Cognitive Science 29 (2005) 103–134

Perception as Abduction: Turning Sensor Data Into Meaningful Representation

Murray Shanahan

Department of Electrical and Electronic Engineering,Imperial College, London, England

Received 26 September 2003; received in revised form 7 May 2004; accepted 21 July 2004

Abstract

This article presents a formal theory of robot perception as a form of abduction. The theory pins down the process whereby low-level sensor data is transformed into a symbolic representation of the external world, drawing together aspects such as incompleteness, top-down information flow, active perception, attention, and sensor fusion in a unifying framework. In addition, a number of themes are identified that are common to both the engineer concerned with developing a rigorous theory of perception, such as the one on offer here, and the philosopher of mind who is exercised by questions relating to mental representation and intentionality.

Keywords: Perception; Abduction; Robotics; Vision; Knowledge representation; Logic

1. Introduction

Philosophers who subscribe to a computational theory of mind and artificial intelligence (AI) researchers who work in the "classical" symbolic style are both vulnerable to awkward questions about how their symbols and representations acquire meaning. The issue dates back at least as far as Franz Brentano, whose concern was how mental states gain their *intentionality,* their "directedness" toward things (Brentano, 1874/1995). Related philosophical specters have surfaced more recently in the form of the Chinese Room argument (Searle, 1980) and the symbol grounding problem (Harnad, 1990).

A typical way to parry such arguments is by appeal to the notion of embodiment (Chrisley & Ziemke, 2003). A system that only interacts with the outside world via a keyboard and screen is said to be *disembodied.* The champions of embodiment are willing to concede that the symbols deployed in such a system lack true intentionality, that they inherit whatever meaning they have from the human designer. By contrast, a robot that enjoys genuine sensorimotor interaction with the same physical world as ourselves is said to be *embodied.* Symbolic representa-

Requests for reprints should be sent to Murray Shanahan, Department of Electrical and Electronic Engineering, Imperial College, Exhibition Road, London SW7 2BT, England. E-mail: m.shanahan@imperial.ac.uk

tions built up by a robot through sensorimotor interaction with the physical world are *grounded* and can be properly said to have intentionality.[1]

This is a promising line of argument, as far as it goes. But it opens a great many questions. In particular, it needs to be filled out with a theoretically respectable account of the process involved in constructing the symbolic representations in question (Barsalou, 1999).[2] This article attempts to plug this gap by offering a formal abductive account of the means by which low-level sensor data is transformed into meaningful representation. The account incorporates detailed treatments of a variety of issues, including incomplete and uncertain sensor data, active perception, and top-down information flow, and it briefly addresses a number of other topics, including attention and sensor fusion.

The methodological context here is robotics, and one of the strengths of the theory on offer is that it has been implemented on a number of robot platforms. No explicit claim is made about the extent to which it applies in the human case. However, at the close of this article various philosophical questions relating to meaning and intentionality are placed, for comparison, alongside answers an engineer might give if the same questions had been asked about a robot built according to the theory of perception being proposed.

Although this article draws heavily on work carried out in the area of logic-based AI, its primary focus is not the challenge of building programs that use explicitly logical forms of representation and inference. Rather, it is to supply an account of robot perception that is couched in a precise theoretical language and that is capable of engaging with a variety of contemporary issues. Naturally the theory *will* appeal to strictly logicist AI researchers, and the article does present work in which logic is used as an explicit representational medium. However, the account also serves as a theoretical yardstick for roboticists working with symbolic AI, but outside the logicist tradition.[3] In addition, it can be viewed as supportive of philosophers of mind who favor a computationalist theory and who are open to the charge that they have failed to supply such an account (Barsalou, 1999, p. 580).[4]

This article is structured into three main parts. In the first part, the abductive theory of perception is outlined, and some of its methodological underpinnings are discussed along with such topics as incompleteness, top-down information flow, active perception, and sensor fusion. In the second part, a number of experiments are sketched in which the abductive theory of perception is put to the test with several real robots ranging from simple mobile platforms to an upper-torso humanoid with stereo vision. These experiments supply the context for a series of worked examples that are presented in more technical detail. In the third part, this article presents a few philosophical reflections on such subjects as the representational theory of mind and intentionality.

2. An abductive account of perception

We begin with a very basic logical characterization of perception as an *abductive* task (Shanahan, 1996b). Given a stream of low-level sensor data, represented by the conjunction Γ of a set of observation sentences, the task of perception is to find one or more explanations of Γ in the form of a consistent logical description Δ of the locations and shapes of hypothesized objects, such that,

$$\Sigma \wedge \Delta \models \Gamma$$

where Σ is a background theory describing how the robot's interactions with the world impact on its sensors. The form of Δ, that is to say, the terms in which an explanation must be couched, is prescribed by the domain. Typically there are many Δs that explain a given Γ, so a preference relation is introduced to order them. When deployed on a robot, a perceptual system conforming to this theoretical description is a perpetual process, embedded in a closed control loop that outputs a constant stream of (sets of) Δs in response to a constant stream of incoming Γs from the robot's sensors, which could be anything from bump switches to a stereo camera.

Observation sentences in Γ will not, in general, describe sensor data in its raw state. The sensor data in Γ is only "low-level" in the following sense. Whatever preprocessing the raw data has undergone to produce it, Γ is never taken as a description of the external world, but rather as a more refined description of the sensor data. For example, Γ might be a description of edges extracted from an image using the Sobel operator, or it might be a sequence of sonar data points that has been through a Kalman filter. Even if the preprocessing applied to the raw sensor data generates information that more obviously purports to describe the world, such as the depth map produced by a stereo-matching algorithm, the abductive theory of perception treats this information simply as a rich form of low-level data. It is strictly the responsibility of the abductive process to decide how to interpret each item of low-level data—whether it should be accepted or rejected, and what it says about the presence, identity, and configuration of objects in the external world.

The real utility of any implementation of perception as abduction will be dictated by the form and content of the background theory Σ—the underlying formalism in which it is expressed, the ontology manifest in its choice of sorts and predicates, and the causal knowledge captured by its constituent formulas. Most of the experiments carried out to date have deployed an underlying formalism capable of representing action, change, and causal relations, and with a rudimentary capability for representing space and shape. As far as content goes, in general Σ has two parts—a set of formulas describing the effect of the robot's actions on the world, and a set of formulas describing the impact of changes in the world on the robot's sensors. In an ideal implementation of the theory, Σ would incorporate a full axiomatization of naive physics (Hayes, 1985), either hand coded or acquired by a compatible learning method, such as inductive logic programming (Muggleton, 1991).

2.1. Incompleteness and uncertainty

Shortly, we will see how the previously mentioned basic abductive framework can be extended to accommodate active perception, attention, and top-down expectation. But even this simple formal description is sufficient to demonstrate several methodological points. In particular, we can lay to rest the false notion that a representational approach to perception assumes the need to build and maintain a complete and accurate model of the environment (Brooks, 1991).

Vision, according to David Marr, is "the process of discovering from images what is present in the world and where it is" (Marr, 1982, p. 3). In the past, researchers have taken this to mean the construction of "a complete and accurate representation of the scene," and to live up to this definition have been forced to make gross idealizations about the environment. For example, an introductory chapter on constraint-based vision from a 1980s AI textbook invites the reader to consider a world of "crack-free polyhedra with lighting arranged to eliminate all shadows" (Winston, 1984, p. 47). This is patently an unsatisfactory starting point if we want to build robots that work in realistic environments. Moreover, it sidesteps two issues central to understanding intelligence, namely incompleteness and uncertainty, which are unavoidable features of what John McCarthy calls the "common sense informatic situation" (McCarthy, 1989).

Uncertainty arises because of noise in a robot's sensors and motors, and it is inevitable because of the physical characteristics of any electrical device. Incompleteness, on the other hand, arises for two reasons, both of which run deeper than the physics. First, incompleteness results from having a limited window on the world. No robot can ever sense more than a small portion of its environment because of the inherently limited reach of any sensor and because of its singular spatial location. Second, incompleteness is an inevitable (and welcome) consequence of the individuality of each robot's goals and abilities. So there is no need for a robot's internal representations to attempt to capture every facet of the external world.

In the same way, an animal's perceptual system is adapted to survival within its particular ecological niche. The subjective visual world of a bee—its *Umwelt,* in the terminology of biologist Jakob von Uexküll—might be thought of as an otherwise meaningless blur in which pollen-bearing flowers stand out like beacons (von Uexküll, 1957). In a similar vein, the *Umwelt* of a robot designed to fetch coke cans might be characterized as a meaningless chaos punctuated by can-shaped objects. So, although it may be a methodological error to presuppose a world of crack-free polyhedra, it still makes sense to build a robot whose perceptual system is tailored to pick out nearly polyhedral objects (such as books), so long as it can do so in a normal, messy working environment.

Mathematical logic is ideally suited to expressing each of the previously mentioned forms of incompleteness and uncertainty, but at the same time possesses a precise semantics. The trick, in the context of the abductive treatment of perception, is to carefully design the permitted form of Δ to take advantage of this. To illustrate, consider the following formula, which constitutes a partial description of an object in the robot's work space.

$$\begin{aligned}&\exists\, c,x,w,v[\mathit{Boundary}(c,x)\wedge \mathit{FromTo}(c,w,v)\wedge\\&\quad \mathit{Distance}(w,\langle -5,5\rangle)<1\wedge \mathit{Distance}(v,\langle 5,5\rangle)<1]\end{aligned}$$

The formula *Boundary*(*c,x*) represents that line *c* is part of the boundary of object *x,* the formula *FromTo*(*c,w,v*) represents that line *c* stretches from point *w* to point *v,* and the term *Distance*(*v,w*) denotes the distance between points *v* and *w.* Both incompleteness and uncertainty are expressed in this formula, thanks to the use of existential quantification. Incompleteness is expressed because the formula only talks about part of the boundary of a single object. It says nothing about what other objects might or might not exist. Uncertainty is expressed because the endpoints of the boundary are only specified within a certain range. Moreover, no commitment to an absolute co-ordinate system is inherent in such a formula, so the locations of these endpoints could be taken as relative to the robot's own viewpoint.

Of course, this formula is purely illustrative, and the design of a good ontology for representing and reasoning about shape and space—perhaps one that is more qualitative and less numerical—is an open research issue. For comparison, here is an example of the kind of formula generated by the implemented abductive system for visual perception, which is described in Section 3.

$$\exists\, x,a,s,r\ [Solid(x,a) \wedge Shape(x,Cuboid) \wedge FaceOf(r,x) \wedge SideOf(E_0,r)]$$

The formula $Solid(x,a)$ represents that a solid object x is presenting aspect a to the viewer, the formula $Shape(x,s)$ represents that the solid object x has shape s, the formula $FaceOf(r,x)$ represents that the two-dimensional surface r is one face of object x, and the formula $SideOf(e,r)$ represents that visible edge e is on the boundary of surface r. The term E_0 denotes a particular visible edge. This is a purely qualitative representation of the simple fact that a certain kind of object is visible and "out there" somewhere.

Even this qualitative representation makes some commitment to the nature of the object in view, through the *Shape* predicate. Yet, as Pylyshyn argued, it is sometimes necessary to provide "a direct (preconceptual, unmediated) connection between elements of a visual representation and certain elements in the world [that] allows entities to be referred to without being categorized or conceptualized" (Pylyshyn, 2001, p. 127). Once again, mathematical logic facilitates this, simply by allowing an existentially quantified variable to be substituted for the term *Cuboid*.

The open-ended versatility of mathematical logic also enables the abductive account of perception to accommodate the idea that to perceive the environment is to perceive what it can afford, "what it provides or furnishes, either for good or ill" (Gibson, 1979, p. 127).[5] Indeed, the representations generated by an abductively characterized perceptual process are only there as intermediate structures to help determine the most appropriate action in any given situation. Their use is only justified to the extent that they enhance rather than hinder the robot's ability to respond rapidly and intelligently to ongoing events.

Respecting this point is, once again, down to the design of the right ontology for the background theory Σ. On the one hand, the design of the background theory Σ might reflect the assumptions of a classical approach to machine vision and describe the relation between completely described three-dimensional objects and the edges and features they give rise to in the visual field. But, on the other hand, it can be designed to reflect the concerns of a more biologically inspired approach to perception and capture the relation between, say, graspable containers and a small number of easily detected visual cues for such objects. In itself, the theoretical framework is entirely neutral with respect to this sort of choice.

2.2. Top-down information flow

Cognitive psychology has long recognized that the flow of information between perception and cognition is bidirectional in humans (Cavanagh, 1999; Peterson, 2003), and the presence of many reentrant neural pathways in the visual cortex reinforces this view (Grossberg, 1999; Ullman, 1996, chap.10). Similarly, in the field of machine vision, although great emphasis was placed on top-down processing in the 1970s (e.g., Waltz, 1975) and on bottom-up techniques in the 1980s (e.g., Marr, 1982), the need for a two-way flow of information has always been acknowledged. Gestalt effects illustrate the point dramatically, as when a participant is presented with a picture that looks like a collection of random spots and blotches until it is pointed out verbally that it depicts, say, a cow, whereupon the spots and blotches resolve themselves, as if by magic, into indisputably animal form (Ullman, 1996, p. 237). In such cases, where linguistically acquired knowledge causes the participant to perceive objects and boundaries where none were seen before, it is clear that high-level cognitive processes have influenced low-level perception.[6]

In the abductive theory of perception on offer here, top-down information flow is accommodated by making use of *expectation.* As with all the proposals advanced here, the aim is not necessarily to achieve psychological or biological plausibility, but to produce a well-engineered mechanism. But in this case, as so often, there is a strong engineering rationale for imitating Nature. Presented with a cluttered, poorly lit scene full of objects occluding one another and camouflaged by their surface patterns, the use of expectations derived from past experience can reduce the task of distinguishing the important objects from the rest to computationally manageable proportions. Indeed, it makes engineering sense to permit the influence of high-level cognition to percolate all the way down to low-level image processing routines, such as edge detection.

The way expectation is exploited is as follows: Suppose a set of Δs has been found that could explain a given Γ. Then, in addition to entailing Γ, each Δ will, when conjoined with the background theory Σ, entail a number of other observation sentences that might not have been present in the original sensor data Γ. These are Δ's expectations. Having determined these, each Δ can be confirmed or disconfirmed by checking whether or not its expectations are fulfilled. This might involve passively processing the same low-level sensor data with a different algorithm, or it might involve a more active test to gather new sensor data, perhaps in a different modality. In Section 3, we see how these ideas are applied to vision.

2.3. Active perception

The use of expectation in the way just described can be thought of as a form of *active perception* (Aloimonos, Weiss, & Bandyopadhyay, 1987; Ballard, 1991). The concept of active perception can be widened from its original application in the work of such pioneers as Aloimonos and Ballard to include any form of action that leads to the acquisition of new information via a robot's sensors, ranging from short duration, low-level actions, such as adjusting the threshold of an edge detection routine or rotating a camera to give a different viewpoint, to longer duration, high-level actions, such as climbing to the top of a hill to see what is on the other side. More subtle "knowledge-producing" actions (Scherl & Levesque, 1993), such as looking up a phone number, or asking another agent a question, can be thought of as lying on the same spectrum.

In general, the varieties of active perception can be classified along two axes. Along one axis, we have the kind of informative action involved—low-level or high-level, short duration or long duration—as detailed previously. Along the other axis, we have the mechanism that triggers the informative action in question. In each of the previous examples, the informative action could have been triggered by a reactive condition–action rule whose side effect is the acquisition of knowledge, it could have been performed as a deliberate, reasoned effort to acquire knowledge in the face of ignorance, or it could have been an automatic attempt to confirm an expectation about this situation. In other words, the triggering mechanism could be habit, intention, or expectation.

As well as providing a theoretical framework for handling uncertainty, incompleteness, and expectation, the basic abductive characterization of sensor data assimilation can be extended to encompass all the previously mentioned forms of active perception. Each is an example of an action or series of actions with the side effect of generating new sensor data, that is to say, new Γs. For each type of active perception, the new sensor data can be interpreted via abduction to yield useful knowledge, that is to say, new Δs. The purpose of action here is to control the flow of Γs. However, the role of action in perception is not only to *maximize useful* information, but also to *minimize useless* information. The basic abductive formula makes this clear. If Γ comprises a large volume of data, the problem of finding Δs to explain it will be intractable.

Part of the responsibility for managing incoming sensor data so that Γ is always small but pertinent belongs to the *attention* mechanism (Niebur & Koch, 1998). In most animals, turning the head toward a loud noise and saccading to motion are instinctive behaviors with obvious survival advantages. An animal needs to gather information from those portions of the environment most likely to contain predators, prey, or potential mates. Analogous strategies have to be found for robots, regardless of their design aims.

In general, a full abductive account of perception has to incorporate three components—attention, explanation, and expectation. The flow of information between these components is cyclical. Thanks to the attention mechanism, the robot's sensory apparatus is directed onto the most relevant aspects of the environment. The explanation mechanism turns the resulting raw sensor data into hypotheses about the world. And in response to expectation, ways of corroborating or refuting those hypotheses are found, thereby influencing the attention mechanism.

2.4. Sensor fusion

The abductive account of perception can also be thought of as supplying a theoretical foundation for addressing the problem of *sensor fusion,* that is to say, the problem of combining potentially conflicting data from multiple sources into a single model of some aspect of the world (Clark & Yuille, 1990). The low-level sensor data in Γ can come from different sensors, or from the same sensor but preprocessed in different ways, or from the same sensor but extracted at different times. If the background theory Σ successfully captures how a configuration of objects in the world gives rise to each kind of low-level sensor data, then the basic abductive characterization of perception also takes care of the data fusion issue. Any Δ conforming to the definition of an explanation has implicitly integrated each item of data in Γ, whatever its source.

However, Σ must be carefully designed to ensure this works, especially in the presence of potentially conflicting data. This is facilitated by the use of *noise* and *abnormality* terms. In addition to describing the "normal" causes of incoming sensor data, namely the robot's interaction with objects in the world, Σ can include formulas that can "explain away" unexpected items of sensor data as mere noise, and formulas that can dismiss the absence of expected sensor data as abnormal. The causal laws expressed in Σ will then include noise or abnormality conditions or both, and Δs must be permitted to include corresponding noise and abnormality terms. It then remains for the preference relation over Δs to prioritize explanations with the fewest such terms, thus ensuring that the explanations that best fit the data rise to the top of the pile.

Looked at in this way, the sensor fusion problem is just an instance of the general problem of finding an explanation for a body of sensor data. Indeed, with a rich sensory modality like vision, the same issue arises even with a single snapshot of a scene preprocessed by a single algorithm, such as edge detection. Both phantom and occluded edges can occur, and the abductive account of perception needs to be able to deal with them. This topic is treated in more detail in Section 3.

3. Elaborating and applying the abductive account

This section describes a number of experiments, carried out over a period of several years, that were designed to evaluate the abductive account of perception using real robots. In each experiment, the basic abductive characterization is elaborated in a different way. So a series of detailed worked examples is also presented to illustrate how the theoretical framework is applied in each experimental context.

3.1. The Rug Warrior experiment

The earliest and simplest of these experiments (Shanahan, 1996a, 1996b) used a low-cost two-wheeled mobile robot (called a "Rug Warrior"), whose only sensors were bump switches that detected collisions with walls and obstacles (Jones & Flynn, 1993). The robot used a reactive exploration strategy, but concurrently ran a sensor data-assimilation process, which, as a side effect of the robot's exploratory movements, built a map of the environment.[7] Although the algorithm for sensor data assimilation was implemented as a straightforward C program, without any form of explicit logical representation, the behavior of the algorithm provably conformed to a precise abductive specification expressed in terms of abduction and predicate logic.

Because the only way for a sensor event to occur with this robot was as a result of its own motion, the map-building process can be viewed as a form of active perception, in the broad sense defined previously. To encompass this form of active perception, the basic abductive account needs to be augmented in such a way that the background theory Σ describes the *effects of actions* and must be conjoined to a description of the *narrative* of robot actions that gave rise to the sensor data Γ. In the work cited, action and change are described using the *event calculus,* a well-known formalism for reasoning about action that incorporates a robust solution to the frame problem (Shanahan, 1997b, 1999).

To get a fuller flavor of how this sort of active perception works, let us make the example of the simple Rug Warrior robot more precise. Suppose we let,

- Γ be a conjunction of observation sentences, each describing an event in which one of the robot's bump switches is tripped,
- Δ_N be a conjunction of sentences, each describing a robot action, such as "turn left," "move forward," and so on, and,
- Σ be a background theory describing both the effect of the robot's actions on the world (specifically the effect of movements on the location of the robot itself) and the impact of

the world on the robot's sensors (specifically how the presence of an obstacle causes bump events).

Then, the task of sensor data assimilation is to find a consistent formula Δ_M describing an arrangement of walls and other obstacles (a map), such that,

$$\Sigma \wedge \Delta_N \wedge \Delta_M \models \Gamma$$

Clearly there will, in general, be many Δ_Ms that conform to this prescription, and a major challenge for any abductive account of perception is to manage multiple explanations in a principled way. This topic is dealt with in more detail toward the end of this section, in the context of vision.

3.2. Reinventing Shakey

In the Rug Warrior experiment, the gap between specification and implementation was considerable. Although the map-building algorithm conformed to a specification expressed in terms of abduction, no explicit abductive reasoning was carried out by the robot, and no explicit logical representations were built. The aim of the next set of experiments, involving miniature Khepera robots (Mondada, Franzi, & Ienne, 1993), was to demonstrate that a practical robot control system could be made in which the basic unit of computation is a proof step, and the basic unit of representation is a sentence of logic (Shanahan, 2000; Shanahan & Witkowski, 2001). In such a system, computation is *transparent,* in the sense that intermediate computational states have declarative meaning—each being a logical sentence in an ongoing derivation—and each step of computation can be justified by the laws of logic.

The attraction of such a foundation from an engineering point of view is clear.[8] As in so many branches of engineering, elegance in design goes hand-in-hand with an elegant theoretical basis, and leads to artifacts that are easy to understand, to modify, and to maintain. However, one of the finest early attempts to use logic and theorem proving in the way described, namely Green's resolution-based planner (Green, 1969), fell foul of computational complexity problems. This lead the Shakey project to compromise the ideal of using logic and theorem proving to control a robot (Nilsson, 1984), and no research group took up the challenge again until the cognitive robotics revival of the mid 1990s (Lespérance et al., 1994). So the question at the top of the agenda was whether a robot control system based on logic is computationally viable.

The Khepera experiments demonstrated this through the use of a logic programming metainterpreter with *anytime* properties (Zilberstein & Russell, 1993), in other words one capable of suspending computation at any time, but still yielding useful partial results (Kowalski, 1995). This capacity facilitates a design that departs from the classical AI approach exemplified by Shakey in a very important way. In an architecture such as Shakey's, logical inference is the driving force behind the selection of actions. By contrast, in the Khepera experiments, logical inference is inserted into a closed-loop control mechanism that only allows it to carry out a small, bounded amount of computation in each cycle before the robot's sensors are consulted again. So the robot can never get stuck in a thinking rut while the world changes around it.

The miniature Khepera mobile robots used in the experiments were each equipped with two drive wheels and a suite of infrared proximity sensors. The robots were placed in a miniature officelike environment comprising walls, rooms, and doorways. The only low-level behaviors the robots were programmed with were wall-following and corner-turning. The first experiment centered on a navigation task (Shanahan, 2000). Given a description of the layout of rooms and doorways, and knowing its initial location, the robot had to find its way to a different specified room. But as the robot proceeded, the experimenter could unexpectedly block any of the doorways to hinder its progress.

Once again, an abductive approach to perception was adopted, although this time true abductive inference was carried out in real time, using the sort of anytime metainterpreter described previously. The job of the abductive sensor data-assimilation process was to explain the ongoing stream of sensor events in terms of what might be going on in the robot's environment. When sensor events conformed to expectations, this was a trivial task. But sometimes anomalous sensor events could occur, such as the detection of an unexpected obstacle. If the only explanation for the anomaly was a blocked doorway, a process of replanning was triggered.

Using this technique, the robot was able to navigate successfully from room to room, without perceptible computational delays, in spite of unexpected and unfavorable changes to its environment. So the experiment showed that, in principle, a robot might be able to deal with a changing world using the technology of classical AI adapted for closed-loop control, albeit in a simple environment.

The second experiment with Khepera robots was another map-building task (Shanahan & Witkowski, 2001). Once again, the crucial difference from the earlier Rug Warrior experiment was that actual abductive inference was carried out in real time. The abductive characterization of the map-building task was similar to that for the Rug Warrior. But for the Khepera experiment, a more sophisticated map-building strategy was deployed, in which the robot reasoned about its own ignorance and carried out a planned sequence of exploratory actions to fill in gaps in its knowledge. To have accurate knowledge of its own ignorance, the robot had to be able to recognize novel locations and to re-identify locations it had previously visited, which is also an important aspect of the "symbol anchoring problem" (Coradeschi & Saffiotti, 2000).

3.3. A simple worked example of active perception as abduction

Based on the navigation task in the Khepera experiments, this section presents a worked example of how a simple form of active perception can be formalized as an abductive process. We begin with a short introduction to the event calculus, the formalism for reasoning about action deployed in all the work reported in this article. The version of the event calculus used here has been widely adopted and has become absorbed into the AI canon (Russell & Norvig, 2002, chap. 10). In addition to robotics, it has found application in such areas as natural language processing (Lévy & Quantz, 1998), intelligent agents (Jung, 1998), and commonsense reasoning (Morgenstern, 2001; Mueller, 2004; Shanahan, 2004). A more detailed presentation can be found in Shanahan (1999).

The ontology of the event calculus includes events (or actions), time points, and fluents. A *fluent* can be anything whose value changes over time, such as the location of a robot. The basic language of the calculus is presented in Table 1.

Table 1
The language of the event calculus

Formula	Meaning
Initiates(e,f,t)	Fluent f starts to hold after action e occurs at time t
Terminates(e,f,t)	Fluent f ceases to hold after action e occurs at time t
Initially(f)	Fluent f holds from time 0
Happens(e,t)	Action or event e occurs at time t
HoldsAt(f,t)	Fluent f holds at time t
Clipped(t^1,f,t^2)	Fluent f is terminated between times t^1 and t^2

The relation between these predicates is governed by the following three axioms, whose conjunction will be denoted EC.

$$HoldsAt(f,t) \leftarrow Initially(f) \wedge \neg\, Clipped(0,f,t) \tag{EC1}$$

$$HoldsAt(f,t_2) \leftarrow \tag{EC2}$$

$$Happens(e,t_1) \wedge Initiates(e,f,t_1) \wedge t_1 < t_2 \wedge \neg\, Clipped(t_1,f,t_2)$$

$$Clipped(t_1,f,t_3) \leftrightarrow \tag{EC3}$$

$$\exists\, e,t_2\ [Happens(e,t_2) \wedge t_1 < t_2 < t_3 \wedge Terminates(e,f,t_2)]$$

The frame problem is overcome using circumscription (Lifschitz, 1994; McCarthy, 1986). The predicates *Happens, Initiates,* and *Terminates* are minimized, corresponding to the default assumption that the known event occurrences are the only event occurrences, and the known effects of actions are the only effects of actions. By circumscribing separate parts of a theory independently, the difficulties that beset early attempts to apply circumscription to the frame problem are avoided (Shanahan, 1997b, 2003). In what follows, given a formula Φ and a set P of predicates, the circumscription of Φ minimizing P will be denoted CIRC[Φ; P].

In practice, the deductive application of these axioms performs prediction, that is to say, reasoning from causes to effects. Given a narrative of events and actions in the form of a set Δ of *Initially* and *Happens* formulas, and a description of the effects of actions in the form of a set Σ of *Initiates* and *Terminates* formulas, the consequences of,

$$\mathrm{CIRC}[\Sigma;\ Initiates,\ Terminates] \wedge \mathrm{CIRC}[\Delta;\ Happens] \wedge \mathrm{EC}$$

will include a set of *HoldsAt* formulas describing which fluents hold when, in other words capturing the effects of the events and actions in Δ.

However, the concern of this article is not with prediction, but with explanation, that is to say, reasoning from effects to causes, where the effects are the raw sensor data and the causes are features of the world external to the robot. In event calculus terms, given,

- a set Γ of *HoldsAt* formulas describing sensor data,
- a set Δ_N of *Initially* and *Happens* formulas describing the robot's initial state and subsequent actions, and

- a description of the effects of actions in the form of a set Σ of *Initiates* and *Terminates* formulas,

the basic abductive task is to find a set Δ_M of formulas describing features of the world such that,

$$\Delta_M \wedge \text{CIRC}[\Sigma; \mathit{Initiates}, \mathit{Terminates}] \wedge \text{CIRC}[\Delta_N ; \mathit{Happens}] \wedge \text{EC} \models \Gamma.$$

Many variations on this basic formulation are possible. For instance, in the example to follow, Γ includes *Happens* formulas describing sensor events.

Now, consider Figure 1. Suppose the robot possesses a topological map and knows the layout of rooms and doorways in its world, but it has no way to keep track of whether doors are open or shut. Because it also has no knowledge of the distances between the various corners, as it proceeds along the wall, the robot cannot possibly know that D_1 is in fact shut until it encounters corner C_4.

Let the robot's knowledge of the corners and doorways be captured by the following set of formulas, whose conjunction will be denoted Δ_{M1}. The predicate names should be self-explanatory.

$\mathit{NextCorner}(C_1,C_2)$	$\mathit{NextCorner}(C_2,C_3)$
$\mathit{NextCorner}(C_3,C_4)$	$\mathit{Inner}(C_1)$
$\mathit{Outer}(C_2)$	$\mathit{Outer}(C_3)$
$\mathit{Inner}(C_4)$	$\mathit{Door}(D_1,C_2,C_3)$

Next, we need a set of formulas describing the effects of the robot's actions. The only action we will consider here is *FollowWall,* which takes the robot to the next detectable corner. In this example, that corner will be C_2 or C_4, depending on whether door D_1 is open or shut.

$$\mathit{Initiates}(\mathit{FollowWall},\mathit{AtCorner}(c_2),t) \leftarrow$$
$$\quad \mathit{HoldsAt}(\mathit{AtCorner}(c_1),t) \wedge \mathit{NextVisibleCorner}(c_1,c_2)$$
$$\mathit{NextVisibleCorner}(c_1,c_2) \leftarrow$$
$$\quad \mathit{NextCorner}(c_1,c_2) \wedge \neg\, \mathit{InvisibleCorner}(c_2)$$
$$\mathit{NextVisibleCorner}(c_1,c_3) \leftarrow$$
$$\quad \mathit{NextCorner}(c_1,c_2) \wedge \mathit{InvisibleCorner}(c_2) \wedge \mathit{NextVisibleCorner}(c_2,c_3)$$

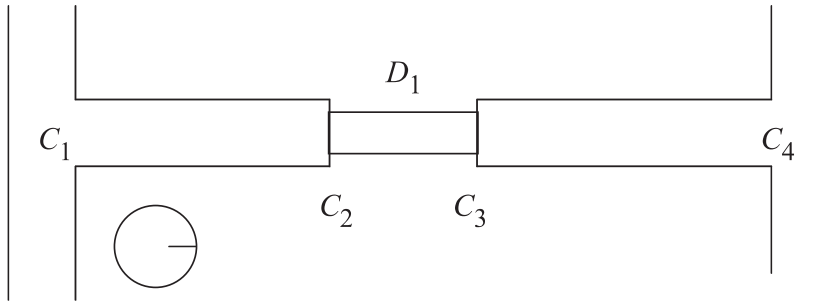

Fig. 1. A Simple Mobile Robot Scenario.

$$InvisibleCorner(c_1) \leftrightarrow \exists\, d,c_2\, [Door(d,c_1,c_2) \wedge DoorShut(d)]$$

$$DoorOpen(d) \leftrightarrow \neg\, DoorShut(d)$$

Let Σ be the conjunction of the previous formulas. Now we require a set of formulas that captures the causes of robot sensor events. In this example, the only sensor events we consider are *GoesHigh*(*Front*), which occurs when the robot's front infrared proximity sensors detect an obstacle ahead, and *GoesLow*(*Left*), which occurs when the robot's sensors encounter a gap to its left.

$$Happens(GoesHigh(Front),t) \leftarrow$$

$$Happens(FollowWall,t) \wedge$$

$$Initiates(FollowWall,AtCorner(c),t) \wedge Inner(c)$$

$$Happens(GoesLow(Left),t) \leftarrow$$

$$Happens(FollowWall,t) \wedge$$

$$Initiates(FollowWall,AtCorner(c),t) \wedge Outer(c)$$

Let Δ_N be the conjunction of the previously *Happens* formulas with the following narrative formulas, which describe the relevant episode in the robot's history.

$$Initially(AtCorner(C_1))$$

$$Happens(FollowWall,T_1)$$

The sensor data to be explained involves a single event.

$$Happens(GoesHigh(Front),T_2)$$

$$T_2 > T_1$$

Now, if we let Γ be the conjunction of the previously mentioned two formulas, the abductive task is to find some Δ_{M2} composed of only *DoorOpen* and *DoorShut* formulas, such that,

$$\Delta_M \wedge \mathrm{CIRC}[\Sigma; \mathit{Initiates, Terminates}] \wedge \mathrm{CIRC}[\Delta_N ; \mathit{Happens}] \wedge \mathrm{EC} \vDash \Gamma$$

where Δ_M is $\Delta_{M1} \wedge \Delta_{M2}$. In this case, the obvious solution is simply to let Δ_{M2} be the formula *DoorShut*(*d*). The inclusion of this formula on the left-hand-side of the entailment ensures that the next detectable corner after C_1 is C_4, which in turn explains the occurrence of a *GoesHigh*(*Front*) event after the robot executes its *FollowWall* action.

This example shows the robot actively probing its environment to discover more about it, and it demonstrates the need for inference to interpret the resulting inflow of sensor data. It is just about the simplest such example that can be presented in a few pages, but it illustrates a basic theoretical apparatus that can be used to couch much more complex forms of perception in abductive terms. For example, by making the whole of Δ_M abducible, the general problem of map building is reduced to a kind of active perception characterized in abductive terms. Indeed, this is precisely how map building was treated in the Khepera experiments described previously, and presented in detail in (Shanahan & Witkowski, 2001). A more substantial extension of the basic framework is required to deal with the multiple explanations problem, which is addressed in the next worked example.

3.4. Visual perception through abduction

Although the Rug Warrior and Khepera experiments served as a useful "proof of concept" for an updated attempt to achieve robot control through logic, they left open the question of whether the ideas under test would scale up to robots with richer sensors in more complex environments. To address the issue of scaling up, a much more ambitious project has been launched, involving a bench-mounted, upper-torso humanoid robot with stereoscopic vision, known as Ludwig (Figure 2).

Ludwig has two arms, articulated at the shoulder and elbow, with 3 degree-of-freedom each. The forearm is mechanically constrained to move in the same plane as the workbench. The arms terminate with simple prods, rather than grippers. The stereoscopic camera is mounted on a pan-and-tilt head. Ludwig is mechanically very simple, having fewer degrees of freedom than other recently constructed humanoid robots, such as Cog at MIT (Brooks, Breazeal, Marjanovic, Scassellati, & Williamson, 1999). This cuts down on low-level motor control problems and allows the project to concentrate on the central issues of cognition and perception.

Like a human infant, Ludwig's main occupation is play. The specific challenge for Ludwig is to learn about the objects in its work space through physical interaction with them. The primary function of the visual system is to pick out potential objects of interest—those that make up Ludwig's *Umwelt*—and to direct the camera's gaze on them. Ludwig can then use visual servoing in an attempt to nudge them. The resulting changes in orientation and relative location of these and other objects on the workbench lead to new and useful information about their shapes, sizes, and layout.

The Ludwig project embraces both classical AI technology and ideas from biologically inspired robotics. On the one hand, the order in which research topics are tackled is not

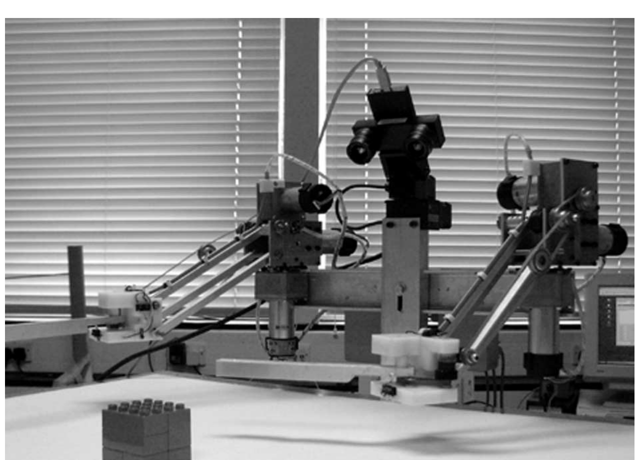

Fig. 2. Ludwig, an upper-torso humanoid.

top-down, as in classical AI, but bottom-up, as in biologically inspired robotics, mirroring ontogeny and phylogeny. Accordingly, rather than viewing human-level intelligence as a disembodied phenomenon and coming at it through language and problem solving, the project considers human-level intelligence to rest on bodily skills, such as nudging and grasping, and to be grounded in (active) perception.

On the other hand, Ludwig's play has an end product that will be familiar to anyone with a classical AI background, namely declarative representations with a denotational semantics—in particular, sentences of logic. (These should not, of course, be confused with sentences of natural language. It is not anticipated that Ludwig will be equipped with a capacity for natural language in the near future.) And as in the Rug Warrior and Khepera experiments, these representations are constructed by an abductive process, which is described here.

To meet the demands of vision—an altogether richer source of sensor data than bump switches or infrared proximity sensors—a much more sophisticated abductive account of perception has been developed (Shanahan, 2002). At the core of this treatment of visual perception are two advances on the basic abductive prescription. First, a precise measure of the *explanatory value* of a hypothesis is defined, based on probability. Crudely put, the key property of this measure is that the more data items a hypothesis explains, the greater its explanatory value. By calculating their explanatory values, competing hypotheses can be compared and unpromising ones discarded.

The precise definition of explanatory value is as follows. Let $\Delta_1 \ldots \Delta_n$ be the set of all hypotheses that explain the sensor data Γ according to the usual abductive definition. Given that one and only one of these hypotheses can be the true explanation, and if none of the hypotheses are subsumed by any other, we can assume $\Delta_1 \oplus \ldots \oplus \Delta_{k-1} \oplus \Delta_{k+1} \oplus \ldots \oplus \Delta_n$.[9] Now, consider any hypothesis Δ_k, and let R be,

$$\Delta_1 \oplus \ldots \oplus \Delta_{k-1} \oplus \Delta_{k+1} \oplus \ldots \oplus \Delta_n.$$

Then, from the laws of probability, we have,

$$P(\Delta_k \mid \Delta_k \oplus R) = \left(1 + \frac{P(R)}{P(\Delta_k)}\right)^{-1} \tag{1}$$

where $P(\Delta_k)$ is the prior probability of Δ_k and $P(R)$ is the prior probability of R. From the form of R, we have,

$$P(R) = \left(\sum_{n}^{i=1} P(\Delta_i)\right) - P(\Delta_k). \tag{2}$$

For any hypothesis Δ of the form $\psi_1 \wedge \ldots \wedge \psi_m$, we have,

$$P(\Delta) = \prod_{m}^{j=1} P(\psi_j) \tag{3}$$

From Equations (1) to (3), the posterior probability of any hypothesis can be calculated, given the set of all hypotheses, and this is taken to be its explanatory value. Now, recall that our measure of explanatory value is supposed to reward a hypothesis for explaining more data. Despite the fact that the abductive definition insists that Γ is explained in its entirety, this criterion can be met by allowing a hypothesis to include *noise terms* (Poole, 1995; Shanahan, 1997a). A noise term "explains away" an item of sensor data as mere noise. If noise terms are assigned a low prior probability, Equation (3) ensures that a hypothesis that properly explains an item of sensor data has a higher explanatory value than one that simply explains it away as noise.

The second advance on the basic abductive story allows for the performance of actions to confirm or disconfirm the *expectations* of competing hypotheses and incorporates the results into an updated measure of explanatory value. A key property of the updated measure is that, other things being equal, the fewer unfulfilled expectations a hypothesis has, the greater its explanatory value.

Overall, the abductive processing of a set of low-level sensor data consists of the following four steps:

1. Using abduction, establish a set of hypotheses that can explain the sensor data.
2. Select the best of these hypotheses according to their explanatory values.
3. Determine the expectations of each selected hypothesis, and reconsult the raw data to check each of these expectations.
4. Recalculate the explanatory values of the selected hypotheses, taking into account the results of Step 3, and reorder the hypotheses accordingly.

To see how these ideas work in the context of vision, consider Figure 3. The central image shows the results of applying a simple edge-detection algorithm, based on the Sobel operator, to the left-hand image. The threshold of the edge-detection algorithm is set high to reduce the volume of spurious data. Now consider the block circled in the upper image. Because of its dark half, only some of this block's edges have been picked out by the algorithm. This leads to two competing initial hypotheses to explain the edge data. According to Hypothesis A, which is incorrect, Ludwig is looking at a short block, the faint edges to the left being noise. According to Hypothesis B, which is correct, Ludwig is looking at a long block, but lots of its edges are missing.

These two hypotheses have different expectations. For example, a consequence of Hypothesis A would be the presence of a vertical edge from the top left-hand corner of the proposed block. A consequence of Hypothesis B is that the topmost edge of the block should extend far-

Fig. 3. An ambiguous block.

ther to the left than it apparently does. Using an edge-checking algorithm that is much more sensitive than the algorithm originally applied to the whole scene, these expectations can be tested. In this example, this leads to a reduction in the explanatory value of Hypothesis A, whose expectations are not fulfilled, and an increase in the explanatory value of Hypothesis B, whose expectations are fulfilled.

3.5. A worked example of top-down information flow

More formally, for the purposes of abduction, low-level image data such as that in Figure 3 is represented by a set of predicate calculus formulas. The low-level data in question could be the product of any off-the-shelf image processing technique, such as color-based segmentation, optical flow, stereo matching, or any combination of these. But for the worked example of this section we stick to edge detection. The edge data in Figure 3 can be represented by a set of formulas that includes the following exemplars:

$Line(1,[238,157],[241,147])$

$Line(3,[249,157],[253,145])$

The formula $Line(w,p_1,p_1)$ represents that there is an edge in the image, labeled w, from point p_1 to point p_2. (The labels are arbitrarily assigned by the low-level edge detection procedure.)

An abductive process can be structured into multiple layers (Josephson & Josephson, 1994; Shanahan, 2002). In effect, each explanation Δ at the nth layer becomes data Γ to be explained at the $n + 1$th layer. Accordingly, formulas that are abducible at layer n can become observable at layer $n + 1$. In this context, it is convenient to adopt a two-layer abductive process. The task of the first layer is to explain one-dimensional edges in terms of two-dimensional visible regions. The task of the second layer is to explain the visible regions in terms of spatially located three-dimensional objects.

Each layer has a separate background theory Σ, which includes three groups of logical formulas. Although every formula enjoys the same status from a declarative point of view, each group plays a different role in the abductive process, and this will be reflected in any implementation. First, we have a set of formulas that can be used to *generate* explanatory hypotheses. Second, we have a set of formulas describing *constraints* on the permissible hypotheses. Third, we have a set of formulas describing the *expectations* of each hypothesis.

Here are two sample first-layer formulas belonging to the first of these three groups. Their style is typical. Each has the form $\Psi \leftarrow \Phi$, where Ψ is a description of sensor data and Φ is a description of conditions under which that sensor data arises.

$$\exists p_1, p_2 \left[Line\left(w, p_1, p_2\right) \right] \leftarrow \exists\, r \left[\mathrm{Re}\,gion\left(r\right) \wedge SideOf\left(w, r\right) \right]$$
$$\exists p_1, p_2 \left[Line\left(w, p_1, p_2\right) \right] \leftarrow Noise\left(w\right)$$

In essence, these two formulas state that a visible edge might be the side of a region or it might simply be noise. The three predicates *Region, SideOf,* and *Noise* (whose meanings should be obvious) are declared to be abducible, so hypotheses can be expressed in their terms. Next, we have a sample first-layer formula from the second group:

$$\left[\mathit{SideOf}\left(w_1,r\right)\wedge \mathit{SideOf}\left(w_2,r\right)\wedge w_1\ w_2\right]\rightarrow$$
$$\left[\mathit{Parallel}\left(w_1,w_2\right)\vee \mathit{Joins}\left(w_1,w_2\right)\right]$$

Again, the style of the formula is typical. It has the form $\Psi \leftarrow \Phi$, where Φ describes conditions that must be met by any hypothesis conforming to Ψ. In this case, the formula presents a constraint that must be met by a hypothesis describing trapezoidal regions, on the (artificial) assumption that the robot's *Umwelt* comprises cuboids whose visual aspect consists of such regions.

We postpone consideration of the third type of formula—one describing the expectations of hypotheses—until the second layer of the abductive process. In the meantime, it can be seen that the first-layer explanations of the low-level image data, according to the formulas previously discussed, include both the hypothesis that there is a region whose sides include Edges 1 and 3 (where Edge 7 is noise), and the competing hypothesis that there is a region whose sides include Edges 7 and 3 (where Edge 1 is noise). Let us call these, respectively, Hypothesis A* and Hypothesis B*. Hypothesis A* includes the following formulas:

$\mathit{Region}(R_0)$ $\mathit{SideOf}(1,R_0)$

$\mathit{SideOf}(3,R_0)$ $\mathit{Noise}(7)$

Hypothesis B* includes the following formulas:

$\mathit{Region}(R_0)$ $\mathit{SideOf}(7,R_0)$

$\mathit{SideOf}(3,R_0)$ $\mathit{Noise}(1)$

According to Equations (1) to (3), these two hypotheses have equal explanatory value, because both postulate a single new region, and both contain the same number of noise terms. Note, however, that hypotheses containing more noise terms will have lower explanatory value. In other words, the best hypotheses are those that postulate only a few regions, each of which explains the presence of many edges.

Now let us consider the second layer of abduction, which moves from visible regions to solid objects. The second-layer formulas for generating hypotheses include the following:

$$\mathit{Region}(r) \leftarrow \exists\, x,a\ [\mathit{Solid}(x,a) \wedge \mathit{IsAspect}(a) \wedge \mathit{FaceOf}(r,x)]$$

$$\mathit{IsAspect}(\mathit{3Face})$$

The predicate $\mathit{Solid}(x,a)$ represents that a solid object x is in the visual field, presenting aspect a. The predicate $\mathit{FaceOf}(r,x)$ represents that the visible region r is a facet of object x that is turned toward the viewer. An *aspect* of an object is an equivalence class of possible views of that object that are invariant with respect to certain topological properties, such as the number of its corners or faces (Figure 4). A cube, for example, has three aspects. The one depicted in Figure 4 is denoted *3Face.*

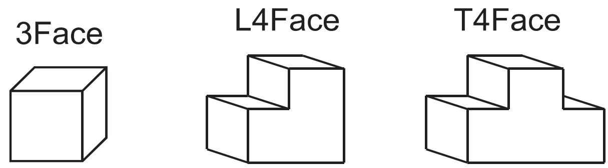

Fig. 4. Aspects of a cube, an L block, and a T-block.

The following formula is an example of a second-layer constraint, which states that every pair of faces in the aspect *3Face* is abutting, and rules out any hypothesis for which this is not the case.

$$[Solid(x,3Face) \wedge FaceOf(r_1,x) \wedge FaceOf(r_2,x) \wedge r_1 r_2] \rightarrow Abuts(r_1,r_2)$$

Now let us consider two competing second-layer hypotheses. Suppose Hypothesis A contains all of Hypothesis A* and also includes the following formulas:

$Region(R_1)$ $SideOf(4,R_1)$

$SideOf(6,R_1)$ $Solid(X_0,3Face)$

$FaceOf(R_0,X_0)$ $FaceOf(R_1,X_0)$

Similarly, suppose Hypothesis B contains all of Hypothesis B* and also includes the following formulas:

$Region(R_1)$ $SideOf(4,R_1)$

$Noise(6)$ $Solid(X_0,3Face)$

$FaceOf(R_0,X_0)$ $FaceOf(R_1,X_0)$

Hypothesis B is, of course, the correct one, as it treats Edge 6 as noise rather than as the boundary of a solid object. However, Hypothesis A has a higher explanatory value according to Equations (1) to (3), because it contains more noise terms.

This brings us, finally, to the topic of expectation and top-down information flow. Hypothesis B is underperforming because of the poor quality of the original low-level data. In particular, the faintly visible face on the left of the block has been ignored by the edge detection algorithm. But Hypothesis B can be promoted to its proper place by actively seeking to confirm those of its expectations that are not already fulfilled by the original data. Here is an example of second-layer "expectation" formulas:

$$[Solid(x,3Face) \wedge FaceOf(r_1,x)] \rightarrow$$

$$\exists\, r_2, r_3\ [FaceOf(r_2,x) \wedge Abuts(r_2,r_1) \wedge$$

$$FaceOf(r_3,x) \wedge r_2\ r_3 \wedge Abuts(r_3,r_1)]$$

This type of formula is superficially similar to a constraint. The difference is that the right-hand-side of the implication describes conditions that, if not already met, can be tested by referring back to the raw image data and reprocessing it in a specialized way. In this case, if the posited cuboid is missing any of its expected faces, this formula will trigger a search for a region having the right geometrical properties, namely, that it abuts the cuboid's known faces. Because these properties are highly specific, the search can be very sensitive. In the example of Figure 3, this will enable the detection of the previously missed facet of the block.

The final step of the reasoning process is to incorporate the results of checking the expectations of each hypothesis into a newly calculated measure of explanatory value. Again, Equations (1) to (3) are used. The updated measure is obtained simply by widening the set of low-level sensor data subject to explanation to encompass both the presence *and* absence of expectations. To allow for the explanation of the absence of an expected item of low-level data, *abnormality terms* are introduced. These are analogous to noise terms—the more of them a hypothesis is forced to contain, the less its explanatory value. This results in an increase in the explanatory value of a hypothesis whose expectations are fulfilled, and a corresponding decrease in the explanatory value of a hypothesis with unfulfilled expectations.

In this example, each hypothesis entails the expectation of a missing third region that abuts both R_0 and R_1. This is a complex expectation involving several edges, which are actively sought out using a sensitive spatially guided edge-checking algorithm. For Hypothesis A this expectation is not met, whereas for Hypothesis B it is. Now both the presence of new edges on the left of the image and the absence of looked-for edges in the center become additions to the low-level sensor data, and this results in Hypothesis B overtaking Hypothesis A in explanatory value.

A prototype abductive inference mechanism has been constructed that incorporates these ideas of explanatory value and top-down expectation. As with the Khepera experiments, the implementation is in the form of a logic programming metainterpreter. The system has been applied to visual data and has been shown to work for examples similar to that in Figure 3. The next step is to interface the system directly to Ludwig's camera, and to let the best hypotheses it generates lead the arm to nudge interesting objects on the workbench. A suitable attention mechanism is also required to guide the camera and to select portions of the visual field for processing. The present solution is for the system to seek out motion. The experimenter is able to draw attention to interesting objects by pushing them or waving them about, much as a parent does with a young infant.

3.6. Active visual perception

The preceding examples have shown how the abductive treatment can be given to a simple sensory modality in a dynamic setting (Section 3.3), and to a richer sensory modality, namely vision, in a static setting (Section 3.5). To complete the picture, this section gives an abductive account of vision in a dynamic setting (Shanahan & Randell, 2004). This is largely a matter of bolting the required logical apparatus for handling actions and change (from Section 3.3) to the abductive machinery for dealing with top-down information flow (from Section 3.5).

Consider Figure 5, which shows two images from Ludwig's camera alongside the results of applying edge detection. Between the two frames, the block has been nudged slightly. In the

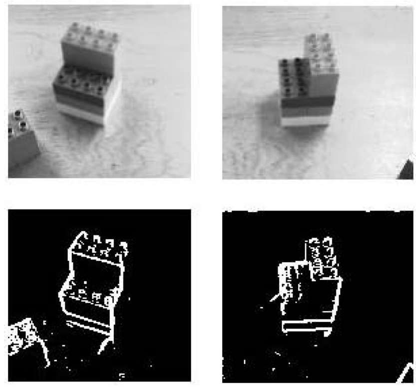

Fig. 5. Nudging a block.

first frame, the bright highlight on the right of the block has resulted in the loss of detectable edges. From the edge information in this frame alone, there is no way to determine whether the object is an L block or a T block (Figure 4). However, the ambiguity is resolvable with the aid of the extra information obtained by nudging the block and thereby gaining a view of it from a different angle. The challenge now is to formalize this process of knowledge increase in abductive terms.

To meet this challenge, a third layer of abductive interpretation is added in addition to the two layers described in the previous section. This layer has the responsibility of explaining sequences of momentary snapshots in terms of persistent solid objects. The first step that needs to be taken is to turn the predicates previously used to describe static images into fluents. This is achieved through "reification," which amounts to wrapping them in the *HoldsAt* predicate. So instead of writing *Solid*(x,a) we now write *HoldsAt*(*Solid*(x,a),t) where t is a time point.

Now suppose that two competing second-layer hypotheses have been found to explain the edge data in the first frame of Figure 5, taken at time T_1. According to Hypothesis A, the object is presenting an aspect of an L block.

$$HoldsAt(Solid(X_0,L4Face), T_1)$$

According to Hypothesis B, the object is presenting an aspect of a T block.

$$HoldsAt(Solid(X_0,T4Face), T_1)$$

Top-down information has helped to narrow down the candidate explanations to these two. But this is the best it can manage, and because the edges on the right of the block are completely obliterated, both hypotheses are accorded equal explanatory value.

The arrival of the next frame, taken at time T_2, is an opportunity to resolve the ambiguity. However, thanks to the block's surface features, edge detection cannot distinguish a clear boundary at the top of the block. So the application of static abductive interpretation to this frame would again reveal two equally good competing hypotheses. According to one hypothesis, the block is presenting an aspect of a cuboid:

$$HoldsAt(Solid(X_0,3Face),T_2)$$

According to the other hypothesis, the object is presenting a different aspect of an L block:

$$HoldsAt(Solid(X_0,L2Face),T_2)$$

But because the two frames are linked, a static analysis of the second one would be inappropriate. Instead, the abductive process must attempt to extend each of the ongoing hypotheses (in this case A and B) to accommodate the data in the new frame. To realize this, we can import a useful and well-established concept from computer vision, namely, that of an *aspect graph* (Koenderink & van Doorn, 1979). The aspect graph of a three-dimensional shape is the set of all possible aspects it can present to the viewer linked by arcs to indicate legitimate transitions from one aspect directly to another (Figure 6).

In what follows, the formula $Shape(x,s)$ represents that object x has shape s, where a shape is defined by its aspect graph. In the third layer of abduction, *Shape* formulas are made abducible. The formula $Arc(s,a_1,a_2)$ represents that there is an arc from aspect a_1 to aspect a_2 in the aspect graph of shape s. So the set of formulas required to represent the shapes illustrated in Figure 6 includes the following:

$$Arc(LBlock,L4Face,L2Face)$$

$$Arc(Cuboid,3Face,2Face)$$

Now we can use the event calculus to describe the possible aspect transitions that can result from nudging an object. The results are incorporated in the third-layer background theory Σ. Let the term $Nudge(x)$ denote an act of nudging object x, and assume that no single nudge is dramatic enough to cause an object to pass through more than one aspect transition. We need to formalize the fact that a *Nudge* action might result in no aspect transition, but it might result in the transition from the current aspect a_1 to any new aspect a_2 such that an arc exists from a_1 to a_2 in the relevant aspect graph.

$$Initiates(Nudge(x),Solid(x,a_2),t) \leftarrow$$
$$HoldsAt(Solid(x,a_1),t) \wedge Shape(x,s) \wedge$$
$$Arc(s,a_1,a_2) \wedge ByChance(TransTo(x,a_2),t)$$

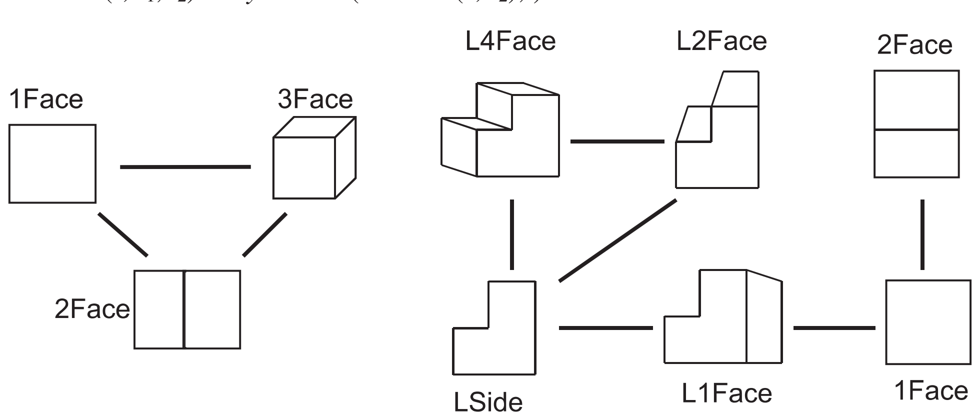

Fig. 6. Partial aspect graphs of a cube and an L block.

The *Nudge* action is an example of an action with a nondeterministic effect, and this formula illustrates one way the event calculus can be used to represent such actions (Shanahan, 1997b). The formula *ByChance*(f,t) represents that fluent f holds at time t. Unlike *HoldsAt,* there are no axioms to constrain the *ByChance* predicate, so the fluent f, which is known as a *determining fluent,* can either hold or not hold at any time. In the context of abduction, *ByChance* formulas are abducible, but make no contribution to the calculation of explanatory value. For more details, see Shanahan (1997a) or Shanahan and Randell (2004).

The *Nudge* action that was carried out between the two frames is represented by a *Happens* formula:

$$Happens(Nudge(X_0),T)$$

$$T_1 < T < T_2$$

Because we have *Arc*(*LBlock,L4Face,L2Face*), Hypothesis A can now be extended to form a third-layer hypothesis that explains the two successive frames in terms of a persistent solid object. The extended hypothesis includes the following formula:

$$Shape(X_0,LBlock)$$

By contrast, Hypothesis B cannot be extended to accommodate the data in the new frame. This is because there is no second-layer hypothesis that includes the formula $HoldsAt(Solid(X_0,a),T_2)$, such that *Arc*(*s,T4Face,a*) for any shape *s.* This is, of course, the desired outcome.

4. Related work

The idea that perception can be viewed as a form of abduction can be traced back to the philosophical writings of Peirce (1903), who carried out an extensive study of abductive reasoning. (Indeed, the term *abduction* was coined by Peirce.) The essential insight is clearly present in Peirce, although his discussion is somewhat informal. Within psychology, Rock's theory of *indirect perception,* which in turn derives from the ideas of the 19th-century scientist Hermann von Helmholtz, is the most closely related to the abductive approach of this article (Rock, 1983). Although Rock does not use the term *abduction,* that is precisely what he describes, as his theory of perception involves the generation of hypotheses about the external world to explain visual data and their testing through the acquisition of further data. Rock's theory is informally presented and unimplemented, but has the virtue of empirical support.

Returning to AI and computer science, the treatment of visual perception offered here has some points of contact with work on model-based vision dating back to the early 1980s (Brooks, 1981), and aspects of the abductive processing of image data are analogous to the matching operation at the heart of contemporary model-based object recognition systems (Jain, Kasturi, & Schunk, 1995, chap. 15). But, in general, these systems neither use logic as a representational formalism, nor carry out any form of logical inference. Moreover, the role of high-level knowledge in model-based vision is usually confined to the representation of the structure of the objects to be recognized.

The combination of abductive inference with the confirmation of expectations, here applied to visual perception, echoes the hypothetico-deductive model of scientific reasoning proposed by philosophers of science. In Kakas and Evans (1992) we find a formal description of hypothetico-deductive reasoning in a logic programming context, although the authors do not apply their ideas to perception. The hypothetico-deductive model is applied to perception in Josephson and Josephson (1994, chap. 10), although the treatment there is informal, and the authors offer few hints about how their ideas might be realized on a computer.

A rare example of an attempt to apply logic directly to image interpretation can be found in Reiter and Mackworth (1989), where the authors provided a formal definition of the relation between an image and the scene it depicts. Their definition, like its abductive counterpart in this article, relies on an axiomatic description of the way different objects in the environment give rise to raw image data. Reiter and Mackworth's treatment was purely theoretical, and their ideas were never implemented in a real vision system. Moreover, they largely neglected the issues of noise and incompleteness, which are a central concern here. However, their work is a clear precursor to the framework of this article. To my knowledge, the only other contemporary attempt to bring logic to bear on the problem of perception is the ongoing work of Pirri and her colleagues (Pirri & Finzi, 1999; Pirri & Romano, 2002).[10]

Another example of a methodologically related approach to perception is the recent work of Chella, Frixione, and Gaglio (1997, 2000), who have developed an architecture for vision with many of the same features as the approach to perception advocated here. Their architecture employs off-the-shelf techniques for low-level vision, but uses declaratively represented knowledge and high-level reasoning to suggest hypotheses that might explain the low-level data and form the expectations of those hypotheses. These expectations influence an attention mechanism that can seek out components of a specified shape in designated parts of an image.

Although the efforts of Chella et al. (1997, 2000) perhaps bear the closest resemblance in the literature to this work, as far as perception is concerned, there are still many differences. In particular, logic and logical inference play a much smaller part in their approach than in the framework of this article, and instead they make liberal use of neural networks. Their work lacks a theoretical counterpart to the abductive treatment of perception that forms the centerpiece of the research reported here. On the other hand, they use a wider range of techniques, from both AI and machine vision, and have a more sophisticated approach to shape.

5. Philosophical afterthoughts

The motivation for developing the theory of robot perception presented in this article is the belief that perception, action, and cognition form an inseparable trinity and that success in AI will not result from studying any one of those topics in isolation. More specifically, the aim is to supply a solid theoretical foundation for the design of robot perceptual systems that incorporate action, in the form of active perception, and cognition, in the form of top-down information flow. As such, the context is engineering, not philosophy.

Yet it is engineering with a philosophical edge. At every turn, concepts are deployed that derive from centuries of thought in the philosophy of mind. Time and again, aspects of the abductive theory of perception echo well-known philosophical debates and conundrums. So it seems appropriate to make some philosophical remarks. But rather than confronting the philosophy head-on, what follows are brief caricatures of a number of issues set alongside the reflections of an imaginary engineer. To bring the issues into proper focus, it is necessary to envisage a more advanced humanoid robot than we have at present, albeit built using the principles set out in this article, so the ensuing discussion has a science-fictional flavor.

5.1. Representational theories of mind

The engineer in question, being a keen amateur reader of philosophy journals, has for years been interested in the debate surrounding "intentionality" (Crane, 2003; Searle, 1983). The philosophers, she has learned, are very struck by the fact that mental states are *directed* toward things—beliefs are always *about* something, as are other so-called *propositional attitudes,* such as desires, intentions, and so on. In the 19th century, Brentano's (1874/1995) challenge to the materialists was to find a theory of mind that could account for this property of intentionality. Not only would it have to account for the fact that propositional attitudes are directed at real, physical objects, it would also have to accommodate their capacity for directedness at *nonexistent* objects, such as unicorns. Moreover, beliefs can be false. So the same account must allow for *error.*

Many of the philosophers trying to meet Brentano's (1874/1995) challenge have adhered to some form of *representational theory of mind,* in which mental states are characterized by appeal to sentence-like representations (Fodor, 1975). Much is made of the *functional role* of these representations in the overall organization of the mind. In some variants of the theory, mental processes are identified with kinds of *computation* over the sentence-like representations.

The engineer also knows a fair amount of neuroscience. So she has noted the considerable gap between the philosophers' notion of representation and the messy gray stuff that is actually found inside people's heads. She is also aware that certain brands of philosopher have themselves remarked on this sort of discrepancy and have expressed their skepticism about the whole idea of a representational theory of mind (Clark, 1998). Some of these philosophers claim that folk-psychological concepts such as the propositional attitudes will, in due course, be *eliminated* altogether from our scientific understanding of the mind (Churchland, 1979).

At this point, the engineer cannot help remarking to herself that her robots, if only they were a little more advanced, would make excellent candidates for a representational theory of mind—far more clear candidates, in fact, than human beings. Unlike the philosophers, she has little difficulty pinning down exactly what she means by an internal representation. These are sentences of first-order predicate calculus. Pressed for further details, she can easily explain how these are encoded in standard data structures familiar to any computer scientist, such as arrays, linked lists, and so on, which in turn are stored in the computer's memory in binary form. Among the various computational processes carried out by the computer are inferences from one such representation to another, and the algorithms executed by these processes provably respect the laws of logic.

5.2. Intentionality and symbol grounding

As she persists in her reading of the philosophical literature, the engineer soon discerns a number of overarching themes in the objections raised to representational theories of the human mind. One of the overarching themes goes to the heart of the whole intentionality issue. How is it that the mind's supposed representations, being composed of nothing but symbols, can actually be said to *represent* anything? The more attractive replies to this question all seem to appeal, in some way or other, to the mind's "connectedness" to the world. The symbols that go to make up a representation acquire their intentionality because they are *grounded* in the perception of objects in the physical world (Barsalou, 1999; Harnad, 1990).

Setting aside her wider concerns about representational theories of the *human* mind, the engineer is struck afresh by the comparative ease with which she can answer this sort of question for her robots, because they have all been built according to the abductive theory of perception. Take the case of the upper torso humanoid robot recently constructed (on a pitifully small budget) in her lab. Where do its internal representations come from? Well, taking visual perception as exemplary, a very clean causal story can be recounted that begins with light entering the lenses of the robot's cameras and ends with the storage in the computer's memory of structures standing for sentences of predicate calculus.

This causal story can be told at the purely physical level. But thanks to the abductive theory of perception to which the humanoid's design conforms, a more valuable, parallel narrative can be told at the level of symbols, representations, and logic. Suppose a student places a white cup on the robot's workbench, and the robot's attention mechanism, drawn by the motion, has directed its gaze on this cup. The perpetual stream of pixel arrays delivered by the robot's camera is subjected to various forms of preprocessing using off-the-shelf vision algorithms. These off-the-shelf techniques supply a symbolic description Γ of many aspects of the image—edges, corners, patches of uniform color, texture, or depth, motion cues, tracked features, and so on.

What happens next is the most important part. This symbolic description Γ is submitted to a computational process that carries out abduction. This process produces a prioritized list of explanations Δ that conform to the abductive prescription given in this article. These are lodged in the computer's memory, and their disjunction contributes to the robot's internal representation of its workbench environment. At the top of the list of predicate calculus sentences is one describing the presence on the workbench of a white vessel with a handle:

$$\exists \times [On(x, Workbench) \wedge Is(x, Cup)]$$

This sentence plays a vital role in the selection of the robot's next action. Working forward from its updated model of this situation, the robot's planner soon identifies an executable action that will bring it closer to one of its chief goals, namely, the production of tea. It carefully positions the teapot it is holding over the cup, using visual servoing, and proceeds to pour. The onlookers applaud.

The engineer is then able to explain to the audience how the robot was able to build up a "correct" representation of the scene. The robot "knows there is a cup on the table," because its system of perception conforms to an abductive theory of perception. This theory pins down the process by which the base metal of raw sensor data is transmuted into the gold of meaningful representation, and it does so in such a way that the correctness (or otherwise) of that process can be logically verified. At the same time, it is clear how the process reduces to a causal story at the level of computer science, electronics, or physics, as preferred.

So when pressed to justify the ascription of meaning to the symbols in a predicate calculus sentence such as the one previously mentioned, the engineer can appeal to more than their suggestive English names. Ideally, she would be able to claim that the previously mentioned sentence is inserted into the robot's model of this situation if and only if there is indeed a cup on the table. In practice, of course, the robot is sometimes fooled into taking noncups for cups and cups for noncups, thanks to poor lighting, unusual surface patterns, and so on. But it generally recovers the correct Δ as soon as it tries to grasp the item in question.

So the engineer's description of the robot's representations makes it sound remarkably as if they have intentionality. The engineer can explain their apparent directedness toward things in the world, under usual operating conditions. She can point to pathological cases in which they are not, in this way, directed at the things they should be, and can legitimize the sense in which these cases are indeed pathological and can thus be said to be examples of "error." Finally, she can point to the hypothetical situations that are represented, using novel combinations of the same predicates and terms, in the course of the robot's planning processes, as examples of directedness toward something nonexistent.

5.3. Empiricism and idealism

Unfortunately, during one such demonstration there is a philosopher in the audience, who asks the engineer how the terms and predicates in the background theory Σ acquire their meaning. Surely any intentionality that can be ascribed to the sentences in Δ piggybacks on the intentionality of the sentences in Σ, and their intentionality in turn is parasitic on the intentionality of the designer's mental states.

The engineer—after pointing out that, as a humble engineer, she is not really interested in intentionality as such, just in designing and building robots according to sound engineering principles—then reiterates the point that a predicate symbol such as *Cup* appears in the robot's model of the world as a causal consequence of there being a cup on the table, and this symbol is subsequently causally implicated in the robot's treating the cup as a cup. It would fulfill the same role, whether it had been put there by a human designer, or acquired by a learning algorithm, or had appeared through an evolutionary process.

But the philosopher is not entirely satisfied by this reply. He would love to ascribe intentionality to the robot's representations, he says, but he is still bothered by the origin of Σ. The engineer patiently explains that, with this generation of humanoids, it is indeed the case that the background theory Σ has been largely hand-coded by legions of hardworking PhD students. But work in, for example, inductive logic programming, has shown that similar theories can be learned, right down to the invention of new predicate and function symbols. Future generations of humanoids are likely to acquire more and more of their background theories through such learning techniques.

That is all very well, says the philosopher, but surely there will always be something in the original background theory, something to get the whole process off the ground. The engineer cheerfully concedes that this is true. The background theory will include at least a basic set of sorts for time, objects, and spatial location, and basic predicates for representing the occurrence of events, cause and effect, spatial relations, and so on. Without this foundation at least, the learning algorithms will never work. She pulls a dog-eared commentary on Kant from the shelf and concludes by quoting "the natural world as we know it … is thoroughly conditioned by [certain] features: our experience is essentially experience of a spatio-temporal world of law-governed objects conceived of as distinct from our temporally successive experiences of them" (Strawson, 1966, p. 21).

6. Concluding remarks

One way to think of the theory of perception on offer here is to contrast it with a strictly classical approach to the problem of sensing, in which some physical quantity needs to be measured using an imperfect instrument. From this standpoint, machine vision is analogous to doing graphics backward. The initial challenge is the "graphics problem"—to find a mathematical model that takes a description of the solid objects in a scene, the lighting conditions, and the position of the viewer, and renders the scene for the viewer in question. The vision problem is then just a matter of inverting the equations so the model works backward, yielding a precise description of the solid objects in the scene, given what the viewer sees. Under this approach, uncertainty is treated in the same way that any engineer trying to measure a physical quantity using a sensor would treat it, namely, by factoring noise into the mathematical model.

The theoretical paradigm for robot perception supplied by this article is very different from this classical approach, but remains entirely compatible with it. Rather than assuming that the ideal final product of the perceptual process is a quantitative, numerical description of the scene, the abductive framework assumes the final product is a qualitative, symbolic description of the scene. The former type of description is now relegated to the intermediate stage between the raw sensor data and the final product. The bridge between the quantitative and the qualitative, between the numerical and the symbolic, is made by abduction, which, drawing on high-level knowledge, is the final arbiter in the fixation of belief for the perceiving agent.

The result is an account of perception that should be attractive to engineers who want to build sophisticated robots working in realistic environments.[11] The main achievement is the provision of a precise theoretical framework in which the relation between world and acquired representation is clear. Moreover, this theoretical framework can account for several aspects of robot perception in a uniform way. These include the processing of incomplete and uncertain data, the use of top-down information flow, sensor fusion, and the interplay of perception and action. With the aid of a number of logic-based knowledge representation techniques, each of these has been assimilated into the notion that the transformation of raw sensor data into meaningful representation can be cast as a form of logical abduction.

Whether or not the abductive theory of perception has genuine philosophical implications is hard to say. But it is striking that the theory echoes so many themes in the philosophical debate on representation and intentionality, and it may help to expose philosophers engaged in that debate to unfamiliar ideas, such as top-down information flow and active perception, which are crucial to filling out any theory of perception that purports to deliver sentence-like representations.

Notes

1. It can also be argued that symbols can be grounded through interaction with a virtual environment, or through disembodied interaction with the real world (through the Internet, for example). The concern of this article, however, is grounding through embodiment.
2. Barsalou's own way of tackling this problem (for humans) is very different from the classical AI solution for robots offered here. Both approaches may be valid for their respective domains of application.
3. This distinction between logic as a representational medium and logic as a specification tool was first clearly articulated by Newell (1982).
4. For the record, my view is that the logicist approach is a viable way to engineer good robots. However, this does not entail endorsement of a computational theory of the *human* mind. From a philosophical point of view, this article shows that if the computational theory of mind is flawed, this is not for want of an account of the transition from sensor data to meaningful representation.
5. Gibson himself would hardly approve of this theory. But this does not prevent us from lifting this particular insight from his work.
6. The precise nature of the interplay between cognition and perception is controversial (Pylyshyn, 1999). That such an interplay exists is less so.
7. In this experiment, the robot's exploratory movements are not the result of planning and are not influenced by what the robot already knows. In the next section, a map-building experiment is described in which the robot reasons about its own ignorance to decide where to move next.
8. The arguments for using logic as an explicit representational medium are well known (Hayes, 1977; Moore, 1982; Nilsson, 1991). They include its possession of a mathematically precise *semantics,* its putative *universality* as a representational formalism, and the *reusability* of (declarative) logical sentences in multiple contexts.
9. In other words, one and only one hypothesis is true. Strictly speaking, this is an abuse of notation, because the $\oplus$ (exclusive or) operator does not compose in the intended way.
10. Van der Does and Van Lambalgen (2000) presented a "logic of vision" that draws on Marr's (1982) ideas. Although their main concern is to supply a semantics of perceptual reports rather than to formalize the perceptual process itself, it would be interesting to compare the relevant aspects of their logic to this approach.
11. In addition to robotics, the theoretical framework of this article has been influential in cognitive vision (Cohn, Magee, Galata, Hogg, & Hazarika, 2003) and spatial reasoning (Hazarika & Cohn, 2002; Remolina & Kuipers, 2004).

Acknowledgments

The work described in this article was supported by EPSRC grants GR/L20023 "Cognitive Robotics", GR/N13104 "Cognitive Robotics II", and GR/S02891 "Spatial Reasoning and Perception in a Humanoid Robot". Thanks to Mark Witkowski, Dave Randell, and Tim Guhl. Thanks also to the reviewers for many helpful suggestions.

References

Aloimonos, J., Weiss, I., & Bandyopadhyay, A. (1987). Active vision. In *Proceedings 1st International Conference on Computer Vision* (pp. 35–54). Washington, DC: Computer Society Press of the IEEE.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence, 48,* 57–86.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

Brentano, F. (1995). *Psychology from an empirical standpoint* (A. C. Rancurello, D. B. Terrell, & L. L. McAlister, Trans.). London: Routledge.

Brooks, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence, 17,* 285–348.

Brooks, R. A. (1991). Intelligence without reason. In J. Mylopoulos & R. Reiter (Eds.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (pp. 569–595). San Francisco: Kaufmann.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. (1999). The cog project: Building a humanoid robot. In C. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents* (pp. 52–87). New York: Springer.

Cavanagh, P. (1999). Top-down processing in vision. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 844–845): Cambridge, MA: MIT Press.

Chella, A., Frixione, M., & Gaglio, S. (1997). A cognitive architecture for artificial vision. *Artificial Intelligence, 89,* 73–111.

Chella, A., Frixione, M., & Gaglio, S. (2000). Understanding dynamic scenes. *Artificial Intelligence, 123,* 89–132.

Chrisley, R., & Ziemke, T. (2003). Embodiment. In L. Nadel (Ed.), *The Macmillan encyclopedia of cognitive science* (Vol. 1, pp. 1102–1108). New York: Macmillan.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind.* Cambridge, England: Cambridge University Press.

Clark, A. (1998). *Being there: Putting brain, body, and world back together.* Cambridge, MA: MIT Press.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Dordrecht, The Netherlands: Kluwer.

Cohn, A. G., Magee, D., Galata, A., Hogg, D., & Hazarika, S. (2003). Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction. In C. Freska, C. Habel, & K. F. Wender (Eds.), *Spatial cognition III* (pp. 232–248). New York: Springer.

Coradeschi, S., & Saffiotti, A. (2000). Anchoring symbols to sensor data: Preliminary report. In Proceedings of the Seventeenth National Conference on Artificial Intelligence and The Twelfth Annual Conference on Innovative Applications of Artificial Intelligence (pp. 129–135). Menlo Park, CA: AAAI Press.

Crane, T. (2003). *The Mechanical mind: A philosophical introduction to minds, machines, and mental representation* (2nd ed.). London: Routledge.

Fodor, J. (1975). *The Language of thought.* Brighton, England: Harvester.

Gibson, J. J. (1979). *The Ecological approach to visual perception.* Boston: Houghton Mifflin.

Green, C. (1969). Applications of theorem proving to problem solving. In D. E. Walker & L. M. Norton (Eds.), *Proceedings of the 1969 International Joint Conference on Artificial Intelligence* (pp. 219–240). San Francisco: Kaufmann.

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision, 12,* 163–186.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42,* 335–346.

Hayes, P. J. (1977). In defence of logic. In *Proceedings of the 1977 International Joint Conference on Artificial Intelligence* (pp. 559–565). San Francisco: Kaufmann.

Hayes, P. J. (1985). The second naive physics manifesto. In J. R. Hobbs & R. C. Moore (Eds.), *Formal theories of the commonsense world* (pp. 1–36). Norwood, NJ: Ablex.

Hazarika, S., & Cohn, A. G. (2002). Abducing qualitative spatio-temporal histories from partial observations. In D. Fensel, F. Giunchiglia, D. L. McGuinness, & M-A. Williams (Eds.), *Proceedings of the Eighth International Conference on the Principles of Knowledge Representation and Reasoning* (pp. 14–25). San Francisco: Kaufmann.

Jain, R., Kasturi, R., & Schunk, B. G. (1995). *Machine vision.* New York: McGraw-Hill.

Jones, J. L., & Flynn, A. M. (1993). *Mobile robots: Inspiration to implementation.* Wellesley, MA: Peters.

Jospehson, J. R., & Josephson, S. G. (1994). *Abductive inference: Computation, philosophy, technology.* Cambridge, England: Cambridge University Press.

Jung, C. G. (1998). Situated abstraction planning by abductive temporal reasoning. In H. Prade (Ed.), *Proceedings 13th European Conference on Artificial Intelligence* (pp. 383–387). Chichester, England: Wiley.

Kakas, A. C., & Evans, C. (1992). Hypothetico-deductive reasoning. In ICOT staff (Eds.), *Proceedings of the 1992 International Conference on Fifth Generation Computer Systems* (pp. 546–554). Amsterdam: IOS Press.

Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211–216.

Kowalski, R. A. (1995). Using meta-logic to reconcile reactive with rational agents. In K. R. Apt & F. Turini (Eds.), *Meta-logics and logic programming* (pp. 227–242). Cambridge, MA: MIT Press.

Lespérance, Y., Levesque, H. J., Lin, F., Marcu, D., Reiter, R., & Scherl, R. B. (1994). A logical approach to high-level robot programming: A progress report. In B. Kuipers (Ed.), *Control of the physical world by intelligent systems: Papers from the 1994 AAAI Fall Symposium* (pp. 79–85). Menlo Park, CA: AAAI Press.

Lévy, F., & Quantz, J. J. (1998). Representing beliefs in a situated event calculus. In H. Prade (Ed.), *Proceedings 13th European Conference on Artificial Intelligence* (pp. 547–551). Chichester, England: Wiley.

Lifschitz, V. (1994). Circumscription. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *The handbook of logic in artificial intelligence and logic programming: Vol. 3. Nonmonotonic reasoning and uncertain reasoning* (pp. 297–352). Oxford, England: Oxford University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York: Freeman.

McCarthy, J. (1986). Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence, 26,* 89–116.

McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In R. Thomason (Ed.), *Philosophical logic and artificial intelligence* (pp. 161–190). Dordrecht, The Netherlands: Kluwer Academic.

Mondada, F., Franzi, E., & Ienne, P. (1993). Mobile robot miniaturisation: A tool for investigation in control algorithms. In T. Yoshikawa & F. M. Yazaki (Eds.), *Experimental Robotics III: The 3rd International Symposium* (pp. 501–513). New York: Springer.

Moore, R. C. (1982). The role of logic in knowledge representation and commonsense reasoning. In *Proceedings 1982 American Association for Artificial Intelligence Conference* (*AAAI 82*) (pp. 428–433). Cambridge, MA: MIT Press.

Morgenstern, L. (2001). Mid-sized axiomatizations of commonsense problems: A case study in egg cracking. *Studia Logica, 67,* 333–384.

Mueller, E. T. (2004). Event calculus reasoning through satisfiability. *The Journal of Logic and Computation, 14,* 703–730.

Muggleton, S. (1991). Inductive logic programming. *New Generation Computing, 8,* 295–318.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18,* 87–127.

Niebur, E., & Koch, C. (1998). Computational architectures for attention. In R. Parasuraman (Ed.), *The attentive brain* (pp. 163–186). Cambridge, MA: MIT Press.

Nilsson, N. J. (Ed.). (1984). *Shakey the robot.* SRI Technical Note No. 323. Menlo Park, CA: SRI International.

Nilsson, N. J. (1991). Logic and artificial intelligence. *Artificial Intelligence, 47,* 31–56.

Peirce, C. S. (1903). Pragmatism as the logic of abduction. In Peirce Edition Project (Eds.), *The essential Peirce: Selected philosophical writings* (Vol. 2, pp. 226–242). Bloomington, IN: Indiana University Press.

Peterson, M. A. (2003). Vision: Top-down effects. In L. Nadel (Ed.), *The Macmillan encyclopedia of cognitive science* (Vol. 4., pp. 500–504). New York: Macmillan.

Pirri, F., & Finzi, A. (1999). An approach to perception in theory of actions: Part I. *Electronic Transactions on Artificial Intelligence, 3,* 19–61.

Pirri, F., & Romano, M. (2002). A situation-Bayes view of object recognition based on SymGeons. In C. Baral & S. McIlraith *Proceedings Third International Cognitive Robotics Workshop* (pp. 88–95). Menlo Park, CA: AAAI Press.

Poole, D. (1995). Logic programming for robot control. In C. S. Mellish, *Proceedings 1995 International Joint Conference on Artificial Intelligence (IJCAI95)* (pp. 150–157). San Francisco: Kaufmann.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22,* 341–364.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80,* 127–158.

Remolina, E., & Kuipers, B. (2004). Towards a general theory of topological maps. *Artificial Intelligence, 152,* 47–104.

Reiter, R., & Mackworth, A. K. (1989). A logical framework for depiction and image interpretation. *Artificial Intelligence, 41,* 125–155.

Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

Russell, S., & Norvig, P. (2002). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Scherl, R., & Levesque, H. (1993). The frame problem and knowledge producing actions. In R. Fikes & W. Lehnert, *Proceedings 1993 American Association for Artifical Intelligence Conference* (*AAAI 93*) (pp. 689–695). Cambridge, MA: MIT Press.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3,* 417–424.

Searle, J. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, England: Cambridge University Press.

Shanahan, M. P. (1996a). Noise and the common sense informatic situation for a mobile robot. In D. Weld & B. Clancey, *Proceedings 1996 American Association for Artificial Intelligence Conference* (*AAAI 96*) (pp. 1098–1103). Menlo Park, CA: AAAI Press.

Shanahan, M. P. (1996b). Robotics and the common sense informatic situation. In W. Wahlster, *Proceedings 1996 European Conference on Artificial Intelligence* (*ECAI 96*) (pp. 684–688). Chichester, England: John Wiley & Sons.

Shanahan, M. P. (1997a). Noise, nondeterminism and spatial uncertainty. In B. Kuipers & B. Webber, *Proceedings 1997 American Association for Artificial Intelligence Conference* (*AAAI 97*) (pp. 153–158). Menlo Park, CA: AAAI Press.

Shanahan, M. P. (1997b). *Solving the frame problem: A mathematical investigation of the common sense law of inertia.* Cambridge, MA: MIT Press.

Shanahan, M. P. (1999). The event calculus explained. In Wooldridge, M. J. & Veloso, M. (Eds.), *Artificial Intelligence Today* (pp. 409–430). New York: Springer-Verlag.

Shanahan, M. P. (2000). Reinventing Shakey. In J. Minker (Ed.), *Logic-based artificial intelligence* (pp. 233–253). New York: Academic.

Shanahan, M. P. (2002). A logical account of perception incorporating feedback and expectation. In D. Fensel, F. Giunchiglia, D. L. McGuinness, & M-A. Williams (Eds.), *Proceedings of the Eighth International Conference on the Principles of Knowledge Representation and Reasoning* (pp. 3–13). San Francisco: Kaufmann.

Shanahan, M. P. (2003). The frame problem. In L. Nadel (Ed.), *The Macmillan Encyclopedia of Cognitive Science* (Vol. 2., pp. 144–150). New York: Macmillan.

Shanahan, M. P. (2004). An attempt to formalise a nontrivial benchmark problem in common sense reasoning. *Artificial Intelligence, 153,* 141–165.

Shanahan, M. P., & Randell, D. (2004). A logic-based formulation of active visual perception. In D. Dubois, C. Welty, & M.-A. Williams, Ninth International Conference on the Principles of Knowledge Representation and Reasoning (pp. 64–72). Menlo Park, CA: AAAI Press.

Shanahan, M. P., & Witkowski, M. (2001). High-level robot control through logic, In *Intelligent Agents VII* (pp. 104–121). New York: Springer.

Strawson, P. F. (1966). *The bounds of sense.* London: Methuen.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, MA: MIT Press.

Van der Does, J., & Van Lambalgen, M. (2000). A logic of vision. *Linguistics and Philosophy, 23*(1), 1–92.

Von Uexküll, J. (1957). A stroll through the worlds of animals and men. In C. H. Schiller (Ed.), *Instinctive behaviour: The development of a modern concept* (pp. 5–80). Guilford, CT: International Universities Press.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 19–91). New York: McGraw-Hill.

Winston, P. H. (1984). *Artificial intelligence* (2nd ed.). Boston: Addison-Wesley.

Zilberstein, S., & Russell, S. J. (1993). Anytime sensing, planning and action: A practical model for robot control. In R. Bajcsy, *Proceedings 1993 International Joint Conference on Artificial Intelligence* (*IJCAI 93*) (pp. 1402–1407). San Francisco: Kaufmann.