Linköping Electronic Articles in
Computer and Information Science
Vol. (1999): nr

An approach to Perception in Theory of Actions: Part I

Fiora Pirri
Alberto Finzi

Dipartimento di Informatica e Sistemistica (DIS)
University of Rome "La Sapienza"
Rome, Italy

Linköping University Electronic Press
Linköping, Sweden

http://www.ep.liu.se/ea/cis/1999//

Published on , 1999 by
Linköping University Electronic Press
581 83 Linköping, Sweden

Linköping Electronic Articles in
Computer and Information Science
ISSN 1401-9841
Series editor: Erik Sandewall

Typeset by the author using LaTeX
Formatted using étendu *style*

Recommended citation:
<Author>. <Title>. Linköping Electronic Articles in
Computer and Information Science, Vol. (1999): nr .
http://www.ep.liu.se/ea/cis/1999//. *, 1999.*

This URL will also contain a link to the author's home page.

Abstract

We present a preliminary approach to the coordination and control of sensing for action oriented perception in the Situation Calculus. We essentially consider visual perception, although the approach can be generalized to other forms of perception. We show that perception is an inference process that uses different levels of refinement, namely, direct perception, selective sense perception and, finally, meaningful perception. We take a cautious point of view to perception which means that we see it as a reasoning process that can be ambiguous and mistaken. Agents can be committed to be attentive so as to make no mistakes. When the theory does not entail that a mistake occurred then perception is considered veridic. Ambiguity is resolved via hypothetical reasoning; we call this non monotonic reasoning process that combines sensing, logical inference and hypotheses, meaningful perception.

1 Introduction

Preamble We present an approach to the coordination and control of sensing for action oriented perception in the situation calculus [53, 67]. In this paper we present the first part of the work, in which we consider perception as an inference process that uses data and information acquired via sensing actions and dialectically interacting with inferences based on prior information. In this first part we treat sensing actions as queries to a sensory system without giving any formal detail on the nature of the sensory system. We shall only assume that it is a visual system. Here and there we will give some hints on the sensory system of reference and on a possible way to represent the sense actions/sensory system interaction: in this paper no precise theory is behind these hints.

Perception is the ultimate purpose of sensation; it is the reasoning or mental process that governs the interpretation of the environment using senses, inner memory, previous knowledge and previous perceptions: it is the ability to extract from sensory information meaningful and useful representation about the world. A single simple act or event like that of recognizing that *there is a red cube on the top of the table in the studio* hides a very complex process involving many specific analytic and physiological processes of different essential qualities: the human perceptual system is constructed from many interacting modules, each dealing with a different aspect of the sensory task. This multi modular nature of perception makes it so difficult and at the same time so challenging to deal with.

An enormous amount of work has been done in the effort of understanding perception from different research areas: psychology [29, 32, 36, 41, 52, 60, 71], cognitive neuroscience (for an overview see[77]), logic [70, 84], visual recognition [12, 59, 83], cognitive science [15, 35] and cognitive robotics [3, 4, 22, 23, 24, 25, 28, 46, 47, 48, 55, 73, 74] and many others (for an overview and a deep treatment of the different approaches to visual perception, excluding the cognitive approaches, see Edelman [30] and also Ulmann [83]). Many research areas are directly connected with perception, in particular with visual perception: the mathematical models for pattern analysis and recognition, topology and mereotopology as logics for spatial reasoning [78, 44, 16, 31, 18, 38, 39, 2, 37, 7, 8, 20, 21, 1], computer graphics [34], geometry and theories of shape [79, 43] (for an overview on the mathematical theories of shape see [57]). This extraordinary amount of work makes one feel uneasy to talk about perception without providing a straight point of view for all the above mentioned research areas or without having the ability to compile all them into a powerful interpretation of perception in the way e.g. of Marr [51].

Nevertheless we deal with cognitive agents that have to achieve complex tasks in which they need to use visual perception and other senses to actively interact with the environment and they have to organize their sensing actions so as to correctly respond to the different needs of each specific task. The kinds of problems an agent is faced with are:

1. Getting data from the environment in a suitable form.
2. Using prior knowledge of the world to interpret these data;
3. Suitably adding these data to the knowledge of the world to solve the task.

The above three items require coordination between acting and perceiving in which not every action is necessarily under the control of the agent's reasoning. For example (see Treisman [80]) the simple detection of features can be interpreted as a preattentive and automatic, non controllable process. E.g. detecting whether the environment is indoors or outdoors, whether it is a studio or a soccer field. On the other hand the perception of an object *a pot* or a fact *the pot is on the table* requires different kinds of bindings (see [81]) such us property binding, part binding, conditional binding – the interpretation of one property depends on another – temporal binding, location binding and so forth. In other words it requires a specific, attentive, sensing in which some reasoning process, using prior knowledge about the environment, is involved. If we consider at a very basic level two layers of perception, a preattentive and an attentive one, then we could divide the interpretation of the image from the perception of the scene. By the interpretation of the image we intend the preliminary phase in which the low level data, from the vision system, are suitably transformed in terse high level descriptions. This phase also has to take into account memory and prior knowledge. On the other hand, by the perception of the scene we intend the process in which a core role is played by the sensing actions that query the vision system using the previously mentioned high level descriptions. Sensing actions, in fact, are under the control of the agent that uses them to precisely identify the elements of a scene and their dynamics. For instance there are tasks like collision avoidance in path-planning [58] for which in general there is no need for a perceptual model that accounts for a precise perception of what the obstacle is about.

In this work we are concerned with this second basic level of perception and we treat sensing actions as non deterministic actions that query the vision system. A very preliminary problem to resolve is to *interface* sensing actions with the sensory system. A sensory system like a vision system can provide a very sophisticated elaboration of the input image, up to the recognition of single objects in a scene. Despite this ability, the most sophisticated vision system cannot be a useful perceptual system for cognitive agents, if it does not account for dynamic knowledge of the environment. We cope with the idea [70, 71] that the recognition of a fact or object in a scene needs previous knowledge of the context and previous information about the recognized fact/object or about some similar fact/object. The core of the argument is the following. The recognition task, or the interpretation of an image, is facilitated by the information a cognitive agent has about the environment, and by the representation of these data in a high level formalism. Moreover, following Biederman's approach [9], we suppose that the terse high level description from the interpretation of an image specifies parts, attributes and relations independently and explicitly. Such a structural representation, relies on a compositional view of what can be perceived.

In cognitive robotics, the purpose of object recognition is to assign a high level meaning or definition of an object to some data sensed in the environment of a robot. Commonly used sensors in robotics for generating 3D data are multiple camera stereo rigs, laser range finders and structured light range cameras. We have actually a vision system called PerceptONE, which is a modular system for the recognition of simple 3D objects. We call simple those shapes that can be approximated by convex polyhedra and that can be used in a suitable combination to get more complex objects. The system, implemented in C++, uses a color stereo video system with two Sony camera modules, two frame-grabbers from Matrox-Meteor and a pan-tilt of DirectedPerception. The system acquires the image from the camera, using a MIL-LITE library, performs noise elimination with a filter based on SUSAN filter, developed by Brady and Smith [76], uses a particular method for segmentation, based on HSV saturation and, finally, it extracts the borderline of a simple object using an extension of Canny algorithm [14],[56]. From the borderline the vertices and the edges are identified and in particular special kinds of points, called multivertices [26]. These last are used for the alignment of the image object with a model obtained by a classification process, [6], based on Huttenlocher and Ulmann's [40] method [19]. The alignment leads to the recognition of the simple object.

The role of sensing actions is to query the vision system to choose distinctive features in a scene, which we call *perceptibles* and combine and group them to get more complex perceptibles, using primary knowledge, inner memory and prior perception. We propose that sensing actions are performed in a control loop setting: sensory information provides feedback that is controlled by sensing actions. Sensing actions affect percepts and, as percepts are contextually defined, prior percepts affect other percepts in a chain of causality [32]. Following the theory of indirect perception [70, 71], where perception is interpreted as a problem-solving process using prior perception, and the structural and compositional approach of Biederman [12], we propose a logical framework that accounts for both the organization of sensing actions in a hierarchy of percepts and the perceptual inference process. We take Rock's approach in its full generality to treat perception in the theory of actions: our goal is to provide a reasoning mechanism that allows an agent to correctly fill in the lack of information, by perceiving facts from the environment and making hypotheses.

We divide the truths about the world into analytic truths and accidental truths (in the sense of Aristotle *Second Analytic*, 71b 9-35). That is, analytic truths are the first principles or prior knowledge about the world the agent is endowed with together with those conclusions entailed by them. Accidental truths are what the agent gains through perception. Our view, which we call *cautious* as opposed to *brave* (a discussion on this is in Section 3) considers the outcome of sensing actions subject to mistakes therefore when the perception of a fact is in disagreement with the analytic truth then the analytic truth is always chosen. On the other hand the brave approach would choose the outcome of sensing. Accidental truths, however, will be added as hypotheses to the prior knowledge whenever no mistake about perception can be entailed. In this setting the information the agent gets from the world can be dynamically added and made useful in order to accomplish tasks otherwise impossible, when the agent has incomplete knowledge of the environment, that is, her environment is an *open world.*

For this purpose we distinguish among three levels of perception. First we formulate *Direct perception* (Section 2.2) as an early inference mechanism that uses prior perception, can be ambiguous and is definitely independent of the agent's reasoning process. This early inference admits mistakes, that is, an agent can mistake a glass door for an open door even if her sensors are functioning perfectly. Selective sense perception (Section 4) coalesces both perceptual inference and analytical inference, still can admit mistakes but only when not accurate. When it is accurate sense perception is considered veridic. When ambiguity cannot be resolved, the inference can resort to hypotheses bared by veridic sense perception, obeying a parsimonious principle. The perceptual process culminates in final perception, which we call *meaningful perception*, following [29] (see Section 5). We formalize it as a form of hypothetical reasoning that uses veridic percepts as hypotheses.

There are many different approaches to visual perception, from cognitive science to cognitive neuroscience. In cognitive science, Gärdenfors in [35] proposes, as opposed to the *symbolic* approach one in which perception is layered into three levels. The first, or sub symbolic level, is concerned with the direct elaboration of data as they are extracted from sensors: neural networks are considered suitable for this level. The intermediary level is the conceptual one and finally the last level is the symbolic (or linguistic) level. Gärdenfors approach is very comprehensive and it has many similarity with both Marr and Kosslyn approaches [51, 59]. In cognitive robotics sensing actions have been considered in the context of the Feature and Fluents framework (e.g. [72, 22, 23, 24, 25]), in the context of the Situation Calculus (e.g [48, 46, 47, 62]), in the context of the action description language $\mathcal{A}$ (e.g. [4, 50]) and in the context of the Event Calculus (for an overview see [74]). In particular Coradeschi in [23] considers, similarly as we does, the *misperception problem*, although in her approach misperception is caused by contradictory observations that rise the problem of revising the knowledge base, this point is better discussed in Section 3. Following [55], Scherl and Levesque in [73] have been the first in the framework of theories of actions to give an ontology for sensing actions, as actions affecting only the agent's mental state. They also provide a solution to the frame problem for sensing actions. In general the approaches to sensing so far introduced in the Situation Calculus (e.g. [48, 46, 28, 47]) give a logical framework for sensing actions, also in the context of planning and on-line execution, but disregard the problem of perception. The difference between a knowledge and a perceptual approach is concerned not only with the problem of the interpretation of the scene but also with the role played by knowledge. Perception is the inference process that gives meaning to sensation elaborating the accidental truths together with prior knowledge, being thus a layered process in which many essential aspects of the visual representation are involved. A knowledge approach is more clean but hardly deals with all these aspects as far as it coerces the outcome of sensing actions directly into knowledge.

In cognitive neuroscience and in vision there are a great variety of theories from the ecological [36] to the computational [82, 52, 51, 41, 59] (for an overview [29, 60, 77, 30]). Recently Edeleman in [30] has developed a new notion of visual representation based on second order isomorphism, between representation and their models, which can be used to preserve second-order qualities such as similarity. Edelman's approach, as he also notices, neglects compositionality. If compositionality might not be needed for tasks such as object identification it is certainly essential in theories of actions in which components of elements appearing in an image can be moved and modified. Although there are many formal approaches at the object-recognition level, it seems to us that so far there is no formal model, at the epistemological level, that accounts for an agent's inference process that coordinates sensing and control actions to interact with a changing environment. We believe that logical formalisms provide the right tool to account for the problem solving approach of [71], as far as agents' perception in the theory of actions is concerned. This paper lends support to this view and we think that the Situation Calculus is well suited for this approach.

2 Perception in the Situation Calculus

The Situation Calculus [53, 67, 68] is a well known formalism for dealing with the theory of actions. We assume the reader to be familiar with the essential features of situation calculus and with first-order logic of which Situation Calculus is a dialect. We refer the reader to [49, 69, 61] for an exhaustive presentation of the the Situation Calculus.

2.1 Preliminaries

We recall that the language $\mathcal{L}_{sitcalc}$ of Situation Calculus has three disjoint sorts: *action* for actions, *situation* for situations and *object* for everything else depending on the domain of application. We use $\wedge$, $\neg$ and $\exists$, with the usual definitions of a full set of connectives and quantifiers. In universally quantified formulae, quantifiers are omitted. We refer the reader to [69], for a detailed presentation of the alphabet of $\mathcal{L}_{sitcalc}$ and for the metalanguage adopted to denote terms and formulae of the language. The alphabet includes relations and functions called Fluents, because their truth value depends on the history of actions performed by an agent: a history is designated by a situation. A situation is the last argument of Fluents, e.g $Color(block1, green, do(paint(block1, green), S_0))$, whose situation is $s = do(paint(block1, green), S_0)$. Fluents can be relational and functional. To simplify the presentation, we omit functional Fluents, we retain only one special functional fluent for the visual image of the scene. In this paper all Fluents will be *observable Fluents*. By observable we mean the the sensory system can observe the fact or property or object the fluent is designating. For example $On(block5, block3, s)$ is a relation between $block5$ and $block3$ that holds in s and can be observed, as far as it holds, by the sensory system. Related to the observable Fluents are the *perceptibles*, that is, terms that denote what is perceived about facts, properties or objects. Since we consider only one sense, namely the vision system, we identify the perceptibles with the data of the image. In fact, to account both for these data and for the geometric description of object we shall assume that the object domain includes the reals $\mathbb{R}$. We do not provide any axiomatization, but rely instead on the standard interpretation of the reals and their operations (addition, multiplication, etc.) and relations ($<, \leq$ etc.), (see [69]). We add a new sort *images* for 3-D points, so that $A \subset \mathbb{R}^3$ is a set of sort images, where A is intended to represent the matrix of points of the bitmap of the image delivered by the vision system. For example the closed set $(0,4) \times (0,4) \times (0,0)$ which is a $2D$ square. Observe that in the image the z coordinate is fixed at 0. In the same spirit as for the reals, we shall not axiomatize sets, but instead rely on the standard interpretation of sets and their operations (union, intersection, etc.) and relations (membership, subset, etc.) (see [69] for a discussion about sets). Moreover, in this work we will also skip any axiomatization for topology and mereotopology (see [2, 7, 8, 20, 21]) which are required to account for the relations holding among shapes and regions in the scene. We shall refer the reader to the vast literature on this topic, to which we shall come back later. We shall also add the sort *degree of truth* that we denote with pr, whose domain, in this presentation, is limited to $\{0, 1\}$.

To the alphabet of $\mathcal{L}_{sitcalc}$ we add new function symbols of sort images, the *perceptibles*, and of sort action, *sensing actions*, the special relational Fluents $Percept$ and $Occluded$, and the special functional fluent $Scene$. We shall also introduce relational Fluents, by definitions, like $Align$, $Mistaken$,

$SePercept$, $Connected$ and so forth. The advantage of introducing concepts by definitions is that we can rely on the compositional laws of logic to get new properties on the basis of simpler ones. The disadvantage is that we cannot take into account all the exceptions that naturally enter into the explanation of a concept, as in general we may lack either sufficient or necessary conditions (see e.g. the discussion on natural kinds in [54]).

As Fluents account for physical events that are affected by *control* actions, perceptibles account for the mental events– the agent's inner denotation of the physical events – that are affected both by sensing and control actions. Formally a perceptible is a term of a general sort *perceptible* which should be the result of the union of many sorts that account for the different data that sensing can interpret. Since in this work we limit the notion of sensing to that of vision we assume that a perceptible is of sort images, nevertheless we shall specifically refer to it as sort *perceptible*. A perceptible differs from a fluent also because it lacks the argument of sort situation: $isF : ((action \cup object)^n) \mapsto images$.

A perceptible is taken as argument by the fluent $Percept$, which traces the sensory experience of the agent, and is of the form:

$$Percept : (perceptible \times pr \times situation).$$

The fluent $Percept$ is called a *percept*, e.g $Percept(isColor(block1, green), 1, s)$. Likewise a perceptible is taken as argument by the fluent:

$$Occluded : (perceptible \times situation).$$

Sensing actions are terms of sort action of the form: $sense : perceptible \times pr \mapsto action$, with $pr \in \{0, 1\}$: 0 and 1 are threshold, 1 means that a given property or fact is sensed to hold and 0 that it sensed not to hold. For example in $sense(isColor(block1, green), pr)$, $pr = 1$ means that the sensory system is in the condition to accept the sensing action and the answer of the vision system is that the perceptible holds in the scene. If $pr = 0$ then either the preconditions for sensing data in the scene are not satisfied or the perceptible does not hold true in the scene.

The fluent symbol $Scene(s)$ is of the form $Scene : situation \mapsto images$, and it is used for the description of the perceptibles in the scene. So, for example, $isCube(block5) \subset Scene(s)$ is true, in the situation s, if there is a region in the scene matching the set of points denoted by the perceptible $isCube(block5)$ and this set, after suitable transformations, can be aligned with the model of the cube.

2.2 A basic theory of action and perception

A basic theory of action $\mathcal{D}$ is the set:

$$\mathcal{D}_{S_0} \cup \mathcal{D}_{ap} \cup \mathcal{D}_{ss} \cup \mathcal{D}_{una} \cup \Sigma \tag{1}$$

Here $\mathcal{D}_{S_0}$ is the initial database: the set of formulae that mention only the situation term S_0 and no other term of sort situation. $\mathcal{D}_{ap}$ is the set of action precondition axioms, one for each term of sort action; $\mathcal{D}_{ss}$ is the set of successor state axioms, one for each fluent; $\mathcal{D}_{una}$ is the set of unique name for actions and Σ is the set of domain independent axioms specifying the properties of the domain of sort situation. We refer again to [49, 69, 61] for details on each element in the above set.

A basic theory of action and perception is defined analogously: the set $\mathcal{D}_{ss}$ includes, together with the successor state axioms for each fluent in the language, also a successor state axioms for the Fluents *Percept* and *Occluded* and for the functional Fluents *Scene*, for each perceptible in the language. $\mathcal{D}_{ap}$ includes action precondition axioms for sensing actions and $\mathcal{D}_{una}$ comprehend together with the set of unique name for actions unique name axioms for perceptibles. To these sets we shall add a set $\mathcal{D}_{da}$ which are the axioms for describing the representation of objects in the scene, and the set $\mathcal{D}_{gd}$ which should describe the class of primitive objects (see for example Biedermann's *geons* [9]) . The above set $\mathcal{D}_{da} \cup \mathcal{D}_{gd}$ together with a set $\mathcal{T}_{\ G}$ of foundational axioms for the geometric description of primitive object and for the topology of their relations holding in the image will be presented in a future work, which is in progress. The presentation here is mainly concerned with the notion of perception itself and its formal dialectic with control actions and analytic truth.

Conventions In the ongoing presentation we shall adopt the *normal form* for successor state axiom as described in [69] together with the notions, definitions and solution to the frame problem, from [67, 68] as described in [69]. In particular we shall make use of symbols like $\gamma_F^+(\vec{x}, a, s)$, $\gamma_F^-(\vec{x}, a, s)$, $\delta_{isF}^+(\vec{x}, a, s)$ and $\delta_{isF}^-(\vec{x}, a, s)$ to denote first order formulae playing a particular role in the successor state axioms in normal form. The role they play is the following: the truth value of a fluent F or *Percept*, in the situation $do(a, s)$ is affected by the truth of certain sentences which, together with the event of action a, explain and cause the change in the fluent value. These sentences play the role of a context for the action a and we say that the context is *positive*, and we denote it with either $\gamma_F^+(\vec{x}, a, s)$, or $\delta_{isF}^+(\vec{x}, a, s)$ when it explains, together with a, a positive change in the fluent value and it is negative otherwise. In this last case we denote it with either $\gamma_F^-(\vec{x}, a, s)$ or $\delta_{isF}^-(\vec{x}, a, s)$, depending on the successor state axiom where they are used. Therefore we shall call the $\gamma(\vec{x}, a, s)$'s and $\delta(\vec{x}, a, s)$'s, that capture these conditions, the *contexts* for the action a. We shall distinguish between *context sensitive* actions and *context free actions*. In the case of the context $\delta_{isF}(\vec{x}, a, s)$ we shall assume that it mentions only sensing actions related to the perceptible isF. Moreover both $\gamma_F^+(\vec{x}, a, s)$ and $\gamma_F^-(\vec{x}, a, s)$ cannot mention sensing actions. We shall make a completeness assumption on the contexts under which an action can lead a fluent to become true (resp. false) in the successor state (see [67, 68]) and we impose some consistency conditions that ensure the contexts to be coherently formulated so that, for example, we would not admit $\gamma_F^+(\vec{x}, a, s) = \gamma_F^-(\vec{x}, a, s)$ and $\gamma_F^+(\vec{x}, a, s) \rightarrow F(do(\vec{x}, a, s))$ and $\gamma_F^-(\vec{x}, a, s) \rightarrow \neg F(\vec{x}, do(a, s))$ (see [67, 68]).

Consistency conditions for contexts

$$\begin{array}{l}\neg\exists(\gamma_F^-(\vec{x}, a, s) \wedge \gamma_F^+(\vec{x}, a, s)),\\ \neg\exists(\delta_{isF}^+(\vec{x}, a, s) \wedge \delta_{isF}^-(\vec{x}, a, s)),\\ \neg\exists(\delta_{isF}^{\pm}(\vec{x}, a, s) \wedge \gamma_F^{\pm}(\vec{x}, a, s))\end{array} \tag{2}$$

Where $\neg\exists(\delta_{isF}^{\pm}(\vec{x}, a, s) \wedge \gamma_F^{\pm}(\vec{x}, a, s))$ is intended for any combination of $+$ and $-$, and $\exists$ quantifies over all the free variables in $\psi(\vec{x}, a, s)$ for ψ either γ or δ. For example, from the above assumptions follows that:

$$\forall\ p\ pr\ a\ s.\ a = sense(p, pr) \rightarrow \neg\gamma^+(a, s) \wedge \neg\gamma^-(a, s) \tag{3}$$

In accordance with problem solving approach to perception [70, 71], in which perception is interpreted as a mental event subject to causal laws, therefore based on prior perceptions, implying a perception-perception chain of causality, we define causal laws for perception. We use the formulae $\delta^+_{isF}(a,s)$ and $\delta^-_{isF}(a,s)$ to account for the correlation and causal laws among Percepts: perceived orientation is the suitable context for perceived shape, perceived adjacency explains the perception of the composition of an object and so forth [71].

Following the same ideas of [67, 68] causal laws for perception can be formulated as negative and positive contexts.

Example 1 Consider the following causal laws:

1. $a = sense(isCube(x), 1) \rightarrow Percept(isCube(x), 1, do(a, s))$;
2. $Fragile(x, s) \wedge a = drop(x) \rightarrow Broken(x, do(a, s))$;
3. $\exists y Percept(isClassRoom(y), 1, s) \wedge \exists w Percept(isWall(w), 1, s) \wedge$ $Percept(isPart(w, y), 1, s) \wedge Percept(isAttached(x, w) 1, s) \wedge$ $a = sense(isBlackboard(x), 1) \rightarrow Percept(isBlackboard(x), 1, do(a, s))$.

In the first item the sensing action $sense(isCube(x), 1)$ is context free, in the second item the action $drop(x)$ is context sensitive and in the third the sensing action $sense(isBlackboard(x), 1)$ is context sensitive. Observe that $\gamma^+_{Broken}(x, a, s)$ is $(Fragile(x, s) \wedge a = drop(x))$, $\delta^+_{isCube}(a, s)$ is $a = sense(isCube(x), 1)$ and $\delta^+_{isBlackboard}(x, a, s)$ is:

$$\exists y Percept(isClassRoom(y), 1, s) \wedge \exists w Percept(isWall(w), 1, s) \wedge \\ Percept(isPart(w, y), 1, s) \wedge Percept(isAttached(x, w), 1, s)$$

□

In the last item, the context plays the role of deciding whether to accept or not the answer from the vision system. When the vision system acknowledges a 1 to the query $sense(isBlackboard(x), 1)$, the context decides whether the perception should hold or not.

2.3 Descriptions of Perceptibles

As we have argued in the Preliminaries, perceptibles are terms of sort image, that is, they are nothing else than sets of points of the bitmap that constitutes the image of the scene, given a coordinate system of reference. A perceptible, as we formulate it, could be either a simple object, e.g. $isCube$, or a more complex one like $isHouse$, or a fact like a relation holding between objects, e.g. $isOn$. The representation of perceptibles as points in the image and the relations holding among them have to be formulated according to a rigorous topological or mereotopological formalization that treats these sets either as regions or sets of points (for a discussion see [2, 7, 8, 20, 37] and also [38, 39, 31] and the approaches on standardized sets operations taken in the graphics literature, e.g [65])). The topological relations would account for the different levels of composition required in the description of a perceptible.

Our view is to combine the alignment method (see [6, 40, 83]) together with the structural one (see e.g. [9, 10, 11, 12, 13, 52]) in which the object parts and their spatial relations are represented categorically, that is, primitive components of objects are metrically invariant and are brought together via rigorous laws of composition. The choice of primitive components is essential because, as a second level of description, in order to use the alignment method, models have to be provided for the primitives. A model is simply a category which prototypically represents all the possible instances of a perceptible in the scene. For example in PerceptONE a geometric model is just a b-rep (see [34, 66]).

The role of a description of a perceptible is to decompose the perceptible according to the composition laws until perceptibles of primitive elements are found. At this point the task of the alignment [40, 5, 83] consists in matching the primitive elements with their models. All the above specifications are not treated here and, as we have elsewhere argued, they will be treated in future work. Nevertheless, since the preconditions for sensing actions depend on what is sensed in the scene, we shall give just a hint on a possible definition of the description of a perceptible in the scene. So, for example, the description of a composed perceptible like e.g. $isOn(x, y)$ could be as follows:

$$\begin{aligned} isOn(x,y) \subset Scene(s) \equiv \;& \exists z_1\, \exists z_2.\; z_1 = isCube(x) \wedge z_2 = isCube(y) \wedge \\ & z_1 \neq z_2 \wedge Percept(isCube(x), 1, s) \wedge \\ & Percept(isCube(y), 1, s) \wedge EC(z_1, z_2, s). \end{aligned}$$

where EC is a fluent that specifies that two perceptibles are externally connected in situation s, and its definition would be based on a suitable topological axiomatization. On the other hand, if a perceptible is a simple object, e.g. a cube, then a possible description could be

$$\begin{aligned} isCube(x) \subset Scene(s) \equiv \;& \exists r\; \Psi(isCube(x), r, s) \wedge \\ & \exists m\; Gmodel(isCube(x), m) \wedge \\ & Align(isCube(x), m). \end{aligned}$$

$\Psi(isCube(x), r, s)$ is a sentence that accounts for the process undergone by the vision system in which the features of the cube have been systematized in vertices and edges of a specific region, or set of points, and the correspondence with the perceptible is established via the alignment between the perceptible and the model. Observe that since there might be many primitive elements corresponding to $isCube$ then Ψ would include an anchoring (see [24, 25]) that would determine the specific one under examination. The concept of *Gmodel* is to provide a representation for the category or model of a primitive component. This concept is easy to define for example for regular polyhedra but quite hard for example for complex shapes as Biederman' s geons. Moreover methods like the orthographic projection from an image, as in [83], or prototyping, as in [6], cannot be used in a straightforward way to construct models when primitive components are considered abstract categories. Many theories of shape spaces have been proposed for representing models (for a detailed discussion that takes also into account the problem of similarity between different models see [30]), we refer to them for a possible choice of representation. Finally the predicate *align* accounts for the process that, after suitable rigid transformations – scaling rotation, translation etc. – would match the model.

The functional fluent $Scene(s)$ and the fluent $Occluded(p, s)$, with p a perceptible, in order to be introduced would need a clear notion of anchoring a denotation of an element with its representation in the image, and of what a dynamic transformation of a scene is. In fact we expect that a perceptible p, in a given situation, is occluded by another perceptible p' when some essential parts of p' would overlap p, however in a previous situation this fact might not hold. However, the transformation of the image, even if it is induced by a control action of an agent, has to be accounted as a shifting of a shape in the image. In other words, a specific treatment of which kind of transformations can occur in the image of the scene and how they are connected with both control and sensing actions will be required.

3 Direct Perception

In this section we introduce the basic formalization for a theory of action and perception combined with the notion of perception, that relies on non deterministic sensing actions, and a basic theory of actions, based on deterministic successor state axioms. The basic formalization serves to define direct perception.

Preconditions axioms for control actions

$$Poss(A(\vec{x}), s) \equiv \Pi_A(\vec{x}, s) \tag{4}$$

Here $\Pi_A(\vec{x}, s)$ is a formula whose free variables are only $\vec{x}$ and s, does not mention other situations, nor *do*, *Poss* nor $\sqsubset$, and it might contain the fluent *Percept*. For example:

$$\begin{aligned} Poss(grasp(x), s) \equiv \exists x_0\ y_0\ x_1\ y_1.&Position(x, x_0, y_0, s) \wedge \\ &Arm(x_1, y_1, s) \wedge dist(x_0 - x_1, y_0 - y_1) < 10. \end{aligned}$$

Preconditions axioms for sensing actions For all perceptibles p:

$$Poss(sense(p, pr), s) \equiv \Pi_{sense(p,pr)}(x, a, s) \tag{5}$$

Observe that a sensing action is to be interpreted as a query to the vision system. Therefore we are assuming that the raw data acknowledged by the vision system, have already been elaborated to a sufficiently high level of details that only excludes the recognition of the objects in the image. The sensing actions asks to the vision system whether a certain fact holds in the image according to the compositional laws formulated in the description of the perceptibles themselves. Sensing actions, in turn, are possible if the perceptibles themselves are accessible, i.e. no significant part is occluded.

For example:

$$\begin{aligned} Poss(sense(isCube(x), s) \equiv\ &isCube(x) \subset Scene(s) \wedge \\ &\neg Occluded(isCube(x), s). \end{aligned}$$

Where the description of perceptible, i.e. $isCube(x) \subset Scene(s)$, the fluent *Occluded* and *Scene* have been discussed in Section 2.3.

Successor state axioms for Fluents For any fluent except for *Percept*, the normal form for the successor state axiom is, as in [67], given by the following schema:

$$F(do(a, s)) \equiv \gamma_F^+(a, s) \vee (F(s) \wedge \neg\gamma_F^-(a, s)) \tag{6}$$

We maintain the idea that Fluents, interpreting the changes in the world are only caused by control actions and other truths about the world, independently from the mental state of an agent. Observable Fluents and their dynamic, described in the successor state axioms, are taken as analytic truths, as opposed to accidental truths drawn by perception, as we argued in the Introduction.

On the other hand, the Fluents for perception are affected only by sensing actions and previous percepts, as we are trying to model *direct perception* through the fluent *Percept*. The reason can be illustrated by the following example.

Example 2 Consider a computer game like FIFA2000, a classic for soccer games. There are, indeed two levels of control actions: those executed by the player playing the game and corresponding to pushing the arrow keys and space, if no joystick is available, and those corresponding to the effect of pushing the arrow keys. That is, those executed by the soccer players in the game. The arrow keys cause, in some order, the actions *jump, turn left, turn right* and *kick*, the space corresponds to *run.* If we suppose a single agent is playing FIFA200, we can forget the distinction between pushing the arrow keys and the actions they induce and suppose that the actions are just *jump, turn left, turn right* and *kick.* The decision to kick the ball or run or turn are left to the player while the consequences of the actions are implemented in the game. So we can imagine that the position of the ball after a kick action is determined by physical laws that take into account the direction of the kick and so on. These laws can be compiled into suitable contexts of a successor state axiom for the fluent $PositionBall$. On the other hand, the perception of the position of the ball depends from specific conditions like whether the player is looking at the screen, whether the ball is or not occluded in the image, the color of the ball and the color available on the screen and so on. All these conditions are obviously independent of the control actions. In other words, despite an autonomous agent playing FIFA200 would decide what to do on the basis of her perception of the scene, at each specific situation, her decision is independent from the laws that represent the position of the ball or the perception of where the ball is in the screen.

□

The upshot of the above example is that successor state axioms for Percept or observable Fluents draw general, natural, physical and commonsense laws about changes in mental or real worlds states, therefore they are reciprocally independent.

□

For $Percept$ we have the following schema, for each n-ary perceptible term isF, which is obtained by closing the causal laws for $Percept$ together with the consistency conditions (2) and the completeness assumption:

$$\begin{aligned} Percept(isF(\vec{x}), pr, do(a,s)) \equiv \quad & (pr = 1 \wedge (\delta^{+}_{isF}(a,s) \vee \\ & (\neg\delta^{-}_{isF}(a,s) \wedge Percept(isF(\vec{x}),1,s)))\ ) \vee \\ & (pr = 0 \wedge (\delta^{-}_{isF}(a,s) \vee \\ & (\neg\delta^{+}_{isF}(a,s) \wedge Percept(isF(\vec{x}),0,s)))\ ) \end{aligned} \tag{7}$$

The above successor state axiom states that the perception of a perceptible– an object, an its part, or a fact – through *direct sense perception* is achieved by directly sensing elements of a scene according to the compositional properties and the causal connections defining the structure of the objects and the scene in their appearance (see Section 2.3). Observe that the syntax of the axiom is just a suitable expansion of the normal form successor state axioms for Fluents (6), by taking care of the nondeterministic outcome of sensing actions.

Example 3 Consider the following domain axioms:

$$\begin{aligned} Holding(x,y,do(a,s)) \quad & \equiv a = grasp(x,y) \wedge Cube(x,s) \wedge Arm(y) \vee \\ & (a \neq drop(x,y) \wedge Holding(x,y,s)) \end{aligned} \tag{8}$$

$$
\begin{aligned}
Percept(isHolding(x,y), pr, do(a,s)) \equiv & \ (\ pr = 1 \wedge \\
& (a = sense(isHolding(x,y),1) \vee \\
& (a \neq sense(isHolding(x,y),0) \wedge \\
& Percept(isHolding(x,y),1,s)) \vee \\
& (\ pr = 0 \ \wedge \\
& (a = sense(isHolding(x,y),0) \vee \\
& (\ a \neq sense(isHolding(x,y),1) \wedge \\
& Percept(isHolding(x,y),0,s) \ )
\end{aligned} \tag{9}
$$

The above axiom (9) says that an agent y perceives that she is holding an object x iff either of the following are verified: (a) she performs the action $sense(isHolding(x,y),1)$ in the situation s; (b) she was already perceiving that she was holding the object and she is performing some action that is not concerned with sensing whether she is holding or not the object. For the perception that she is not holding the object the circumstances are similar, but reversed, as stipulated by the axiom for *Percept*.

For a coherent behaviour of *Percept* we require that in the initial situation S_0, the following holds:

$$
\forall p \neg(Percept(p,1,S_0) \wedge Percept(p,0,S_0)) \tag{10}
$$

Lemma 1 Conditions (2) together with (10) imply:

$$
\forall p \ s. \neg(Percept(p,1,s) \wedge Percept(p,0,s))
$$

Proof. By induction. For S_0 we have it by definition. Let $s = do(a', s')$, $S_0 \sqsubseteq s'$, and suppose that for some perceptible p, $Percept(p,1,s) \wedge Percept(p,0,s)$. It follows, by the successor state axiom for *Percept* that the following will hold:

$$
\begin{aligned}
&(\delta_F^+(a',s') \vee (\neg\delta_F^-(a',s') \wedge Percept(isF(\vec{x}),1,s'))) \wedge \\
&(\delta_F^-(a',s') \vee (\neg\delta_F^+(a',s') \wedge Percept(isF(\vec{x}),0,s')))
\end{aligned}
$$

By the consistency conditions and the induction hypothesis, i.e., $\neg(Percept(p,1,s') \wedge Percept(p,0,s'))$, follows a contradiction.

□

By the above lemma we get that:

$$
\forall (Percept(p,1,s) \rightarrow \neg Percept(p,0,s))
$$

$$
\forall (Percept(p,0,s) \rightarrow \neg Percept(p,1,s))
$$

But we do not get $\neg Percept(p,1,s) \rightarrow Percept(p,0,s)$ nor $\neg Percept(p,0,s) \rightarrow Percept(p,1,s)$. Therefore the negation of a percept does not necessarily coincide with the perception that a perceptible does not hold. In some sense, the value of $pr \in \{0,1\}$ is a weak truth value because it depends on the vision system, as opposed to the truth value of *Percept* which depends on the semantics.

The inference mechanism of sense perception is based on the hierarchy of perception, which we handle with the successor state axioms for *Percept*. While physical events, caused by control actions might affect the agent's mental events, mental events do not affect physical events, as the causal laws governing physical events are distinguished from the perceiving subject, as we have highlighted with Example 2 This is the content of the next lemma.

Lemma 2 For any fluent F, $F \neq Percept$:

$$\mathcal{D} \models \forall\ (a = sense(p, pr) \rightarrow\ \ F(s) \equiv F(do(a, s))\ )$$

Proof. By contradiction. Let F be some n-ary fluent, $a = sense(p, pr)$, for some perceptible p and degree of truth pr and suppose $\mathcal{D} \not\models \forall(\ F(s) \equiv F(do(a, s))\ )$. There is a model $\mathfrak{A}$ of $\mathcal{D}$ and an assignment v s.t. $\mathfrak{A}, v \models F(\vec{x}, s)$ and $\mathfrak{A}, v \models \neg F(\vec{x}, do(a, s)$. By the successor state axiom for F follows that $\mathfrak{A}, v \models \neg\gamma^+(a, s) \wedge \gamma^-(a, s)$. By the completeness assumption $\neg\gamma^+(a, s) \wedge \gamma^-(a, s)$ are all the conditions that make F become false. Since $sense(p, pr)$ cannot be mentioned neither in $\gamma^+(a, s)$ nor in $\gamma^-(a, s)$, by definition, it follows that $a \neq sense(p, pr)$, for any perceptible p and degree of truth pr. A contradiction. Analogous proof holds if we have a model $\mathfrak{A}$ of $\mathcal{D}$ and an assignment v s.t. $\mathfrak{A}, v \models \neg F(\vec{x}, s)$ and $\mathfrak{A}, v \models F(\vec{x}, do(a, s)$.

□

On the other hand, perception can be affected just by sensing actions and other Percepts, independently of physical events. Therefore we have the analogous of Lemma 2 for perceptibles:

Lemma 3 For any perceptible isF:

$$\begin{array}{ll} \mathcal{D} \models & \forall\ (\ a \neq sense(p, pr') \rightarrow \\ & Percept(isF(\vec{x}), pr, s) \equiv Percept(isF(\vec{x}), pr, do(a, s))\ ) \end{array}$$

Proof. Analogous as for Lemma 2, but using the successor state axiom for $Percept$.

□

We get, therefore, two parallel inference processes: the inference that affects the real world and the inference that affects the perception of the real world. Both are logically consistent. Despite the fact the two processes are logically consistent they continuously diverge. In fact, by Lemma 2 we have that whenever the contexts for a fluent to change its truth value is realized, in the next situation the fluent and the perceptible will diverge since the perceptible will maintain its previous degree of truth, by Lemma 3. Clearly, if the initial knowledge base, $\mathcal{D}_{S_0}$ entails some $Percept$ these divergence would be even more problematic because it would not be caused by sensing actions, since in $\mathcal{D}_{S_0}$ no sensing action has yet been performed. A natural choice is to impose that in the initial knowledge base nothing about $Percept$ can be concluded. The next lemma justify this choice.

Lemma 4 Let $\circ pr \in \{0, 1\}$ be a fixed value for pr and $\bullet pr$ its complement.

$$\begin{array}{lll} (1) & \mathcal{D} \models & \forall s\ \forall\vec{x}\forall pr.Percept(isF(\vec{x}), pr, s) \wedge \\ & & \quad \forall s''\ s'a'(s'' \sqsubseteq s \wedge s'' = do(a', s') \rightarrow \\ & & \qquad a' \neq sense(isF(\vec{t}), pr)) \rightarrow Percept(isF(\vec{t}), pr, S_0) \\ (2) & \mathcal{D} \models & \forall s\ \forall\vec{x}.\neg Percept(isF(\vec{x}), \circ pr\ , s) \wedge \\ & & \quad \forall s''\ s'a'(s'' \sqsubseteq s \wedge s'' = do(a', s') \rightarrow \\ & & \qquad a' \neq sense(isF(\vec{x}), \bullet pr\ )) \rightarrow \neg Percept(isF(\vec{x}), \circ pr\ , S_0) \end{array}$$

Proof. We show the proof by induction on s. For $s = S_0$ the claim holds trivially. Let $s = do(a, s')$, $s' \sqsupseteq S_0$. We show the proof for $\circ pr = 1$ as $\circ pr = 0$ is symmetric. Let $\mathfrak{A}$ be a model of $\mathcal{D}$ and v an assignment.
(1) Suppose:

$$\mathfrak{A}, v \models Percept(isF(\vec{x}), 1, s) \wedge \forall s''\ s'(s'' \sqsubseteq s \rightarrow s'' \neq do(sense(isF(\vec{x}), 1), s')).$$

By the successor state axiom for *Percept*, $\mathfrak{A}, v \models \delta^+_{isF}(a, s') \vee (\neg\delta^-_{isF}(a, s') \wedge Percept(isF(\vec{x}), 1, s'))$. Since by hypothesis $a \neq sense(isF(\vec{x}), 1)$ then $\mathfrak{A}, v \not\models \delta^+_{isF}(a, s')$, hence $\mathfrak{A}, v \models \neg\delta^-_{isF}(a, s') \wedge Percept(isF(\vec{x}), 1, s')$ and by induction hypothesis the claim follows.
(2) Suppose:

$$\mathfrak{A}, v \models \neg Percept(isF(\vec{x}), 1, s) \wedge \forall s''\ s'(s'' \sqsubseteq s \rightarrow s'' \neq do(sense(isF(\vec{x}), 0), s')).$$

By the successor state axiom for *Percept*, then:

$$\mathfrak{A}, v \models \neg\delta^+_{isF}(a, s') \wedge (\delta^-_{isF}(a, s') \vee \neg Percept(isF(\vec{x}), 1, s')).$$

Since by hypothesis $a \neq sense(isF(\vec{x}), 0)$ then $\mathfrak{A}, v \not\models \delta^-_{isF}(a, s')$, hence $\mathfrak{A}, v \models \neg\delta^+_{isF}(a, s') \wedge Percept(isF(\vec{x}), 1, s')$ and by induction hypothesis the claim follows.

$\square$

The above lemma says essentially that if *Percept* about a certain perceptible is entailed in the initial knowledge base and no sensing action relative to this specific perceptible occurs then the *Percept* will persist. Which we do not want, because if the agent never opens her eyes there is no reason to entail any direct perception. Therefore we require:

$$\mathcal{D}_{S_0} \not\models \exists p\ pr \pm Percept(p, pr, S_0) \tag{11}$$

The above statement however does not avoid mistakes in any situation after S_0. We introduce therefore the notion of the agent being *Mistaken*, that should trace the different truth-behaviour of a specific fact of the world and of their perception by the agent.

The definition of ***Mistaken*** From the above axiom for *Percept*, the perception of a fact persists in the memory of an agent if no sensing action occurs to modify the memory of her direct perception. When a sensing action occurs different possible outcomes might take place. For example, let σ be a sequence of actions and suppose it is extended with the action $sense(isRed(block5), 1)$. We have different possibilities:

- The sensing action confirms what the agent infers.
  That is, $\mathcal{D} \models Red(block5, \sigma)$ so, by Lemma 2, $\mathcal{D} \models Red(block5, do(sense(isRed(a), 1), \sigma))$;
- The sensing action has no counterpart with the agent's knowledge of the world. That is $\mathcal{D} \not\models Red(block5, \sigma)$, hence $\mathcal{D} \not\models Red(block5, do(sense(isRed(block5), 1), \sigma))$. If as a consequence of the sensing action it follows that $\mathcal{D} \models Percept(isRed(block5), 1, do(sense(isRed(block5), 1), \sigma))$, then the sensing action is truly informative;
- The sensing action is in disagreement with the agent's knowledge of the world. That is, $\mathcal{D} \models \neg Red(block5, \sigma)$ hence, by Lemma 2, $\mathcal{D} \models \neg Red(block5, do(sense(isRed(block5), 1), \sigma)$.

The last item above determines an incoherence between the agent's direct sense perception and her reasoning process. To trace this incoherence we introduce the following definition:

$$\begin{aligned} Mistaken(isF(\vec{x}), s) \stackrel{\text{def}}{\equiv}\ & (\neg F(\vec{x}, s) \wedge Percept(isF(\vec{x}), 1, s)) \vee \\ & (F(\vec{x}, s) \wedge Percept(isF(\vec{x}), 0, s)). \end{aligned} \tag{12}$$

The above definition says that the agent is mistaking a perceptible isF whenever she gets a misalignment between a truth about an observable fluent and her perception of it.

The definition takes a cautious point of view on perception. To illustrate why we give an example that can cause ambiguous or at least contrasting interpretations.

Example 4 Assume the following successor state axioms:

$$\begin{array}{rl} Position(x, x_0, y_0, do(a, s)) \equiv & a = move(x, x_0, y_0) \vee \\ & (\forall\ w\ z.\ a \neq move(x, w, z) \wedge \\ & \quad Position(x, x_0, y_0, s)). \end{array} \tag{13}$$

$$\begin{array}{l} Percept(isPosition(x, x_0, y_0), pr, do(a, s)) \equiv \\ \quad (pr = 1 \wedge (a = sense(isPosition(x, x_0, y_0), pr)) \vee \\ \quad\ (a \neq sense(isPosition(x, x_0, y_0), 0) \wedge \\ \quad\quad \forall\ w\ z. w \neq x_0 \wedge z \neq y_0 \rightarrow a \neq sense(isPosition(x, w, z), 1) \wedge \\ \quad\quad\quad Percept(isPosition(x, x_0, y_0), pr, s) \\ \quad)\ ) \vee \\ \quad (pr = 0 \wedge (a = sense(isPosition(x, x_0, y_0), pr) \vee \\ \quad\quad \exists\ w\ z.\ a = sense(isPosition(x, w, z), 1)\ \wedge w \neq x_0 \wedge z \neq y_0) \vee \\ \quad\ (a \neq sense(isPosition(x, x_0, y_0), 1) \wedge \\ \quad\quad Percept(isPosition(x, x_0, y_0), pr, s) \\ \quad)\ ). \end{array} \tag{14}$$

Let σ be some sequence of actions and suppose $\mathcal{D}$ entails

$$Position(block5, 1, 4, \sigma)$$

Suppose that σ is extended with $sense(isPosition(block5, 3, 4), pr)$, with $pr = 1$. Then $\mathcal{D}$ entails:

$$\begin{array}{l} Percept(isPosition(block5, 3, 4), 1, do(sense(isPosition(block5, 3, 4), 1), \sigma)), \\ \neg Position(block5, 3, 4, do(sense(isPosition(block5, 3, 4), 1), \sigma)), \\ Percept(isPosition(block5, 1, 4), 0, do(sense(isPosition(block5, 3, 4), 1), \sigma)), \\ Position(block5, 1, 4, do(sense(isPosition(block5, 3, 4), 1), \sigma)), \\ Mistaken(isPosition(block5, 3, 4), do(sense(isPosition(block5, 3, 4), 1), \sigma)), \\ Mistaken(isPosition(block5, 1, 4), do(sense(isPosition(block5, 3, 4), 1), \sigma). \end{array}$$

□

If we consider the above example from the point of view of what is entailed by the theory $\mathcal{D}$ then the interpretation depends only on the choice of pr in $sense(isF, pr)$ as an hypothesis which might be or not in agreement with the observable fluent F and, obviously, one of the two choices is always the wrong one. If, on the other hand, we consider the example as *what happened at execution time* then the agent is committed to the choice of pr as it is delivered by the sensory system, and it relies on the outcome of a truth interaction of the agent with the world. In this case the Example well interpret the pro and contra of choosing a cautious approach versus a brave one. The cautious approach, in fact, *prefers* the truth value of the fluent whenever fluent and perceptible diverge. While the brave approach would prefer the *Percept*.

Consider the following two possible interpretations of the above example:

First Interpretation: Block5 is, indeed, at position $(1,4)$ in σ and the outcome of sensing is wrong, so that the *Percept* is mistaken and *Mistaken* correctly interpret the situation, independently from what causes the error.

Second Interpretation: Imagine the following scenario: the computation of the position of the block is correct, i.e. $Position(block5,1,4),1,\sigma)$, but during the execution of σ the arm shifted the position of the blocks along the x-axis and the sensing action $sense(isPosition(block5,3,4),1)$ is correct. In this case *Mistaken* does not capture the real situation.

The above example show two symmetrical interpretations, one in which the agent has a wrong perception of the real world and one in which sensing is correct and the misalignment between perception and observable Fluents is not a mistake. A brave approach would always consider sensing correct, eventually making the outcome of sensing action even change the truth of observable Fluents. A cautious one would consider the derivable truth about the observable Fluents the measure of correctness. Both approaches, in our opinion, are ontologically valid, they would interpret two different philosophical traditions.

The choice of a cautious or brave approach should be, in general, carefully motivated as they should depend on the use one wants to make of the dynamics of the system. If the system is used for off-line planning then the preference accorded to the deterministic axioms for observable Fluents, w.r.t. the non deterministic ones for percepts, is justified. In fact, in off-line reasoning an *ideal behaviour* of actions is undertaken and thus a percept has to be treated as an hypothesis. If the system is used on-line, instead, the representation problem is reversed: sensing has the role of confirming the adherence to the real world context. In this case the reliance on the the vision system and the percepts is auspicable and thus the preference would be accorded to the percepts.

The theory of action and perception should basically account for both the approaches – this the reason why we introduced mistaken as a definition.

In this work we take the cautious approach because it is simpler, and as we shall see, it requires just to add hypotheses to the theory, while the brave approach is much harder as it would require to revise the theory. A brave approach based on revision is taken by Coradeschi in [23] in the framework of *Feature and Fluents* [72]. Observe that in [23] the brave approach is justified by the fact that observations – as opposed to sensing actions – can be contradictory in a given scenario. In our formalization it is easy to see that this is not possible, since perceptibles are terms and because of Lemma 1

For example, see e.g. [27], if exogenous actions would be introduced, like for instance the shifting of the position of the blocks in Example 4, then the entailed $Position(block5,1,4,\sigma)$ could be revised and thus the outcome of the sensing action believed. In fact, with exogenous actions the choice between the cautious and brave approach would depend on the specific case and thus would be less coarse than the one taken here.

Clearly both approaches, the cautious and the brave one, account for non monotonicity. The reason is that the recognition process behind a perceptible is only partially reliable because of occlusion or, more generally, because of the great number of exceptions to be considered that cannot fit into definitions. For example Biedermann in [12] tells us that

> Although a picture of a sparrow is classified as bird rather than a sparrow, a picture of a penguin is classified more quickly as

a penguin than as a bird (Jolicoeur, Gluck and Kosslyn [42]). The same holds true for ostrich, duck and a number of atypical instance of basic-level categories.

And the same can be said about successor state axioms for observable Fluents. So when the two inference processes are combined or are dialectically interleaved then one of the two has to prevail. In other words, the large number of exceptions that enter the recognition and the reasoning processes have to be taken into account somewhere in the formalization of a theory of actions and perception: either in the description of the scene and of the representation model – where we prefer a monotonic descriptions based on definitions – or in the treatment of the outcomes, as hypotheses or revisable observed facts.

Clearly either suitable heuristics or specific strategies or probabilities (see e.g. [3] and also [22]) can help in deciding when it is the case to believe in the outcome of a sensing action or in the truth of an observable fluent.

In the following we shall always assume that if a misalignment between the outcome of a sensing action and the truth of an observable fluent occurs then the outcome, i.e. the hypothesis, is wrong. The following theorem gives to the outcome of sensing actions the status of hypothesis.

Theorem 1 Let F be an $n+1$-ary fluent and isF the n-ary perceptible associated with F. Then

$$\mathcal{D} \models \forall\ (\pm F(s) \wedge \mp F(do(a,s)) \rightarrow Mistaken(isF, s) \vee Mistaken(isF, do(a,s)).$$

Proof. By Lemmata 2 and 3 and the definition of $Mistaken$. $\square$

4 Selective Sense Perception

In the previous section we have seen that, whenever the context of a control action changes the truth value of a fluent, we get a mistake either in the previous or in the current situation, since direct perception is not affected by the execution of a control action. This is not problematic because it is natural that *if the world had changed and the agent cannot see the changes occurred then she cannot have direct perception of them.*

What if the last action to occur is a sensing action? Whenever a sensing action is, so to say, mistaken, the two inference process are not aligned, and the inference mechanism will cause the inability to associate perception of a fact with the cognition of the same fact[1].

We introduce, now, selective sense perception to ensure that the two processes are always aligned, but when sensing is not correct.

Notational Convention We use $\circ pr$, in conjunction with $\pm$, to denote the condition:

$$\circ pr \;=\; \begin{cases} 1, & \text{if } \pm F \text{ is } F \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$\bullet pr$ denotes the reversed above condition, i.e.

$$\bullet pr \;=\; \begin{cases} 1, & \text{if } \pm F \text{ is } \neg F \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

[1]In cognitive science this disorder is called *associative agnosia.* Farah [33], tells that agnosia caused a patient to believe that someone extraneous was staring at him intently and, in fact, the patient had been looking at himself in a mirror.

Following the *constructivist* view [29, 32, 51, 60, 70, 71, 77, 82]) we propose selective sense perception as an early inference mechanism that selects for further processing that stimulus information which is relevant for current processing goals. We want, in agreement with a cautious point of view, that sensing actions are performed when necessary, otherwise $SePercept$ chooses the available truth. This is the intended meaning of the following definition.

The definition of Selective Sense Perception The role of selective sense perception is to align the two inference processes, the one that determines the truth value of an observable fluent and the one that determines the truth-degree of a perceptible. After a sensing action selective sense perception chooses direct perception, otherwise when no sensing action occurs, selective sense perception adds to the visual memory the truth information that holds in the current situation. So selective sense perception behaves greedily, by choosing the truth-degree of a perceptible according to the last action executed.

$$\begin{array}{l} SePercept(isF(\vec{x}), pr, s) \stackrel{\text{def}}{\equiv} \\ \qquad (s = S_0 \wedge (pr = 1 \wedge F(\vec{x}, s)) \vee (pr = 0 \wedge \neg F(\vec{x}, s))) \\ \qquad \vee(\exists s'\ a.s = do(a, s') \wedge \\ \qquad\quad (\exists p'\ pr'\ a = sense(p', pr') \wedge Percept(isF(\vec{x}), pr, s)) \vee \\ \qquad\quad (\forall p'\ pr'.a \neq sense(p', pr') \wedge \\ \qquad\qquad (pr = 1 \wedge F(\vec{x}, s)) \vee (pr = 0 \wedge \neg F(\vec{x}, s))). \end{array} \tag{17}$$

□

Observe that in $\mathcal{D}_{S_0}$ the above definition reduces to:

$$SePercept(isF(\vec{x}), pr, S_0) \stackrel{\text{def}}{\equiv} \quad (pr = 1 \wedge F(\vec{x}, S_0)) \vee (pr = 0 \wedge \neg F(\vec{x}, S_0)))$$

The intuition behind the selective sense perception of a perceptible isF is the following.

> If a is a sensing action, take the current value of the direct perception of isF, even if it can be mistaken. Otherwise, if a is a control action, believe in the reasoning inference process that uses analytic truth.

Lemma 5 For any perceptible isF

$$\mathcal{D} \models \forall \neg\ (SePercept(isF(\vec{x}), 1, s) \wedge SePercept(isF(\vec{x}), 0, s))$$

Proof. For $s = S_0$ the statement is a consequence of the definition of $SePerecept$. Let $s \sqsupseteq do(a, s')$, with $s' \sqsupseteq S_0$. Suppose there is a model $\mathfrak{A}$ of $\mathcal{D}$ and an assignment v such that $\mathfrak{A}, v \models SePercept(p, 1, s) \wedge SePercept(p, 0, s)$. We shall consider two cases:

1. a is a sensing action;
2. a is a control action.

1. If a is a sensing action and $\mathfrak{A}, v \models SePercept(isF(\vec{x}), 1, s) \wedge SePercept(isF(\vec{x}), 0, s)$ then:

$$\mathfrak{A}, v \models \quad Percept(isF(\vec{x}), 1, do(a', s)) \wedge Percept(isF(\vec{x}), 0, do(a, s))$$

And we get a contradiction with Lemma 3.

2. If a is a control action then $\mathfrak{A}, v \models SePercept(p, 1, s) \wedge SePercept(p, 0, s)$ together with the definition of $SePercept$ implies:

$$\mathfrak{A}, v \models F(\vec{x}, s) \wedge \neg F(\vec{x}, s)$$

A contradiction.

□

Observe, however that, whenever the last action in σ is not a sensing action it is possible to get the following:

$$SePercept(isF(\vec{t}), 1, \sigma) \wedge Percept(isF(\vec{t}), 0, \sigma)$$

This is a consequence of Theorem 1. To better understand the behaviour of $SePercept$ we state the following.

Lemma 6 Let F be an n-ary observable fluent and isF its associated perceptible. Let $\wp r$ denote the condition stated in (15).
1. For any control action a:

$$\mathcal{D} \models \quad \forall\ (\pm F(\vec{x}, do(a, s)) \rightarrow SePercept(isF(\vec{x}), \wp r, do(a, s)));$$

2.

$$\begin{aligned} \mathcal{D} \models \quad & \forall\ (\pm F(\vec{x}, s) \wedge a = sense(isF\vec{x}, \wp r\ ) \rightarrow \\ & (SePercept(isF(\vec{x}), \wp r, do(a, s)) \rightarrow \neg Mistaken(isF(\vec{x}), do(a, s))); \end{aligned}$$

3. If a is any sensing action:

$$\begin{aligned} \mathcal{D} \models \quad & \forall\ (\pm F(\vec{x}, s) \wedge Percept(isF(\vec{x}), \wp r\ , do(a, s))) \rightarrow \\ & (SePercept(isF(\vec{x}), \wp r\ , do(a, s)) \wedge \neg Mistaken(isF(\vec{x}), do(a, s)))). \end{aligned}$$

Proof.

1. By the definition of $SePercept$.
2. Using Lemma 2, the definition of $SePercept$ and the definition of $Mistaken$
3. By the previous item.

□

The above results give us an intuition about the behaviour of selective sense perception. To better understand its meaning, consider a sequence of control actions $\sigma = [a_1, \ldots, a_k]$[2]. $\mathcal{D}_{S_0} \not\models \exists p\ pr\ Percept(p, pr, S_0)$, because no sensing action occurred in $\mathcal{D}_{S_0}$ and because of (11). It follows that $\mathcal{D} \not\models \exists p\ pr\ Percept(p, pr, \sigma)$, by Lemma 4. On the other hand, by the definition of $SePercept$ we have that, for all observables s.t. $\mathcal{D} \models \vec{Q}\vec{x} \pm F(\vec{x}, \sigma)$, with $\vec{Q}$ a sequence of quantifiers binding the sequence $\vec{x}$:

$$\mathcal{D} \models \vec{Q}\ \vec{x}\ SePercept(isF(\vec{x}, \wp r\ , \sigma)$$

In other words even if an agent has no direct perception of the state of the world after a sequence of control actions $\sigma = [a_1, \ldots, a_k]$ her selective sense perception is such that she updates her selective sense memory with the conclusions drawn by reasoning. Suppose now that the history $[a_1, \ldots, a_k]$ is augmented with a sequence of sensing actions $[a_{k+1}, \ldots, a_{k+m}]$, so that the

[2]For this notation see e.g. [69], it is just a compact way to denote $do(a_k, \ldots, do(a_1, S_0))$

actual situation is σ'. Suppose that the agent's sensing does not contradicts the analytic truths that she holds about the world then for all observables s.t. $\mathcal{D} \models \vec{Q}\vec{x} \pm F(\vec{x}, \sigma)$, with $\vec{Q}$ a sequence of quantifiers binding the sequence $\vec{x}$:

$$\mathcal{D} \models \vec{Q}\ \vec{x}\ SePercept(isF(\vec{x}), \wp r\ , \sigma')$$

In this case we would say that *selective sense perception is* **veridic**.

In general, however we do not need veridic perception, but something weaker. Consider the following example.

Example 5 Ben enters in a room, with the eyes covered. In the room all his friends are waiting for him with a cake to celebrate his birthday. The cake is decorated with a big soccer field. Ben has no idea what there is in the room, but as soon as he takes away the mask he immediately recognizes his best friends, the cake and the wonderful decoration. Whatever he perceives about the cake decoration cannot be match against any prior knowledge he has, as it was a surprise.

In the above example the surprise is paradigmatic for those contexts in which the lack of prior knowledge about a given fact makes perception the only mean to convey information from the environment to the agent. In this case we would say that sense perception is *weakly veridic.*

Weakly veridic sense perception can be characterized only semantically as follows. Weakly veridic sense perception holds when there exists at least one model $\mathfrak{A}$ of $\mathcal{D}$ and an assignment v such that, whenever a is a sensing action, for the perceptible isF and its associated observable:

$$\mathfrak{A}, v \models SePercept(isF(\vec{x}), \wp r\ , do(a, s)) \rightarrow\ \pm F(\vec{x}, do(a, s)) \tag{18}$$

The above statement, which is obviously true when a is a control action, says that if there is a model of $\mathcal{D}$ in which selective perception implies the associated observable, then it is possible to believe in the outcome of selective sense perception. Weakly veridic sense perception requires that a misalignment between perceptibles and observable Fluents cannot be drawn. This implies that when weakly veridic sense perception is satisfied then $Mistaken$, for such a specific perceptible, is not entailed by the theory. The difference, with respect to veridic sense perception, between $\mathcal{D} \not\models Mistaken(isF(\vec{t}), \sigma)$ and $\mathcal{D} \models \neg Mistaken(isF(\vec{t}), \sigma)$ is illustrated in the following example.

Example 6 Let σ be a sequence of actions, suppose that $\mathcal{D} \models \neg\ Mistaken(isF, \sigma)$ and let $\pm = +$ and $\wp r\ = 1$. Consider two models of $\mathcal{D}$, $\mathfrak{A}$ and $\mathfrak{B}$ s.t. $\mathfrak{A} \models F(\sigma)$ and $\mathfrak{B} \models \neg F(\sigma)$. Since both satisfy $\neg\ Mistaken(isF, \sigma)$ then we have:

1. $\mathfrak{A} \models F(\sigma) \wedge \neg Percept(isF, 0, \sigma)$
2. $\mathfrak{B} \models \neg F(\sigma) \wedge \neg Percept(isF, 1, \sigma)$

Let $a = sense(isF, 1)$ and suppose that $\delta^{+}_{isF}(a, \sigma)$ is satisfied in both models, then we will have

1. $\mathfrak{A} \models F(do(a, \sigma)) \wedge Percept(isF, 1, do(a, \sigma))$
2. $\mathfrak{B} \models \neg F(do(a\sigma)) \wedge Percept(isF, 1, do(a, \sigma))$

Therefore $\neg Mistaken(isF, do(a,\sigma))$ is no more entailed by $\mathcal{D}$, after the sensing action. In this case selective sense perception is weakly veridic but not veridic.

$\square$

The difference between veridic and weakly veridic sense perception can be exemplified by the two following sentences:

Veridic perception: "This is a piece of marble and I can both prove it and perceive it". In this case analytic and accidental truth coincide.

Weakly veridic perception: "I have no idea about what this material is, I see that is marble" In this case accidental truth is the only available truth, the lack of misalignment is due only to the lack of prior knowledge.

Senses are obviously useful in the case of weakly veridic perception. How do we make use of them, well because, in general, we take for good what we perceive unless there is some contradiction. A cautious approach would consider the contradiction from the side of analytic truth, a brave approach would take the contradiction from the side of accidental truth.

The dynamic or the dialectic between perception and analytic truth is acknowledged in the notion of *meaningful perception* that we are going to introduce in the next section.

5 Meaningful Perception

As we have argued in the previous section, selective sense perception behaves greedily in the sense that it chooses the truth from the current situation whether it is determined by a control or a sensing action. In the case of a sensing action it could fall into a mistake. We called that form of sense perception *selective* because it aligns the two parallel reasoning processes simply by selecting what is drawn by one or the other, depending on the last action occurred. We have also shown that weakly veridic sense perception can be ensured whenever $Mistaken$ is not entailed. In this sense selective sense perception is an early stage of the whole perceptual inference. What we want as a late perception inference is the following:

> Sense!, if the outcome of sensing does not induce a contradiction with analytic truth, that is, no mistake can be drawn with observable Fluents, then add to analytic truth what has been sensed, in the form of hypotheses.

This behaviour just requires that no mistake is derived in the history of actions. The late reasoning process turns out to be similar to hypothetical or abductive reasoning because it exploits hypotheses that are consistently added to the initial database $\mathcal{D}_{S_0}$. These hypotheses, besides consistency, have to respect a parsimonious principle, add the least hypotheses or the more *normal*. However, differently from other forms of commonsense reasoning, these hypotheses can be looked for in the history of actions performed, as are selected by direct sense perception.
To make the foregoing ideas more precise we introduce some definitions that will be used to define what meaningful perception is about.

Definition 1 (Sensing history) Let $\sigma = [\alpha_1, \ldots, \alpha_k]$ and $isP_1(\vec{t_1}), \ldots, isP_m(\vec{t_m})$ be all the perceptibles mentioned in σ. We denote with $isP^{\propto pr}_{j,\sigma_m}(\vec{t_j})$ the perceptible s.t.

$$\sigma \sqsupseteq \sigma_m = do(sense(isP_j(\vec{t_j}), \propto pr\ , \sigma_{m-1})$$

(where $\propto pr\ \in \{0,1\}$ and its meaning, in accordance with $\pm$, is the one given in (15)).

The list $isP^{\propto pr}_{1,\sigma_k}(\vec{t_1}), \ldots, isP^{\propto pr}_{m,\sigma_p}(\vec{t_m})$, with $\sigma_k \sqsubset \sigma_p$, which we abbreviate with $\sigma(\mathbb{H})$ is called the *sensing history* of σ.

□

Definition 2 (Last perceptible) Given two perceptibles $isP^{\propto pr}_{i,\sigma_p}(\vec{t_i})$, $isP^{\propto pr'}_{i,\sigma_m}(\vec{t_i}) \in \sigma(\mathbb{H})$ we say that

$$isP^{\propto pr}_{i,\sigma_p}(\vec{t_i}) \preceq isP^{\propto pr\ '}_{i,\sigma_m}(\vec{t_i}) \quad \text{iff} \quad \sigma_p \sqsubseteq \sigma_m \text{ and they agree on everything but } \propto pr$$

We say that $isP^{\propto pr}_{i,\sigma_m}(\vec{t_i})$ is *last* in $\sigma(\mathbb{H})$ if there is no other perceptible $isP^{\propto pr'}_{i,\sigma_p}(\vec{t_i}) \in \sigma(\mathbb{H})$ such that $isP^{\propto pr}_{i,\sigma_m}(\vec{t_i}) \prec isP^{\propto pr\ '}_{i,\sigma_p}(\vec{t_i})$.

□

For example let $\sigma_p \sqsubseteq \sigma_q$ then

$$isOn^1_{\sigma_p}(block3, block5) \prec isOn^0_{\sigma_m}(block3, block5) \text{ and}$$
$$isOn^1_{\sigma_p}(block3, block5) \not\prec isOn^0_{\sigma_m}(block3, block27)$$

Definition 3 (Safe history) Let $\sigma(\mathbb{H})$ be the sensing history of σ. Let $isP^{\propto pr}_{1,\sigma_k}(\vec{t_1}) \ldots isP^{\propto pr}_{m,\sigma_p}(\vec{t_m})$ be all the perceptible terms mentioned in $\sigma(\mathbb{H})$. We say that σ is *safe from mistakes* if:

- For each observable fluent P_i associated with a perceptible $isP^{\propto pr}_{i,\sigma_k}(\vec{t_i})$ mentioned in $\sigma(\mathbb{H})$ which is *last*:

$$\mathcal{D} \models \pm P_i(\vec{t}, \sigma_k) \rightarrow\ \pm P_i(\vec{t}, \sigma)$$

- For each perceptible $isP^{\propto pr}_{i,\sigma_k}(\vec{t_i})$ mentioned in $\sigma(\mathbb{H})$, $\mathcal{D} \not\models Mistaken(isP_i(\vec{t_i}), \sigma_m)$

□

The above definition says that if a is a sensing action taking place at a given σ_m, over a perceptible $isP_i(\vec{t_i})$ with outcome $\propto pr$ and all the following conditions are satisfied:

1. The fluent P_i, associated with isP_i will not change its truth value after that specific sensing action over $isP_i(\vec{t_i})$ in $\sigma(\mathbb{H})$, if $isP_i(\vec{t_i})$ is last.
2. All sensing actions that take place in $\sigma(\mathbb{H})$ do not lead to a mistake.

Then the sensing history is reliable. This is to say that an agent will have a veridic perception of the real state of the world, as it changes under the effect of her control actions, if she does not close her eyes after performing the control actions. We give an example to better understand the above definition.

Example 7 Let $\mathcal{D}_{S_0} = \{Clear(block1, S_0), Clear(block5, S_0)\}$, let $\mathcal{D}_{ss} \cup \mathcal{D}_{ap}$ be the set of classical successor state axioms and action precondition axioms for the blocks world, together with suitable axioms for *Percept*, whose definition we leave to the reader's imagination. Let σ be the following history:

$$\begin{array}{l}[moveToTable(block1), move(block5, block3), \\ \quad sense(isOn(block5, block3), 1), \\ \qquad move(block1, block5), sense(isOn(block1, block5), 1)]\end{array}$$

The sensing history $\sigma(\mathbb{H})$ is

$$\begin{array}{l}\{isOn^1_{\sigma_1}(block5, block3),\ isOn^1_{\sigma_2}(block1, block5)\}. \\ \text{where:} \\ \sigma_1 \text{ is } [moveToTable(block1), move(block5, block3)] \\ \sigma_2 \text{ is } [moveToTable(block1), move(block5, block3), \\ \qquad sense(isOn(block5, block3), 1), move(block1, block5)]\end{array}$$

All the above perceptibles are last because

$$isOn^1_{[move\cdots]}(block5, block3) \not\preceq isOn^1_{[move\cdots]}(block1, block5)$$

Moreover $\sigma(\mathbb{H})$ is safe because the fluent

$$On(block5, block3, [moveToTable(block1)\cdots, sense(isOn(block5, block3), 1)])$$

will not be changed in the sensing history and

$$\begin{array}{ll}\mathcal{D} \not\models & Mistaken(isOn(block5, block3), [moveToTable(block1), \\ & \qquad move(block5, block3), sense(isOn(block5, block3), 1)]) \\ \mathcal{D} \not\models & Mistaken(isOn(block1, block5), [moveToTable(block1), \\ & \qquad move(block5, block3), sense(isOn(block5, block3), 1), \\ & \qquad\ move(block1, block5), sense(isOn(block1, block5), 1)])\end{array}$$

When σ is safe weakly veridic sense perception holds (see (18)). This is what the next lemma is about.

Lemma 7 If $\sigma(\mathbb{H})$ is a safe history then, for each perceptible isF, mentioned in σ, there exists a model $\mathfrak{A}$ and an assignment v such that, if F is associated with isF:

$$\mathfrak{A}, v \models SePercept(isF(\vec{x}), \circ pr\ , \sigma) \rightarrow\ \pm F(\vec{x}, \sigma)$$

Proof. We show the proof only for $\circ pr\ = 1$ as the symmetric case is analogous. For $\sigma = S_0$ is a consequence of the definition of $SePercept$. Let $\sigma = do(a, \sigma')$, $\sigma' \sqsupseteq S_0$. If a is a control action it follows by definition of $SePercept$. Suppose, now, that $a = sense(p, pr)$, for some p and pr and suppose, by contradiction, that for all model $\mathfrak{B}$ of $\mathcal{D}$ and assignment v $\mathfrak{B}, v \models SePercept(isF(\vec{x}), \circ pr\ , \sigma) \wedge \neg F(\vec{x}, \sigma)$. Let σ_{p_i} be the last situation such that $\sigma_{p_i} = sense(isF(\vec{x}), \circ pr\ )$, then $isF(\vec{x})$ is last in σ_{p_i}, hence, since σ is safe, $\mathcal{D} \not\models Mistaken(isF(\vec{x})\sigma_{p_i})$, therefore there is a model $\mathfrak{C}$, s.t. $\mathfrak{C}, v \models \neg Mistaken(isF(\vec{x}), \sigma_{p_i})$ and thus, by definition of both $SePercept$ and $Mistaken$ it follows that $\mathfrak{C}, v \models F(\vec{x}, \sigma_{p_i})$. Since σ is safe it follows that $\mathfrak{C}, v \models F(\vec{x}, \sigma)$ and we get a contradiction.

$\square$

When a perceptible $isP^{\circ pr}_{i,\sigma_k}$ is non last then there is a perceptible $isP^{\circ pr'}_{i,\sigma_m}$ s.t.

$$isP^{\circ pr}_{i,\sigma_k} \prec isP^{\circ pr\ '}_{i,\sigma_m}$$

For example let σ_q be $do(sense(isOn(block3, block5), 1), \sigma_{q-1})$, suppose that afterwards $block3$ is moved on the table and a new situation could be

$$\sigma_p = do(sense(isOn(block3, block5), 0), do(moveToTable(block3), do \cdots, \sigma_q))$$

therefore the first perceptible is not last, because:

$$isOn^1_{\sigma_q}(block3, block5) \prec isOn^0_{\sigma_p}(block3, block5)$$

The requirement that $\mathcal{D} \not\models Mistaken(isOn(block3, block5), \sigma_q)$ means that

either $\mathcal{D} \not\models \neg On(block3, block5, \sigma_{q-1})$ or $\mathcal{D} \models On(block3, block5, \sigma_{q-1})$.

As a corollary of the above lemma, the following will hold for any perceptible isF

$$\mathcal{D} \models \forall\ x \forall s\ (\neg Mistaken(isF(\vec{x}), s) \rightarrow \exists a \neg Mistaken(isF(\vec{x}), do(a, s))) \tag{19}$$

which implies that there exists always a safe history for any perceptible isF. In other words an attentive sense process does not arrive to a dead end. Whenever veridic perception of a fact holds, such a fact can be safely added to the database, which can be updated. This is stated in the following corollary:

Corollary 1 If σ is safe then for each observable F, such that there is a term $isF_{\sigma_{p_i}}(\vec{t})$ in $\sigma(\mathbb{H})$:

$$\mathcal{D} \cup \{\pm F(\vec{t}, \sigma_{p_i})\} \text{ is consistent}$$

So when an inference process involving sensing is weakly veridic then we can say that the fact sensed can be safely added to the database. Observe however that adding to $\mathcal{D}$ a sentence mentioning σ would make $\mathcal{D}$ loose the properties of a basic theory of actions. This problem can be avoided if, instead of adding $\psi(\sigma)$ to $\mathcal{D}$ we add the regression of $\psi(\sigma)$ to the initial database $\mathcal{D}_{S_0}$. For this purpose we briefly recall the keys ideas of regression. Regression [68, 69, 61] is a theorem-proving mechanism for situation calculus: a regressable formula W, of $\mathcal{L}_{\ sitcalc}$, can be transformed into a formula *uniform in* S_0, which is a formula mentioning only the initial situation S_0 and not mentioning $\sqsubseteq$, nor $Poss$ nor do. We recall here that a formula W mentioning situations σ of the form $do(\alpha_k, \ldots, do(\alpha_1, S_0))$, where each α_i is of the form $A_i(\vec{t_i})$ and W does not quantify over situations is *regressable*: there is a regression operator $\mathcal{R}$ which applied to the regressable formula W will transform it into a formula $\mathcal{R}\ (W)$, uniform in S_0, such that:

$$\mathcal{D} \models (\forall)(W \equiv \mathcal{R}\ (W)) \tag{20}$$

where $\forall$ quantifies over all the free variables. The above (20) holds also for a basic theory of action and perception.

Now we can rephrase the statement of Corollary 1 as follows. If σ is safe then for each observable F for which there is a each perceptible $isF_{\sigma_{p_i}}(\vec{t})$ in $\sigma(\mathbb{H})$:

$$\mathcal{D}_{S_0} \cup \mathcal{D}_{una} \cup \{\mathcal{R}\ (\pm F(\vec{t}, \sigma_{p_i}))\}$$

Is consistent.

Definition 4 We say that a formula ϕ of $\mathcal{L}_{\ sitcalc}$ is in *observable form* if it does not mention either $Percept$ or $SePercept$. It is in *meaningful form* if it does not mention $Percept$.

Example 8 The following sentence is in meaningful form:

$$\begin{aligned}\exists y \quad & (On(y, block5, do(move(block5, y), S_0)) \vee \\ & \exists pr\, SePercept(isCube(block5), pr, \\ & \quad do(sense(isCube(block5), pr), do((move(block5, y), S_0))))).\end{aligned}$$

The following are not in meaningful form:

$$\begin{aligned}& On(block5, block3, S_0) \wedge Good(sugar). \\ & \exists pr(Percept(isCube(block5), pr, do(sense(isCube(block5), pr), \\ & \quad do((move(block5, y), S_0)))) \wedge \\ & SePercept(isCube(block5), pr, do(sense(isCube(block5), pr), \\ & \quad do((move(block5, y), S_0)))).\end{aligned}$$

Any sentence of the SC without perception is in observable form.

We do not consider a sentence that mentions *Percept* in meaningful form because *Percept* accounts for direct perception which, as far as it is not aligned with reasoning via *SePercept*, it accounts for changes in the real world only after sensing actions. It is straightforward, however, to transform any sentence mentioning *Percept* into a sentence in meaningful form, using Lemma 6 and the definition of *SePercept*, extending all the situations mentioned with a suitable sensing action.

A sentence in meaningful form is a sentence that records those percepts mediated by *SePercept*. Such a sentence hides implicit hypotheses, namely the facts that have been perceived.

Example 9 Consider the following sentences:

$$Dead(Mario, \sigma) \text{ and } SePercept(isDead(Mario), 1, \sigma)$$

What is the difference between the two? The first asserts a fact, which might be true or false. The second is a mental state, ground on perception, about what can be true of the world – e.g. Mario lays down on the floor in a sea of blood. The second sentence, in fact, hides an implicit hypothesis: something happened in the real world, e.g. Mario was killed, that caused what is currently perceived. Now, suppose that $\mathcal{D} \models SePercept(isDead(Mario), 1, \sigma)$, how can we ensure that the conclusion that *Mario is dead!* can be drawn. If by adding suitable hypotheses, that cannot be rejected (i.e. they are consistent with the initial database), we can conclude that *Mario is dead*:

$$\mathcal{D} \cup \{\text{hypotheses}\} \models Dead(Mario, \sigma)$$

then the perception that Mario is dead, should be considered meaningful. However, because, the information from the real world is acquired by perception, these hypotheses have to be obtained by the relevant percepts gathered with sensing actions that have taken place in the history of σ.

The results presented in the following will give us a way to make explicit the implicit hypotheses hidden in a sentence in meaningful form.

The next definition will define how to translate a sentence in meaningful form into a sentence in observable form.

Definition 5 Let ϕ be a sentence in meaningful form. We define the translation of ϕ into a new formula ψ in observable form inductively, as follows:

1. if ψ is in observable form then $tr(\psi)$ is ψ;

2. if ψ is an atom of the form $SePercept(isF(\vec{t}), \circ pr\ , s)$ then $tr(\psi)$ is $\pm F(\vec{t}, s)$ with $\circ pr$ and $\pm$ defined according to the convention (15);
3. if ψ is $\neg\theta$ then $tr(\phi)$ is $\neg\ tr(\theta)$;
4. is ψ is $\psi_1 \vee \psi_2$ then $tr(\phi)$ is $tr(\psi_1) \vee tr(\psi_2)$;
5. if ψ is $\exists x\ \theta$ or $\exists s\ \theta$, then $tr(\psi)$ is $\exists x\ tr(\theta)$ or $\exists s\ tr(\theta)$;
6. if ψ is $\exists pr\ \theta$ then $tr(\psi)$ is $tr(\theta(pr/0)) \vee tr(\theta(pr/1))$.

Lemma 8 Let $\phi(\sigma)$ be a sentence in meaningful form, s.t. the only situation mentioned is σ and let $\psi(\sigma)$ be the sentence obtained by translation, according to Definition 5, from $\phi(\sigma)$. If σ does not mention any sensing action then:

$$\mathcal{D} \models \phi(\sigma) \text{ implies } \mathcal{D} \models \psi(\sigma)$$

Proof. By definition of $SePercept$ if $a \neq sense(p, pr)$ then:

$$SePercept(isF(\vec{x}), \circ pr\ , do(a, s)) \stackrel{def}{\equiv} \pm F(\vec{x}, do(a, s))$$

The claim follows by induction on $\phi(\sigma)$, using Definition 5. $\square$

Definition 6 (Meaningful perception) Let $\phi(\sigma)$ be a sentence in meaningful form mentioning only σ and s.t $\mathcal{D} \models \phi(\sigma)$.
Let $\sigma(\mathbb{H}) = \{isF^{\circ pr}_{1,\sigma_{p_1}}(\vec{t_1}), \ldots, isF^{\circ pr}_{m,\sigma_{p_m}}(\vec{t_m})\}$, and σ safe. Let also $\psi(\sigma)$ be a sentence in observable form obtained from $\phi(\sigma)$ according to Definition 5.

$\mathcal{D} \approx\!\!\!\models \psi(\sigma)$ means that for each perceptible $isF^{\circ pr}_{i,\sigma_{p_i}}(\vec{t_i})$, which is last, there is an observable $\pm F_i(\vec{t}, \sigma_{p_i})$, associated with it, such that:

$$\mathcal{D} \cup \{\bigwedge \pm F_i(\vec{t_i}, \sigma_{p_i})\} \models \psi(\sigma)$$

We call $\mathcal{D} \approx\!\!\!\models \psi(\sigma)$ the **meaningful perception** of ϕ in σ.

$\square$

As we noted earlier, in order to maintain the properties of a basic theory of action and perception we would add the hypotheses to the initial database and thus meaningful perception of a sentence in meaningful form, is intended to be:

$$\mathcal{D}_{S_0} \cup \{\mathcal{R}\ (\bigwedge \pm F_i(\vec{t_i}, \sigma_{p_i})\} \cup \mathcal{D}_{una} \cup \mathcal{D}_{ss} \cup \mathcal{D}_{ds} \cup \mathcal{D}_{ap} \cup \Sigma \models \psi(\sigma)$$

The following theorem will tell us under which conditions meaningful perception is ensured.

Theorem 2 Let $\phi(\sigma)$ be a sentence uniform in σ and in meaningful form. Suppose also the following:

i. $\mathcal{D} \models \phi(\sigma)$;

ii. $\sigma(\mathbb{H}) = \{isF^{\circ pr}_{1,\sigma_{p_1}}(\vec{t_1}), \ldots, isF^{\circ pr}_{m,\sigma_{p_m}}(\vec{t_m})\}$, σ safe ;

iii. For each $isF^{\circ pr}_{i,\sigma_{p_i}}(\vec{t_i}) \in \sigma(\mathbb{H})$, which is last, $\mathcal{D} \models Percept(isF(\vec{t_i}), \circ pr\ , \sigma_{p_i})$.

iv. $\psi(\sigma)$ is in observable form and $\psi(\sigma) = tr(\phi(\sigma))$.

Let

$$\mathcal{V} = \{\mathfrak{A} \mid \mathfrak{A} \text{ is a model of } \mathcal{D} \text{ and } \mathfrak{A} \models \neg Mistaken(isF_i(\vec{t_i}), \sigma_{p_i}), \text{ for each } isF^{\varphi r}_{i,\sigma_{p_i}}(\vec{t_i}) \in \sigma(\mathbb{H}) \text{ which is last}\}$$

If $\mathcal{V}$ is not empty then:

1. $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge \pm F_i(\vec{t_i}), \sigma_{p_i})\}$ is consistent.
 Where each $\pm F_i(\vec{t_i}, \sigma_{p_i})$ is associated with a $isF^{\varphi r}_{i,\sigma_{p_i}}(\vec{t_i}) \in \sigma(\mathbb{H})$ which is last.
2. For any model $\mathfrak{B}$ of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge \pm F_i(\vec{t_i}, \sigma_{p_i}))$:
 $\mathfrak{B} \models \psi(\sigma)$ iff $\psi(\sigma)$ is true in all models in $\mathcal{V}$.

Proof. First observe that if σ does not mention any sensing action then, by Lemma 8, we have $\mathcal{D} \models \phi(\sigma)$ iff $\mathcal{D} \models tr(\phi(\sigma))$ and since $tr(\phi(\sigma)) = \psi(\sigma)$ both 1. and 2 hold. Therefore suppose that σ mentions some sensing actions and that (i.) through (iv.) above hold.

1. Let $\mathcal{V}$ be not empty. Then there exists a model $\mathfrak{A}$ of $\mathcal{D}$ such that $\mathfrak{A} \models \bigwedge_i \neg Mistaken(isF_i(\vec{t_i}), \sigma_{p_i})$ for each $isF^{\varphi r}_{i,\sigma_{p_i}}(\vec{t_i}) \in \sigma(\mathbb{H})$, which is last. As by hypothesis $\mathcal{D} \models Percept(isF_i(\vec{t_i}), \varphi r\ , \sigma_{p_i})$, for each $isF^{\varphi r}_{i,\sigma_{p_i}}(\vec{t_i}) \in \sigma(\mathbb{H})$, which is last, it follow by the definition of $Mistaken$ that $\mathfrak{A} \models \bigwedge_i \pm F(\vec{t_i}, \sigma_{p_i})$. Since by the regression theorem [61] $\mathcal{D} \models \forall(\mathcal{R}\ (\phi(\sigma)) \equiv \phi(\sigma))$ then $\mathfrak{A} \models \mathcal{R}\ (\bigwedge_i \pm F(\vec{t_i}, \sigma_{p_i}))$. Then $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge \pm F_i(\vec{t_i}), \sigma_{p_i})\}$ is consistent.

2. ($\Rightarrow$). It is straightforward since $\mathcal{V}$ is a subset of the class of models of $\mathcal{D}$.

($\Leftarrow$). Let $\mathcal{V}$ be not empty and for any $\mathfrak{B} \in \mathcal{V}$, $\mathfrak{B} \models \psi(\sigma)$. From (1) above we have that any model in $\mathcal{V}$ is a model of $\mathcal{D}$ and is also a model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$.

We now show that if $\mathfrak{A}$ is any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$, where $\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i})$ conjoins all the Fluents associated with all the terms $isF^{\varphi r}_{\sigma_p}(\vec{t}) \in \sigma(\mathbb{H})$ which are last, $\mathfrak{A} \models \psi(\sigma)$. Observe that by (1), such a model exists. The proof is by induction on σ and the structure of $\phi(\sigma)$ using the translation defined in Definition 5.
If $\sigma = S_0$ then the claim follows by Lemma 8.
Let $\sigma = do(a, \sigma')$, $\sigma' \sqsupseteq S_0$.

1. Let $\phi(\sigma)$ be $\pm F(\vec{t}, \sigma)$ then $\psi(\sigma) \equiv \phi(\sigma)$ and the claim follows from (i.).
2. Let $\phi(\sigma)$ be $SePercept(isF(\vec{t}), \varphi r, \sigma)$ and suppose $\varphi r = 1$, the symmetric case is similar. By the definition of $SePercept$, either $\mathcal{D} \models F(\vec{t}, \sigma)$ or $\mathcal{D} \models Percept(isF(\vec{t}), 1, \sigma)$. In the first case the claim obviously holds, since $tr(\phi(\sigma))$ is $F(\vec{t}, \sigma)$. In the second case we have to consider the following possibilities, for $\sigma_p \sqsubseteq \sigma$.

   (a) $isF^1_{\sigma_p}(\vec{t}) \notin \sigma(\mathbb{H})$;

   (b) $isF^1_{\sigma_p}(\vec{t}) \in \sigma(\mathbb{H})$;

   Consider (a) then by Lemma 4, it must be that
   $\mathcal{D}_{S_0} \models Percept(isF(\vec{t}), 1, S_0)$ contradicting thus (11).
   Consider (b) and let σ_q be the last situation mentioning $isF^1(\vec{t})$ so that $isF^1_{\sigma_q}(\vec{t})$ is last. Since $isF^1_{\sigma_q}(\vec{t})$ is last then $\mathcal{D} \models F(\vec{t}, \sigma_q) \rightarrow F(\vec{t}, \sigma)$. Let $\mathfrak{A}$ be any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$, in particular $\mathfrak{A} \models \mathcal{R}\ (F(\vec{t}, \sigma_q))$, because $isF^1_{\sigma_p}(\vec{t}) \in \sigma(\mathbb{H})$ is last, therefore $\mathfrak{A} \models F(\vec{t}, \sigma)$.

Finally, as by Definition 5, $F(\vec{t}, \sigma)$ is $tr(SePercept(isF(\vec{t}), 1, \sigma))$, then $\mathfrak{A} \models \psi(\sigma)$.

3. $\phi(\sigma)$ is $\neg SePercept(isF(\vec{t}), \wp r, \sigma)$. As above let us suppose $\wp r = 1$. By the definition of $SePercept$ it must be that $\mathcal{D} \models \neg Percept(isF(\vec{t}), 1, \sigma) \wedge \exists s'\ p'\ pr'.a = sense(p', pr') \wedge \sigma = do(a, s')$. Consider the two cases, for $\sigma_p \sqsubseteq \sigma$:

   (a) $isF^0_{\sigma_p}(\vec{t}) \notin \sigma(\mathbb{H})$;

   (b) $isF^0_{\sigma_p}(\vec{t}) \in \sigma(\mathbb{H})$;

   Suppose (a). Then by Lemma 4 it follows that $\mathcal{D} \models \neg Percept(isF(\vec{t}), 1, S_0)$, contradicting thus (11).

   Suppose (b). Let σ_q be the latest situation, w.r.t. σ, s.t. $\mathcal{D} \models (\sigma_q \sqsubseteq \sigma \wedge Percept(isF(\vec{t}), 0, \sigma_q))$. Then $isF^0_{\sigma_q}(\vec{t})$ is last. Let $\mathfrak{A}$ be any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$, $\mathfrak{A}$ is also model of $\mathcal{D} \cup \{\mathcal{R}\ (\neg F(\vec{t_i}, \sigma_q))\}$, because $isF^0_{\sigma_q}(\vec{t})$ is last. And also, because it is last, $\mathcal{D} \models \neg F(\vec{t}, \sigma_{p_i}) \rightarrow \neg F(\vec{t}, \sigma)$. It follows that $\mathfrak{A} \models \neg F(\vec{t}, \sigma)$.
   Finally it is easy to verify that
   $tr(\neg SePercept(isF(\vec{t}), 1, \sigma)) = \neg tr(SePercept(isF(\vec{t}), 1, \sigma))$ and
   $tr(SePercept(isF(\vec{t}), 1, \sigma)) = F(\vec{t}, \sigma_{p_i})$ hence $tr(\phi(\sigma)) = \neg F(\vec{t}, \sigma)$.

4. $\phi(\sigma)$ is $\phi_1(\sigma) \wedge \phi_2(\sigma)$. $\psi(\sigma) = tr(\phi(\sigma)) = tr(\phi_1(\sigma)) \wedge tr(\phi_2(\sigma))$. Since $\mathcal{D} \models \phi_i(\sigma)$, by inductive hypothesis if $\mathfrak{C}$ is any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$ then $\mathfrak{C} \models \psi_1(\sigma)$ and if $\mathfrak{C}'$ is any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$ then $\mathfrak{C}' \models \psi_2(\sigma)$. But since $\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))$ depends only on σ, $\mathfrak{C}$ and $\mathfrak{C}'$ both satisfy $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$ hence both satisfy $tr(\phi_1(\sigma)) \wedge tr(\phi_2(\sigma)) = \psi(\sigma)$.

5. $\phi(\sigma)$ is $\exists x\ \theta(\sigma)$. Then $\psi(\sigma) = tr(\phi(\sigma)) = \exists x\ tr(\theta(\sigma))$. Let $\mathfrak{A}$ be any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$ then, by induction hypothesis, for some assignment v, $\mathfrak{A}, v \models tr(\theta(x, \sigma))$. Therefore $\mathfrak{A}, v \models \exists x\ tr(\theta(\sigma))$, which implies that $\mathfrak{A}, v \models \exists x \psi(\sigma)$.

6. $\phi(\sigma)$ is $\exists pr\ \theta(\sigma)$. Then $\psi(\sigma) = tr(\phi(\sigma)) = tr(\theta(pr/1, \sigma)) \vee tr(\theta(pr/0, \sigma))$. Let $\mathfrak{A}$ be any model of $\mathcal{D} \cup \{\mathcal{R}\ (\bigwedge_i \pm F_i(\vec{t_i}, \sigma_{p_i}))\}$ then, by induction hypothesis, for some assignment v, $\mathfrak{A}, v \models \theta(pr, \sigma)$. As $v(pr) \in \{0, 1\}$ it follows that if $v(pr) = 1$ then $\mathfrak{A}, v \models tr(\theta(1, \sigma))$ hence $\mathfrak{A}, v \models tr(\theta(1, \sigma)) \vee tr(\theta(0, \sigma)))$. On the other hand, if $v(pr) = 0$ then $\mathfrak{A}, v \models tr(\theta(0, \sigma))$ hence $\mathfrak{A}, v \models tr(\theta(1, \sigma)) \vee tr(\theta(0, \sigma)))$. And finally $\mathfrak{A}, v \models \psi(\sigma)$.

□

As a corollary of the above theorem we have

Corollary 2 Let $\mathcal{D} \models \phi(\sigma)$, with $\phi(\sigma)$ in meaningful form, σ safe and the only situation mentioned in ϕ:

$$\mathcal{D} \mathrel{|}\!\approx tr(\phi(\sigma)) \text{ iff } \mathcal{D} \cup \{\bigwedge_i \neg Mistaken(isF_i(\vec{t_i}), \sigma_{p_i})\} \text{ is consistent}$$

Where in $\bigwedge_i \neg Mistaken(isF_i(\vec{t_i}), \sigma_{p_i})$,
$\sigma_{p_i} \sqsubseteq \sigma$ and $\sigma_{p_i} = do(sense(isF_i(\vec{t_i}), \wp r), \sigma_{p_{i-1}})$.

An example will help to understand what meaningful perception is useful for.

6 A simple example for meaningful perception

Example 10 Consider the following scenario. An arm, endowed with wheels is in an unknown desert environment in which there are no obstacles, the ground is reasonably smooth: there are just stones of small dimension. The arm has a trolley attached and its task is to collect only the green stones.

The arm has to look for a close stone, move toward it if it is green pick it up and put it on the trolley.

We give a suitable part of the formalization of the domain, excluding the geometric representation of the primitives, like a stone, and the axiomatization for $Occluded$. We will not write down the definitions of $SePercept$ and $Mistaken$ for each perceptible associated with an observable fluent, as they can be easily figured out by the schemata given in (17) and in (12).

Control Actions

$$moveToward10(x), moveAway10(x), pickUp(x), putInTrolley(x).$$

Where $moveToward10$ is the action of moving the arm toward a stone of ten units. $moveAway10$ is the action of moving away from the position where a stone is, of ten units. $pickUp$ is the action of picking up a stone and $putInTrolley$ is the action of putting a stone in the trolley.

Sensing Actions

$$\begin{array}{l} sense(isGreen(x), pr), sense(isStone(x), pr), \\ sense(isTrolley, pr), sense(isInTrolley(x), pr), \\ sense(isDist(x,z), pr), sense(isHolding(x), pr). \end{array}$$

The above sensing actions implicitly tell us what the perceptibles are, namely:

$$\begin{array}{l} isGreen(x), isStone(x), isTrolley, \\ isInTrolley(x), isDist(x,z), isHolding(x) \end{array}$$

$isGreen(x)$ is the perceptible telling that, in the image, object x has color green. $isStone(x)$ is the perceptible telling that, in the image, object x is a stone. $isTrolley$ is the perceptible telling that, in the image, there is a trolley. $InTrolley(x)$ is the perceptible telling that, in the image, object x is in the trolley. $isDist(x, z)$ is the perceptible telling that, in the image, object x is distance z from the arm, and $isHolding(x)$ is the perceptible telling that, in the image, the arm is holding object x. The above sensing actions, for each of the above mentioned perceptibles, are the actions corresponding to a request to the vision system about each perceptible: does the perceptible corresponds to a specific set of points in the image and does this set of points match, via alignment, with the object geometric description; or does the set of points corresponds to a property holding between objects which, in turn, have been matched via alignment, with the objects representation. And so forth.

Observable Fluents

$$\begin{array}{l} Holding(x,s), InTrolley(x,s), Green(x,s), Dist(x,y,s), \\ Covering(x,s), EmptyTrolley(s) \end{array}$$

$Holding(x, s)$ is the observable fact stating that the agent is holding the stone x in situation s. $InTrolley(x, s)$ is the observable fluent stating that the stone x is in the trolley in situation s. $Green(x, s)$ is the observable property of a stone x to being green. $Dist(x, z, s)$ states that in situation s the distance of stone x is z. $Covering(x, s)$ says that in situation s the arm and the trolley are covering stone x and $EmptyTrolley(s)$ says that in situation s the trolley is empty.

The Initial Database $\mathcal{D}_{S_0}$

$$EmptyTrolley(S_0)$$

The only fact known in the initial database is that the trolley is empty.

Action Precondition axioms for control actions

$$\begin{array}{ll}
Poss(moveAway10(x), s) & \equiv \top \\
Poss(moveToward10(x), s) & \equiv \neg Covering(x, s). \\
Poss(pickUp(x), s) & \equiv \exists z\ Dist(x, z, s) \land z \leq 10). \\
Poss(putInTrolley(x, s) & \equiv Holding(x, s).
\end{array}$$

The above axioms state the preconditions to execute the control actions we have discussed above. The first states that is always possible to move away from a stone. The second states that it is possible to move toward a stone if the arm is not already over the stone. The third sates that it is possible to pick up a stone if the distance between the arm and the stone is at most 10 units, this is the length of the arm. The last states that the stone can be put in the trolley if the arm is holding the stone.

Successor state axioms for Observable Fluents

$$\begin{array}{ll}
Holding(x, do(a, s)) & \equiv a = pickUp(x) \lor \\
& \quad (a \neq putInTrolley(x) \land Holding(x, s))
\end{array}$$

$$\begin{array}{ll}
Dist(x, dist, do(a, s)) & \equiv \exists z.\ Dist(x, z, s) \land \\
& ((a = moveToward10(x) \land dist = z - 10) \lor \\
& (a = moveAway10(x) \land dist = z + 10)) \lor \\
& (a \neq pickUp(x) \land a \neq putInTrolley(x) \land \\
& (a \neq moveAway10(x) \land \forall y (y \neq x \rightarrow \\
& \quad a \neq moveToward10(y)) \land Dist(x, dist, s))). \\
InTrolley(x, do(a, s)) & \equiv a = putInTrolley(x) \lor InTrolley(x, s). \\
Green(x, do(a, s)) & \equiv Green(x, s).
\end{array}$$

The above successor state axioms for observable Fluents state what will be changed of the environment after the execution of a control action, whenever the contexts for the control actions are verified.

Definitions for observable Fluents

$$\begin{array}{ll}
Covering(x, s) & \equiv \exists z Dist(x, z, s) \land z \leq 10. \\
Empty(s) & \equiv \neg \exists x InTrolley(x, s).
\end{array}$$

For the above observable Fluents no successor state axiom is provided since they are definable in terms of other Fluents. Observe that $Covering$ states that the arm is covering a stone if the distance is less than the length of the arm, which we supposed to be about ten units.

Precondition axioms for sensing actions

$$
\begin{array}{l}
Poss(sense(isDist(x,z),pr),s) \equiv \\
\quad (pr = 1 \wedge isDist(x,z) \subset Scene(s)) \vee (pr = 0 \wedge \\
\quad\; (isDist(x,z) \not\subset Scene(s)). \\
Poss(sense(isStone(x),pr),s) \equiv \\
\quad (pr = 1 \wedge isStone(x) \subset Scene(s) \wedge \neg Occluded(isStone(x),s)) \vee \\
\quad (pr = 0 \wedge (isStone(x) \not\subset Scene(s) \vee Occluded(isStone(x),s))). \\
Poss(sense(isTrolley,pr),s) \equiv \\
\quad (pr = 1 \wedge isTrolley \subset Scene(s) \wedge \neg Occluded(isTrolley,s)) \vee \\
\quad (pr = 0 \wedge (isTrolley \not\subset Scene(s) \vee Occluded(isTrolley,s))). \\
Poss(sense(isInTrolley(x,n),pr),s) \equiv \\
\quad (pr = 1 \wedge isInTrolley(x,n) \subset Scene(s) \wedge \\
\quad\; \neg Occluded(isInTrolley(x,n),s)) \vee \\
\quad (pr = 0 \wedge isInTrolley(x,n) \not\subset Scene(s) \vee \\
\quad\; Occluded(isInTrolley(x,n),s)). \\
Poss(sense(isGreen(x),pr),s) \equiv \\
\quad (pr = 1 \wedge Percept(isDist(x,z),1,s) \wedge dist \leq 10 \\
\quad \wedge isGreen(x) \subset Scene(s)) \vee \\
\quad (pr = 0 \wedge Percept(isDist(x,z),1,s) \wedge dist > 10 \vee \\
\quad\; isGreen(x) \not\subset Scene(s)))).
\end{array}
$$

The above precondition axioms for sensing actions are typical in the sense that for each sensing action over a perceptible isF it is required that the outcome of sensing will be 1 if, whenever the precondition for sensing are satisfied, the object is *recognized* in the image: the perceptible belongs to the scene in the situation in which it is sensed and in this very situation it is not occluded. We gave a hint about occlusion in Section 2.3; here we just avoid to insist on the definition of a fluent that requires a specific work on its own to be fully defined.

Definitions for Descriptions of perceptibles

$$
\begin{array}{ll}
isStone(x) \subset Scene(s) & \equiv \exists\ R\ id.\ \Psi_{isStone}(R,s,x) \wedge \\
& \quad align(isStone(x),id) \wedge \\
& \quad Gmodel(id, isStone(x)). \\
isGreen(x) \subset Scene(s) & \equiv Percept(isStone(x),1,s) \wedge \\
& \quad \forall z.z \in isStone(x) \rightarrow H(z) = 120. \\
isDist(x,z) \subset Scene(s) & \equiv Percept(isStone(x),1,s) \wedge \\
& \quad \exists p\ (p \in isStone(x) \wedge dist(p) = z). \\
isTrolley \subset Scene(s) & \equiv \exists\ R\ id.\ \Psi_{isTrolley}(R,s) \wedge \\
& \quad align(isTrolley,id) \wedge \\
& \quad Gmodel(id, isTrolley). \\
isInTrolley(x,n) \subset Scene(s) & \equiv Percept(isTrolley,1,s) \wedge \\
& \quad Percept(isStone(y),1,s) \wedge \\
& \quad PPartOf(isStone(y), isTrolley, s).
\end{array}
$$

In the above definitions $\Psi_{isF}(R,s)$ is a sentence that states the conditions made true by the elaboration process of the vision system that led to the identification of the set of points R corresponding to the perceptible isF. We recall that we have two kinds of descriptions. The geometric descriptions, that here we represent with the predicate $Gmodel(model, perceptible)$, which are given only for primitive objects – assumed as prototypical – and that we shall not treat here at all, and the descriptions for perceptibles. The description for perceptibles, of which those provided above are a coarse approximation, should specify the conditions under which a certain set of points in the scene, which have been isolated after suitable elaboration of the image – filtration, segmentation, identification of vertices and edges and so forth, see Section 2.3 – are matched against the geometric representation of a given prototype. Where the matching is achieved by an alignment process [83, 40] whose outcome is stated here by the predicate $align(perceptible, model)$. In general a description is either one of a primitive object, like for $isStone$ or is one for a relation holding between or among objects, like $isInTrolley$. If the description is concerned only with a primitive object, in the simplest case no relations with other regions of the image are needed to describe the object. If the perceptible is *composite* then the description would account for the *topological* relations holding for the set of points involved in the perceptible. In fact the last description $isInTrolley(x, n) \subset Scene(s)$ states that in the image a stone is in the trolley, in a situation s, whenever both the stone and the trolley have already been recognized and the relation, that we express as being in the trolley, corresponds in the image to the relation between the perceptible $isStone$ and the perceptible $isTrolley$ being one proper part, i.e. $PPartOf$, the other. The descriptions for the perceptibles $isDist$ and $isGreen$ are peculiar because they require a direct outcome from the vision system. For $isDist$ the direct outcome is simply the depth of a point belonging to the set of points which, in the image, have already been matched with a stone. For $isGreen$ is analogous but the pixels associated with stone are required to correspond to a certain function which depends on the color model. For example if the color model is the HSV one, then green corresponds to $120°$ in the hexcone [75], or six-sided pyramid defined within the model.

Successor State Axioms for Percept

$$
\begin{array}{l}
Percept(isHolding(x), pr, do(a, s)) \equiv \\
\quad (pr = 1 \wedge a = sense(isHolding(x), pr) \vee \\
\quad (a \neq sense(isHolding(x), 0) \wedge Percept(isHolding(x), pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isHolding(x), pr) \vee \\
\quad (\forall pr' \; a \neq sense(isHolding(x), pr') \wedge Percept(isHolding(x), pr, s))).
\end{array}
$$

$$
\begin{array}{l}
Percept(isDist(x, z), pr, do(a, s)) \equiv \\
\quad (pr = 1 \wedge a = sense(isDist(x, z), pr) \vee \\
\quad\; (a \neq sense(isDist(x, z), 0) \wedge Percept(isDist(x, z), pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isDist(x, z), pr) \vee \\
\quad\; (a \neq sense(isDist(x, z), 1) \wedge Percept(isDist(x, dist), pr, s))).
\end{array}
$$

$$
\begin{array}{l}
Percept(isInTrolley(x), pr, do(a, s)) \equiv \\
\quad (pr = 1 \wedge a = sense(isInTrolley(x), pr) \vee \\
\quad\; (a \neq sense(isInTrolley(x), 0) \wedge Percept(isInTrolley(x), pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isInTrolley(x), pr) \vee \\
\quad\; ((\forall pr' \; a \neq sense(isInTrolley(x), pr') \wedge \\
\quad\quad Percept(isInTrolley(x), pr, s))).
\end{array}
$$

$$
\begin{array}{l}
Percept(isGreen(x), pr, do(a, s)) \equiv \\
\quad (pr = 1 \wedge a = sense(isGreen(x), pr) \vee \\
\quad\; (a \neq sense(isGreen(x), 0) \wedge Percept(isGreen(x), pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isGreen(x), pr) \vee \\
\quad\; (\forall pr' \; a \neq sense(isGreen(x), pr') \wedge Percept(isGreen(x), pr, s))).
\end{array}
$$

The above successor state axioms for the perceptibles $isHolding$, $isDist$, $isInTrolley$ and $isGreen$ state what changes in the mental state of an agent once a sensing action is performed in the required context.

Observe that for the following perceptibles, namely $isStone$ and $isTrolley$ we do not have associated observable Fluents; this means that we will have neither $Mistaken$ nor $SePercept$. We could easily introduce the two observables $Stone$ and $Trolley$ just by adding two successor state axioms of the form:

$$Trolley(do(a,s)) \equiv Trolley(s) \text{ and } Stone(x, do(a,s)) \equiv Stone(x,s)$$

It would be redundant. Moreover we can assume that there is no risk for the agent to confuse the trolley with a stone –by the position, the colors, etc– and, since the agent knows nothing about the stones the addition of an observable would not help. In this sense having both $SePercept$ and $Mistaken$ as definitions helps, since whenever a perceptible is not associated with a fluent, and vice versa, these definitions can be avoided. Observe also that there are no constants in the language, therefore there is no way to name a new stone as soon as it is sensed. Therefore in the successor state axiom for the perceptible $isStone$ we will add a way to name with a number each stone sensed – i.e. if the outcome of sensing is 1.

$$\begin{array}{l}
Percept(isStone(x), pr, do(a,s)) \equiv \\
\quad (pr = 1 \wedge (a = sense(isStone(x), pr) \wedge \Psi(x,s)) \vee \\
\quad\ (a \neq sense(isStone(x), 0) \wedge Percept(isStone(x), pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isStone(x), pr) \vee \\
\quad ((\forall pr'\ a \neq sense(isStone(x), pr') \vee \neg\Psi(x,s)) \\
\quad\ \wedge Percept(isStone(x), pr, s))).
\end{array}$$

$\Psi(x,s)$ is the following sentence:

$$\begin{array}{l}
(s = S_0 \vee (\forall s'.s' \sqsubseteq s \rightarrow \neg\exists x'\ Percept(isStone(x'), 1, s')) \rightarrow x = 1) \wedge \\
(\exists s'\ x'\ (s' \sqsubseteq s \wedge Percept(isStone(x'), 1, s') \wedge \\
\forall x''\ s''(s' \sqsubseteq s'' \sqsubseteq s \wedge Percept(isStone(x''), 1, s'') \rightarrow x'' = x') \rightarrow x = x' + 1)
\end{array}$$

The above sentence names after each sensing action of the form $sense(isStone(x), 1)$ a stone with a number in a naturally ordered sequence: if $s = S_0$ then the action $a = sense(isStone(x), 1)$ will be the first sensing, therefore the sensed stone will be called 1; afterwards, after each sense action of the form $sense(isStone(x), 1)$ each stone will be named with a natural number which is the successor of the previous one.

$$\begin{array}{l}
Percept(isTrolley, pr, do(a,s)) \equiv \\
\quad (pr = 1 \wedge a = sense(isTrolley, pr) \vee \\
\quad\ (a \neq sense(isTrolley, 0) \wedge Percept(isTrolley, pr, s))) \vee \\
\quad (pr = 0 \wedge a = sense(isTrolley, pr) \vee \\
\quad\ (\forall pr'\ a \neq sense(isTrolley, pr') \wedge\ Percept(isTrolley, pr, s))).
\end{array}$$

□

We want, now, to determine whether

$$\mathcal{D} \models \exists s \exists x. Green(x,s) \wedge InTrolley(x,s). \tag{21}$$

Observe that we require that all the actions are possible, that is, $\mathcal{D}$ entails:

$$\begin{array}{l}
\exists s\ \exists x\ (\ \forall a \forall s'\ s''.s' \sqsubseteq s \wedge s' = do(a, s'') \rightarrow (Poss(a, s'') \wedge \\
\quad Green(x,s) \wedge InTrolley(x,s))).
\end{array} \tag{22}$$

which is not since nothing is known in the initial database S_0 about where the stones are and where the green stones are. The only fact known in S_0 is that the trolley is empty.

It is pretty clear that only sensing can be used for the required task. So let us require only that sensing actions are possible and that

$$\begin{aligned}\mathcal{D} \models\ & \exists s\, \exists x\ (\ \forall a \forall s' \forall s''(s' \sqsubseteq s \wedge s' = do(a, s'') \wedge \\ & \forall p\ pr.\ a = sense(p, pr) \rightarrow Poss(a, s'')) \wedge \\ & SePercept(isGreen(x), 1, s) \wedge SePercept(isInTrolley(x), 1, s)).\end{aligned} \tag{23}$$

If we can prove (23) then we can use meaningful perception, to add the suitable hypotheses to the initial database $\mathcal{D}_{S_0}$ to get

$$\mathcal{D} \models\!\approx \exists s \exists x. Green(x, s) \wedge InTrolley(x, s).$$

with all the actions possible.

Therefore let us assume, to simplify the example, that no perceptible is occluded, that *align* does the required job, that is, the description of stone and trolley are always correctly matched, that the sensing for the perceptibles *isGreen* and *isDist* are correctly performed, that $k = 1$ and let σ be the following sequence:

$$\begin{aligned}&[sense(isStone(1), 1), sense(isDist(1, 25), 1), moveToward10(1), \\ &\quad sense(isDist(1, 15), 1), moveToward10(1), sense(isDist(1, 5), 1), \\ &\quad sense(isGreen(1), 1), \\ &\qquad pickUp(1), putInTrolley(1), sense(isInTrolley(1), 1)].\end{aligned} \tag{24}$$

Under the above assumption, all the sensing actions are possible and

$$\mathcal{D} \models SePercept(isInTrolley(1), 1, \sigma) \wedge SePercept(isGreen(1), 1, \sigma)$$

Moreover:

$$\begin{aligned}
\mathcal{D} \models\ & Percept(isStone(1), 1, [sense(isStone(1), 1)]). \\
\mathcal{D} \models\ & SePercept(isDist(1, 25), 1, [sense(isStone(1), 1), \\
& \quad sense(isDist(1, 25), 1)]). \\
\mathcal{D} \models\ & SePercept(isDist(1, 15), 1, [sense(isStone(1), 1), \\
& \quad sense(isDist(1, 25), 1), \\
& \quad moveToward10(1), sense(isDist(1, 15), 1)]). \\
\mathcal{D} \models\ & SePercept(isDist(1, 5), 1, [sense(isStone(1), 1), \\
& \quad sense(isDist(1, 25), 1), moveToward10(1), \\
& \quad sense(isDist(1, 15), 1), moveToward10(1)]). \\
\mathcal{D} \models\ & SePercept(isGreen(1), 1, [sense(isStone(1), 1), \\
& \quad sense(isDist(1, 25), 1), moveToward10(1), \\
& \quad sense(isDist(1, 15), 1), moveToward10(1), \\
& \quad sense(isDist(1, 5), 1), sense(isGreen(1), 1)]). \\
\mathcal{D} \models\ & SePercept(isInTrolley(1), 1, [sense(isStone(1), 1), \\
& \quad sense(isDist(1, 25), 1), moveToward10(1), \\
& \quad sense(isDist(1, 15), 1), moveToward10(1), \\
& \quad sense(isDist(1, 5), 1), sense(isGreen(1), 1), \\
& \quad sense(isTrolley, 1), sense(isInTrolley(1), 1)]).
\end{aligned}$$

It is easy to verify that σ is safe and for each σ_{p_i} and for each perceptible $isF_{i, \sigma_{p_i}}$, *Mistaken* is not entailed by $\mathcal{D}$. So for example

$$\mathcal{D} \not\models Mistaken(isDist(1, 25), [sense(isStone(1), 1), sense(isDist(1, 25), 1)])$$

Now, in order to gain the hypotheses that have to be added to $\mathcal{D}_{S_0}$, for each $isF_{i,\sigma_{p_i}}$ which is last, and such that $SePercept(isF_i, 1, \sigma_{p_i})$ is entailed by $\mathcal{D}$, we will do the regression of the associated fluent, so for example we do the regression of

$$InTrolley(1,\sigma), Green(1, [sense(isStone(1),1),\cdots],$$
$$Dist(1,5,[sense(isStone(1),1)\cdots],\ldots$$

and we get that the following facts should be added to $\mathcal{D}_{S_0}$:

$$Dist(1,25,S_0), Green(1,S_0)$$

It is easy to see that:

$$\mathcal{D}_{S_0} \cup \{Dist(1,25,S_0), Green(1,S_0)\} \cup \mathcal{D}_{ss} \cup \mathcal{D}_{una} \cup \mathcal{D}_{da} \cup \Sigma$$

will entail (22).

□

It is important to observe that we required above that only sensing actions were possible. This might not be the correct assumption when on-line reasoning and conditional plans are involved. In fact in these cases the set of actions have to be constructively provided. There are different possible ways to get around this problem. The first one is to include in each precondition for control action also the percept, so for example we would rewrite the preconditions for $moveToward$ as follows:

$$\begin{aligned} Poss(moveToward(x), s) \equiv\ & \neg Covering(x,s) \vee \\ & Percept(isStone(x), 1, s) \wedge \\ & Percept(isDist(x,z), 1, s) \wedge z \not\leq 10. \end{aligned}$$

The above action precondition axiom would make the action $moveToward$ possible based on the perception of those facts, i.e. that the distance is not less than ten units, that make the action possible. With this transformation it is easy to see that also $moveToward$ becomes possible with respect to the above sequence of actions.

Another possible solution is to ask whether $\mathcal{D}$ entails the translation of the preconditions of the control action that is not possible in the situation s. So in the above example we could require whether

$$\begin{aligned} \mathcal{D} \models\ & SePercept(isStone(1), 1, [sense(isStone(1), 1), \\ & \quad sense(isDist(1,25),1)]) \wedge \\ & SePercept(isDist(1,25), 1, [sense(isStone(1), 1), \\ & \quad sense(isDist(1,25),1)]) \end{aligned}$$

A third possibility is a combination of the above ones, in which a situation s is said to be *executable for perception* if for each control action a, mentioned in s, the preconditions of $Poss(a, s)$, translated in meaningful form, are entailed in $\mathcal{D}$.

All these hypotheses will be treated in the second part of this work, that will in fact deal with conditional plans and on-line execution. If on-line execution is taken into account or conditional planning then the history has to be constructively determined. Such a history will be a branching one, as it would depend on the outcome of sensing. The interesting aspects here are that, depending on the outcome of sensing and on the fact that no mistake can be entailed about this outcome, a specific hypothesis would be added to the initial database to ensure meaningful perception. However, in general, in on-line execution it would be enough to ensure that mistaken is not entailed, after a sensing action: in other words, the notion of meaningful perception serves as the correct semantics for on-line execution of sensing and conditional planning.

□

We have shown with the above example how meaningful perception works and, at the same time, how useful it can be in those domains in which a task can be accomplished only using perception. Observe that we have emphasized only what follows deductively from the theory $\mathcal{D}$, in which perception is involved.

7 Acknowledgments

We thanks Ray Reiter for his fundamental suggestion of modifying the notion of *Mistaken* so as to replace the previous defined successor state axiom for *Mistaken* into a definition. He did also a big work in playing the nasty role of the devil advocate to improve our notion of perception. We thanks also Gerhard Lakemeyer because his work on SGolog allowed us to implement our version of OpenPerGolog, that accounts for the theory we have introduced here. Although we do not mentioned here OpenPerGolog, it had improved our understanding of the behaviour of perception. We thanks Elvis Ciafrone, Alex Moretto and Sandro D'Ambrosio for all their work on PerceptONE, without a good and working implementation of a vision system none of the above ideas would have been possible. We finally thanks Michael Tielsher for the organization of NRAC99 and for his patience.

References

[1] Aiello, M., van Benthem, J. Logic Patterns in Space. no. PP-99-24, Prepublication Series, *Technical Report*, 1999, Uva, Amsterdam.

[2] Asher, N., Vieu, L. Toward a Geometry of Common Sense: A Semantics and a Complete Axiomatization of Mereotopology. In *Proceedings of IJCAI 95*, pp. 846–852, 1995.

[3] Bacchus, F., Halpern, J.Y., and Levesque, H. *Reasoning about Noisy Sensors in the Situation Calculus.* In *Proceedings of IJCAI 95*, pp. 1933–1940, 1995.

[4] Baral,C. and Son, T.C. Formalizing sensing actions - a transition function based approach. In *Proceedings of ILPS97*, pp. 387–401, 1997.

[5] Basry, R, and Ulmann, S. *The alignment of objects with smooth surfaces.* In *Computer Vision, Graphics, and Image Processing*, 57(3), (1993) 331–345.

[6] Basri, R. Recognition by prototype. *International Journal of Computer Vision*, 19(2), (1996) 147–168.

[7] Bennett, B., Cohn, A.G., and Isli. A. Combining Multiple Representations in a Spatial Reasoning System. In *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)*, Newport Beach, CA, pp. 314–322, 1997.

[8] Bennett, B. Determining Consistency of Topological Relations *CONSTRAINTS*, 3 (2 & 3), Special Issue on Spatial and Temporal Reasoning, Kluwer, (1998) 213-225.

[9] BIEDERMAN, I. Recognition by components. A theory of human image interpretation. *Psychological Review*, 94, (1987) 115-147.

[10] BIEDERMAN, I., COOPER, E.E. Evidence for complete translational and reflectional invariance in visual object priming. In *Perception*, 20, (1991) 585-593.

[11] BIEDERMAN, I., COOPER,. E.E. AND HUMMEL, J.E. Metric Invariance in Object recognition: A review and further evidence. *Canadian Journal of Psychology*, 46(2), (1992) 191–214.

[12] BIEDERMAN, I. Visual Object Recognition. In [59], pp. 121–166.

[13] BROOKS, R. Symbolic reasoning among 3-dimentional models and 2-dimentional images. In *Artificial Intelligence*, 17, (1981) 285–349.

[14] CANNY, J. A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 8, (1986) 679–698.

[15] CHELLA, A., FRIXIONE, M., GAGLIO, S. A cognitive architecture for artificial vision. In *Artificial Intelligence*, 89, (1997) 73–111.

[16] CLARKE, B.L. A calculus of individuals based on 'connection'. *Notre Dame Journal of Formal Logic*, 23(3), (1981) 204–218.

[17] CLARKE, B.L. Individuals and points. *Notre Dame Journal of Formal Logic*, 26(1), (1985) 61–75.

[18] CLEMENTINI, E., SHARMA, J., AND EGENHOFER, M.J Modeling topological spatial relations: strategies for query processing. *Computer and Garphics*, 18(6), (1994) 815–822.

[19] CIAFRONE, E. PerceptONE: A solution to the edges detection problem and 2D/3D alignment. *Master Thesis in Computer Engineering*, University of Rome "La Sapienza", May 1999.

[20] COHN, A.G., AND VARZI, A. Connection Relations in Mereotopology. H. Prade, (Ed.), *Proceedings of 13th European Conf. on AI (ECAI98)*, pp. 150–154, John Wiley, 1998.

[21] COHN, A.G., AND VARZI, A. Modes of Connection. In C. Freksa and D. M. Mark (Eds.), *Spatial Information Theory-Cognitive and Computational Foundations of Geographic Information Science*, Lecture Notes in Computer Science 1661, Springer, Berlin, pp. 299–314, 1999.

[22] CORADESCHI, S. Reasoning with Unreliable Observations in the Features and Fluents Framework. Technical report, Department of Computer and Information Science, Linköping University, 1995.

[23] CORADESCHI, S. Reasoning with Misperception in the Features and Fluents Framework. In *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96)*, pp. 12–16, 1996.

[24] CORADESCHI, S., SAFFIOTTI, A. Anchoring symbolic object description to sensor data. Problem statement. In *Linköping electronic articles in computer and information science*, Vol. 4(1999): nr9.

[25] CORADESCHI, S., SAFFIOTTI, A. Anchoring symbols to vision data by fuzzy logic. in *Qualitative and Quantitative Approaches to Reasoning with Uncertainty*, In *LNAI*, A. Hunter and S. Parsons, (Eds.), pp. 104–115, Springer, Berlin, Germany, 1999.

[26] D'AMBROSIO, S. PerceptONE: Sampling and edges detection with an edges based methodology. *Master Thesis in Computer Engineering*, University of Rome "La Sapienza", May 1999.

[27] DE GIACOMO,G., REITER,R.,SOUTCHANSKI, M. Execution monitoring of high-level robot programs. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, pp. 453–464, 1998.

[28] DE GIACOMO, G. AND LEVESQUE, H.J. An Incremental Interpreter for High-Level Programs with Sensing. In *Logical Foundation for Cognitive Agents: Contributions in Honor of Ray Reiter*, Hector J. Levesque and Fiora Pirri, (Eds.), pp. 86–102, Springer 1999.

[29] DRETSKE, F. Meaningful Perception. In [59], pp 331–352.

[30] EDELMAN, S. Representation and Recognition in Vision. MIT Press, 1999.

[31] EGENHOFER, M. Reasoning about binary topological relations. In O. Gunther and H.J. Schek (Eds.), *Proceedings of the Second Symposium on Large Spatial Databases, SSD'91(Zurich, Switzerland).* Lecture Notes in Computer Science, 525, (1991) 143–160.

[32] EPSTEIN, W. Percept-Percept coupling. In [71], pp. 17–29.

[33] FARAH, M.J. Dissociable systems for visual recognition: a cognitive neuropsychology approach. In [59], pp. 101–119.

[34] FOLEY, J. AND VAN DAM, A. AND FEINER, K.S. AND HUGES, J.E. AND PHILLIPS, R.L. *Introduction to Computer Graphics*, Vol. 2. Addison-Wesley, 1997.

[35] GÄRDENFORS, P. *Conceptual Spaces* MIT Press, Bradford Books, Cambridge, MA 1999, in press.

[36] J. J. GIBSON The ecological approach to visual perception. Houghton Mifflin, Boston, 1980.

[37] GOTTS, N.M. Formalising Commonsense Topology: The INCH Calculus. In *Proceedings of the Fourth International Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, Florida, January 1996.

[38] HERNANDEZ, D. *Qualitative Representation of Spatial Knowledge, Lecture Notes in Atificial Intelligence*, Vol. 804, Springer-Verlag, 1994.

[39] HERNANDEZ, D, CLEMENTINI, E., DI FELICE, P. *Qualitative distances*, In Kuhn, W., Frank, A., (Ed.), *Spatial Information Theory: a theoretical basis for GIS*, number 988 in Lecture Notes in Computer Science, pp. 45–58, Springer-Verlag, Berlin, 1995.

[40] HUTTENLOCHER, D.P. AND ULMANN, S. Recognizing Solid Objects by alignment with an image. *International Journal of Computer Vision*, 5, (1990) 195–212.

[41] IVRY, R.B. AND ROBERTSON, L.C. The two sides of perception. MIT Press, 1998.

[42] JOLICOEUR, P. AND GLUCK, M.A AND KOSSLYN, S.M. Picture and names: making the connection. *Cognitive Psychology*, 16, (1984) 243–275.

[43] Kendall, D.G. Shape manifold, Procustean metrics and complex projective spaces. *Bull. Lond. Math. Soc*, 16, (1984) 81–121.

[44] Kuratowski, K. *Introduction to Set Theory and Topology.* Pergamon Press, 2nd edition, 1972.

[45] Kurbat, M.A. Structural description theories: is RBC/JIM a general-purpose theory of human entry-level object recognition? In *Perception*, 23, (1994) 1339–1368.

[46] Lakemeyer, G., and Levesque, H. AOL:a logic of acting, sensing, knowing and only knowing. In *KR-98*, pp. 316–327, Trento, IRST, 1998.

[47] Lakemeyer, G. On Sensing and Off-Line Interpreting in GOLOG. In *Logical Foundation for Cognitive Agents: Contributions in Honor of Ray Reiter*, Hector J. Levesque and Fiora Pirri (Eds.), pp. 173–189, Springer, 1999.

[48] Levesque, H. *What is planning in the presence of sensing?* In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, AAAI-96, pp. 1139–1146, Portland, Oregon, August 1996.

[49] Levesque, H., and Pirri, F., and Reiter, R. *Foundations for the Situation Calculus. Series: Linkping Electronic Articles in Computer and Information Science*, Vol. 3: nr 018, (1998) 159–178.

[50] Lobo, J., Taylor, S., and Mendez, G., *Adding knowledge to the action description language A* In *Proceedings of AAAI-97*, pp. 454–459, AAAI Press, July 1997.

[51] D. Marr *Vision.* W.H. Freeman and Company, New York, 1982.

[52] Marr, D., Nishihara, H.K. Representation and recognition of the spatial organization of three-dimansional shapes. In *Proceedings of the Royal Society*, London, B200, pp. 269–294, 1978.

[53] McCarthy, J. and Hayes, P. Some philosophical problems from the standpoint of AI. B. Meltzer and D. Michie (Eds.), *Machine Intelligence* 4, Edinburgh University Press (1969) 463–502.

[54] Mervis, C. and Rosh, E Categorization of natural objects. *Annual Review of Psycology*, 32, (1981) 89–115.

[55] Moore, R. C. A Formal Theory of Knowledge and Action. In J. R. Hobbs and R. C. Moore (Eds.), *Formal Theories of the Commonsense World*, Ablex, Norwood, NJ, (1985) 319–358.

[56] Moretto, A. PerceptONE: filtering, noise elimination and segmentation. *Master Thesis in Computer Engineering*, University of Rome “La Sapienza”. May 1999.

[57] Mumford, D. Mathematical Theories of Shape: Do they model perception?. *Geometric methods in Computer Vision*, SPIE Vol. 1570, (1991) 115–147.

[58] Murphi, R.R. *Coordination and Control of Sensing for mobility using action-oriented perception.* In *Artificial Intelligence and Mobile Robots* D. Kortenkamp, R.P. Bonasso and R. Murphy (Eds.), AAAI, The MIT Press, 1998.

[59] Osherson, D.N. and Kosslyn, S.M. *Visual Cognition*, Vol. 2, The MIT Press, 1995.

[60] PALMER, S.E. *The legacy of Irvin Rock in indirect perception.* In [71], Foreword.

[61] PIRRI, F., AND REITER, R. *Some Contributions to the Metatheory of the Situation Calculus. Journal of the ACM*, 46(3), (1999) 325–361.

[62] PIRRI, F., AND FINZI, A. *A preliminary approach to Perception in Theory of Action.* In *Proceedings of the IJCAI99's workshop on Non-monotonic Reasoning, Action and Change (NRAC99)*, pp. 49–56, 1999.

[63] RANDELL,D.A., AND COHN, A.G. Naive Topology: modelling the force pump. Struss, P., and Faltings, B., (Eds.), In *Advances in Qualitative Physics*, MIT Press, (1992) 177–192.

[64] RANDELL, D. A., CUI, Z., AND COHN, A. G. A Spatial Logic based on Regions and Connection. In *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, Boston, October, 1992.

[65] REQUICHA, A.A.G. Representations for rigid solids: Theory, Methods and Systems. *ACM Computing Surveys*, 12(4), (1980) 437–464.

[66] REQUICHA, A.A.G. AND VOELCKER, H.B. Boolean operations in solid modeling: Boundary evaluation and merging algorithms. *Proceedings of IEEE*, 73(1), (1980) 30–44.

[67] REITER, R. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, V. Lifschitz (Ed.), Academic Press, 1991, pp. 359–380.

[68] REITER, R. Proving properties of states in the situation calculus. *Artificial Intelligence Journal*, 64, (1993) 337–351.

[69] REITER, R. Knowledge in Action: logical foundations for describing and implementing dynamical systems. *MIT press*, to appear. Available at http://www.cs.toronto.edu/~cogrobo.

[70] ROCK, I. *The Logic of Perception.* The MIT Press, 1985.

[71] ROCK, I. *Indirect Perception.* The MIT Press, 1996.

[72] SANDEWALL, E. *Features and Fluents*, Oxford Press, 1994.

[73] R. SCHERL AND H.J. LEVESQUE. *The Frame Problem and Knowledge Producing Actions.* In *Proceedings of the National Conference on Artificial Intelligence (AAAI-93)*, AAAI Press/The MIT Press, pp. 689–695, Washington, DC, July 1993.

[74] SHANAHAN, M. What Sort of Computation Mediates Best Between Perception and Action? In *Logical Foundation for Cognitive Agents: Contributions in Honor of Ray Reiter*, Hector J. Levesque and Fiora Pirri (Eds.), pp. 352–369, Springer 1999.

[75] SMITH, A. R. *Color Gamut Transform Pairs. SIGGRAPH 78*, (1978) 12–19.

[76] SMITH, A. R., BRADY, M. Susan - a new approach to low level image processing. In *Inernational Journal of Computer Vision*, 23(1), (1997) 45-78.

[77] Squire, L.R. and Kosslyn, S.M. Findings and current opinions in Cognitive neuroscience. The MIT press, 1998.

[78] Tarsky, A. Sentential calculus and topology. In *Logic, Semantic, Methamatematics*, Oxford Clarendon Press, 1956.

[79] Tarsky, A., and Beth, W. Equilatery as the only primitive notion of Euclidian geometry. *Indagationes Mathematicae*, 18, (1956) 462-467.

[80] Treisman, A. and Vieira, A. and Hayes, A. *Automaticity and preattentive processing. American Journal of Psychology*, 105, (1992) 341-362.

[81] Treisman, A. *The binding Problem.* In [77].

[82] Ulmann, S. *Against direct perception.* In *Behavioral and Brain Science*, 3, (1980) 373–381.

[83] Ulmann, S. High Level Vision The MIT press, 1996.

[84] Van der Does, J., and Van Lambalgen,J. A Logic of Vision. Research report LP-97-07. University of Amsterdam: ILLC, 1997.