. Most of the subjects started trying to predict where the robot would go next so that they would be ready for the its next course change and not lose it. At least one subject used the fact that the focus robot moved "with a purpose" while distractors moved randomly, to identify the focus robot after losing sight of it. Other subjects reported using the focus robot's speed and trajectory to identify it after losing sight of it when following the focus robot into a room with several distractor robots.

Summary

We have described the problem of identifying perceptually indistinguishable objects. Perceptually indistinguishable objects must be identified using reasoning and knowledge since sensory information cannot help. People can sometimes identify perceptually indistinguishable objects effortlessly. We are currently running experiments with human subjects to find out what strategies people use to identify perceptually indistinguishable objects and how well they can do this task. We have discussed some preliminary results from our experiments. People use location, time, object mobility, plan recognition, and even random guessing to identify perceptually indistinguishable objects. We are designing a simulated robot with the ability to identify perceptually indistinguishable objects. The robot will use the strategies that our experiments show that people use to identify perceptually indistinguishable objects.

References

Coradeschi, S. and Saffiotti, A. (2001). Forward. In Coradeschi, S. and Saffiotti, A., editors, *Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems: Papers from the 2001 AAAI Fall Symposium, Technical Report FS-01-01*, page viii, Menlo Park CA. AAAI Press.

Ericsson, K. A. and Simon, H. A. (1984). *Protocal Analysis.* MIT Press.

Henry Hexmoor, Johan Lammens, S. C. S. (1993). Embodiment in GLAIR: a grounded layered architecture with integrated reasoning for autonomous agents. In Dankel, D. D. and Stewman, J., editors, *Proceedings of The Sixth Florida AI Research Symposium (FLAIRS 93)*, pages 325–329. Florida AI Research Society.

Hexmoor, H. and Shapiro, S. C. (1997). Integrating skill and knowledge in expert agents. In P. J. Feltovic, K. M. Ford, R. R. H., editor, *Expertise in Context*, pages 383–404, Cambridge, MA. AAAI/MIT Press.

id Software. http://www.idsoftware.com/.

Jorrit Tyberghein, Andrew Zabolotny, E. S. e. a. (2002). *Crystal Space: Open Source 3D Engine Documentation.* http://crystal.sourceforge.net, .92 edition. ftp://crystal.sourceforge.net/pub/crystal/docs/download/.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving.* Prentice-Hall.

Pollock, J. (1974). *Knowledge and Justification.* Princeton University Press.

Shapiro, S. C. (1989). The CASSIE projects: An approach to natural language competence. In *4th Portugese Conference on Artificial Intelligence*, pages 362–380. Springer-Verlag.

Shapiro, S. C. (1993). Belief spaces as sets of propositions. *Journal of Experimental and Theoretical Artificial Intelligence (JETAI)*, 5(2 and 3):225–235.

Shapiro, S. C. (1998). Embodied Cassie. In *Cognitive Robotics: Papers from the 1998 AAAI Fall Symposium*, pages 136–143, Menlo Park, CA. AAAI, AAAI Press.

Shapiro, S. C. (2000). SNePS: A logic for natural language understanding and commonsense reasoning. In Iwanska, L. and Shapiro, S. C., editors, *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, pages 175–195. AAAI Press/MIT Press, Menlo Park CA.

Shapiro, S. C. and Ismail, H. O. (2001). Symbol-anchoring in Cassie. In Coradeschi, S. and Saffioti, A., editors, *Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems: Papers from the 2001 AAAI Fall Symposium, Technical Report FS-01-01*, pages 2–8. AAAI Press.

Shapiro, S. C. and Rapaport, W. J. (1992). The SNePS family. *Computers and Mathematics with Applications*, 23(2-5):243–275.

Shapiro, S. C. and the SNePS Implementation Group (1999). *SNePS 2.5 User's Manual.* Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo NY.