Reasoning with unreliable observations in the Features and Fluents Framework

Silvia Coradeschi
Email: silco@ida.liu.se

Abstract

In this work we consider a way to deal with the problem of unreliable observations when an agent is reasoning about dynamical systems as they are formalized and systematically studied in Sandewall's approach to reasoning about action and change. The presence of incorrect observations can be detected in case it generates a contradiction. In this case a revision function resolves the inconsistency by constructing alternative consistent descriptions of the system. However, revision is in general expensive and therefore we define a delayed revision that postpones doing revision for some steps and we prove that it gives results similar to immediate revision. We also examine some forms of preferential revision that reduce the number of alternative descriptions of the system. We finally consider the relations between our work and Gärdenfors' approach to belief revision.

IDA Technical Report 1995
LiTH-IDA-R-95-44
ISSN-0281-4250

Department of Computer
and Information Science,
Linköping University,
S-581 83 Linköping, Sweden

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The general field to which this work belongs is *reasoning about actions and changes*, that is reasoning about how actions affect the world and how the world changes over time.

In particular we consider the case in which there are agents that perform actions and the world changes over time as a consequence of these actions.

The actions are characterized by *action laws*, that is axioms which describe what effects an action has or may have in a specific situation.

We consider in particular the problem of including the possibility to have unreliable knowledge in Sandewall's framework.

Sandewall's framework as it is described in [San94] has the aim of giving a uniform account of the various approaches to reasoning about actions and change in a way that could make clear their properties and their range of applicability.

To achieve this result it is not enough just to consider test examples and informal arguments in favour of different approaches, but a systematic and formal framework in which different approaches can be compared and studied is needed.

This formal framework is given by Sandewall in terms of a *taxonomy* of reasoning problems in which the assumptions that, usually implicitly, are made in reasoning about actions and changes are clearly defined. The different approaches can be classified according to the assumptions they make and studying what a certain assumption implies can determine the range of applicability of each approach.

The taxonomy is based on *ontological assumptions* about the actions and the world where the actions occur and on *epistemological assumptions* about the knowledge about them.

This systematic approach has given new impulses to the study of the cases in the taxonomy not yet largely explored, for example the case of concurrent action [Yi95] and it has given the opportunity of studying interesting problems, as planning, from a different point of view and with the support of Sandewall formalism [Kar95].

The work that we present in this thesis gives a solution of one of the cases of Sandewall taxonomy, that is the case in which the agent's knowledge of the world is unreliable. Our agent receives information from the world, but this information can be incorrect. The agent detects the existence of incorrect information by discovering contradictions between one piece of information and another, but in general, even if it is aware that something is incorrect, the agent can't precisely identify what is incorrect. We propose a solution where the agent, in case it finds a contradiction, construct several alternative descriptions of the world, each one containing a consistent subset of the observations received.

Our approach can be considered a coherence-based approach as defined in [BDP95], that is an approach where the aim is revising the knowledge base and restoring consistency. In particular it can be considered near Rescher and Mannor approach to reasoning with inconsistency [RM70].

The problem of unreliable observations is related to the problem of belief revision and in particular to Gärdenfors' work [Gär88]. In the last chapter we consider the relation with this work and we point out similarities and differences.

In chapter 2 we describe the features and fluents framework at least for the part that we use in our work. In chapter 3 we introduce the problem of incorrect observations and we give a solution to it in terms of narrative and revision of narrative. In chapter 4 we examine the relations between our work and belief revision, in particular we discuss how Gärdenfors' postulates apply to our case. Finally in chapter 5 we give our conclusions and a description of future work.

Chapter 2

The Features and Fluents framework

2.1 Introduction

In this chapter the Features and Fluents framework introduced by Erik Sandewall in [San94] is described. Our intention is not to give a deep description of the framework, but to give a general view of the work and to describe more in details the parts that are related to our work. What it is written in this chapter can be found much more accurately and amply described in [San94].

Sandewall's framework addresses knowledge representation and the reasoning about inert and inhabited dynamical systems. The agents inhabiting the system are characterized by their ability to perform *actions*. A dynamical system is inert if things tend not to change unless there is a specific reason for which they should change. The concept of inert system is quite similar to the one expressed in physics by the principle of inertia.

In section 2.2 the concept of dynamical system and some of its properties are defined. In section 2.3 a general inhabited dynamical system with one or more agents is introduced. In section 2.4 the analysis is restricted to single-agent inhabited dynamical system and in this case it is examined the interactions between the "ego", a part of the agent, and the world.

One of the aims of Sandewall's framework is to define a taxonomy which characterizes classes of dynamical systems with specific restrictions on them. The classification serves as a reference system to decide the ranges of applicability of logics of action and change. In Sandewall's book a number of logics are described and analysed using this method. The taxonomy is described in 2.5 and in 2.6.

In 2.6 the notion of scenario is informally introduced and then formally defined. A scenario can be seen as a description of a world and the actions that the agent performs in the world.

Finally in 2.6.2 the sets of models for scenario descriptions are defined in terms of interaction between the ego and the world.

2.2 Dynamical systems

A *system* can be defined as a collection of *objects*. Each object has individual properties and can be in relation with other objects. Examples of properties are color, dimension, and weight; examples of relationships are adjacency, support, and mutual visibility.

A *state* of the system describes the property assignments and the relationships that hold at a point in time.

A dynamical system is a system in which the state changes over time. State changes depend on earlier states of the system, and on the input that the system receives from outside. In a dynamical system we can have *agents* which influence the system state's performing *actions* and *activities*. In this case the dynamical system is said to be *inhabited*. The difference between actions and activities is that the actions are performed with a specific purpose and terminate when the purpose has been achieved, they are therefore time limited, while the activities are terminated due to a separate decision by the agent.

A dynamical system is *inert* when its properties and relationships do not change except when there is a good reason why they must or may do so.

We can assume at least two levels of description for a dynamical system: the *material* level and the *image* level. The material level can be considered closer to what the system really is or it is even identical to it, while the image level is obtained from the material level with a process of perception or classification. In general the material level uses continuous time and continuous-valued variables that change according to the law of physics. The image level uses discrete-valued state variables. As there are two levels of system description there are also two state domains. We represent the material-level state domain as $\mathcal{R}_M$ and the image-level state domain as $\mathcal{R}$. The only constraint we impose is that $\mathcal{R}$ should provide less information than $\mathcal{R}_M$.

We need a mechanism for relating material and image level description. We call this mechanism *perception*. To give the formal definition of perception we first need to introduce the concept of *finite* and *complete history*.

Definition 1 (finite and complete history) *Let $\mathcal{T}$ be a timepoint* *domain and $\mathcal{R}$ a state domain. A finite history for $\mathcal{R}$ is a mapping $R: [0,t] \to \mathcal{R}$ for some $t \in \mathcal{T}$. A complete history for $\mathcal{R}$ is a mapping $R: \mathcal{T} \to \mathcal{R}$.*

$\mathcal{T}$ can, for instance, be the set of natural numbers or a finite set of numbers $\{0, \ldots, n\}$ for some n.

Definition 2 (perception function) *Let a material domain $\mathcal{R}_M$ and an image domain $\mathcal{R}$ be given. A perception function is a mapping* `Perc` *from finite histories for $\mathcal{R}_M$ to finite histories for $\mathcal{R}$ which satisfies the following assumptions:*

1. *The argument and value are histories over the same interval.*
2. *If R is a history for $\mathcal{R}_M$ over $[0,t]$, $s < t$, and $R_{0:s}$ is the restriction of R to the interval $[0,s]$, then $\texttt{Perc}(R)_{0:s} = \texttt{Perc}(R_{0:s})$.*

The second condition means that if we apply the `Perc` function to $\mathcal{R}$ and we take the part of $\texttt{Perc}(R)$ between the time points 0 and s, this gives the same result as applying the perception function to the part of the history between 0 and s.

It is important to notice that what has been perceived before time s is not reinterpreted because of what happens and is perceived later, but the perception during $[s,t]$ for some later time point t may be influenced by what has been perceived before time s.

Definition 3 (truly inert dynamical system) *A dynamical system is truly inert iff its properties and relationships change only in well understood ways, so that one reasons about developments in the system under the assumption that things do not change unless there is a specific reason why they must or may change.*

Definition 4 (perceived inert dynamical system) *A dynamical system is perceived inert iff there is a (useful) perception function such that the image is truly inert.*

It is important to point out that the perception function should be useful, otherwise we could take as perception function the one that maps all members of $\mathcal{R}_{\mathcal{M}}$ to an image without any information at all, and that image of course is truly inert.

The difference between a truly and a perceived inert system can be illustrated by the "goldfish hotel" example. A goldfish hotel consists of a number of goldfish bowls. The fish inside the bowls can move in the bowl but can not leave the bowls. The states in $\mathcal{R}_M$ specify, among a lot of other information, also the exact position of the fishes at each time point. If we choose a image domain that specifies only in which bowl the fishes are we have a system that is not truly inert, but it is perceived inert.

Definition 5 (strictly inert system) *A system is said to be strictly inert if there is only one way that the state can change, namely by the direct effect of an action.*

Each action type is assumed to influence a certain, well defined set of state variables in the system state. It is also assumed that all other state variables remain unchanged even when an action is performed, and that there is no change during those periods when there is no action.

Definition 6 (weakly inert system) *A system is said to be weakly inert if it has in general the properties of inertia, but there can be exceptions of this properties.*

Examples of exceptions can be: "surprises", "miracles", side-effects of action and causal chains of events that override inertia. Surprises consist in unexpected changes in the world that occurs only infrequently, while miracles consist in inexplicable changes in the world.

2.3 Inhabited Dynamical System (IDS)

An IDS consists of one or more *agents* which interact with a *material system* denoted by M in figure 2.1 Each agent consists of two parts, an *ego* K_i and a *vehicle* V_i. The vehicle is a system for perception and actuation. Relative to a given agent, the material system and the other agents are together called its *reality*. Relative to one ego, the reality and its own vehicle are together called its *world*.

The vehicle interacts with the material system and the ego interacts with the vehicle. We have two interfaces, one between vehicle and material system and one between ego and vehicle. We examine in detail the interface between ego and vehicle. The interface between vehicle and material system can be imagined as "physical" and can be characterized in quantitative terms.

We use *feature* for a characteristic aspect of an IDS which changes its value over time. For example "the glass is on the table" or "the glass is not on the table" are features.

The definition of fluent is the same as the one normally used in standard A.I. A *fluent* is a function from timepoints to corresponding

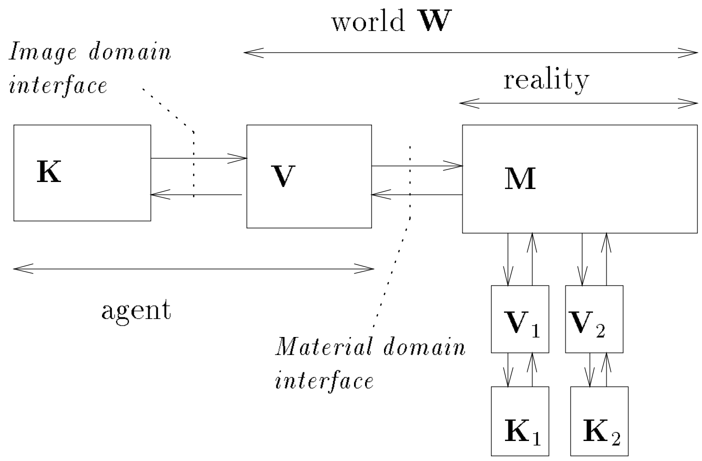

Figure 2.1: Structure of an IDS

values, usually taken from a finite set, and can be thought of as the trace over time of one of the state variables. A fluent is a function and not a function symbol.

An *IDS history* is a function $\mathcal{R}(f, t)$ which for each feature f and time-point t assigns an appropriate value for that feature. If a specific feature f is chosen, then the resulting function of time is a fluent. On the other hand if a specific time-point t is chosen, then the resulting function from features to corresponding values is a *state*.

There are three types of *occurrences*: actions, activities and events. In this work we consider mainly methods for reasoning about actions and not about activities and events, but it is important to distinguish clearly between these three terms because several authors in A.I. treat them as synonyms. An *action* happens over an interval of time through the intervention of an agent. An example of an action is: "the agent fills the glass with the water". An *event* happens in reality over an interval of time. If the event does not include any of the agents, it is called a *material event* and proceeds in accordance with "natural" laws. An example of an event is "the cup falls on the floor". An *activity* is similar to an action except that it can potentially go on for indefinite time, and its termination as well as its start is decided by the ego.

A *process* is a structure consisting of several occurrences (actions, activities and events), sequential or concurrent. A *schedule* is a possibly parameterized description of a process, expressed for example as a set of logical formulae. A *plan* is a schedule which satisfies certain constraints making it executable, and which is intended to be used for achieving a given goal or set of goals.

2.4 Ego/world interaction

We analyze the interaction between the ego and the world as a game where ego and world make alternating moves. We consider just the case of single-agent IDS as illustrated in figure 2.2. In this case the ego interacts with a world consisting of the material system and its own vehicle, and without other agents.

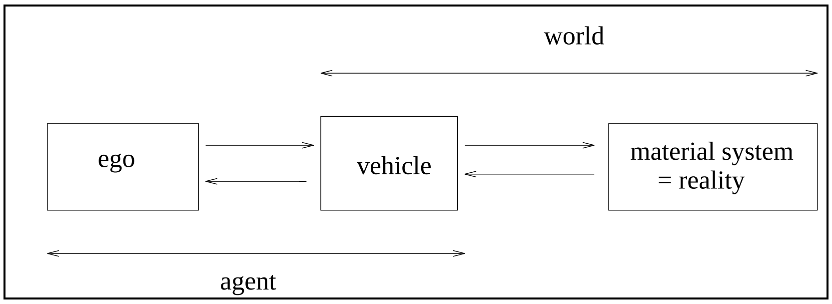

Figure 2.2: Single-agent IDS

The ego receives perceptions of the IDS reality which are produced by the sensory system in its vehicle and requests the vehicle to perform actions. To give a formal definition of the interaction between ego and world we introduce the concept of development.

Let a domain $\mathcal{R}$ of *states* and a domain $\mathcal{E}$ of *occurrence designators* be given. Each state $r \in \mathcal{R}$ assigns a discrete value to each of a number of *features*, for example "the current color shown by the traffic light number 216 is red". An *occurrence* is a tuple $\langle s, E, t \rangle \in \mathcal{T} \times \mathcal{E} \times \mathcal{T}$ where $s < t$ and $\mathcal{T}$ can be the infinite time domain of integers in $[0, \infty]$ or a finite time domain of integer in $[0, n]$. Let also a number of *temporal constant symbols* s_i and t_i, and a number of *object constant symbols* o_i be given. A finite development over $[0, n]$ can be seen as a discrete image of what has happened in the world up to time n.

Definition 7 (finite development) *A finite development is a tuple* $\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C} \rangle$ *where*

- $\mathcal{B} \subseteq \mathcal{T}$ *is a finite subset of* $\mathcal{T}$*, whose members are called breakpoints. The largest member of* $\mathcal{B}$ *is called now and is written as* $n_{\mathcal{B}}$ *or (when the choice of* $\mathcal{B}$ *is clear from context) as* n*. The subset of* $\mathcal{T}$ *with all values between* 0 *and* $n_{\mathcal{B}}$ *inclusive is written* $\mathcal{T}_{\mathcal{B}}$*.*

- *M, the valuation, is a mapping which assigns values to some or all of the temporal constant symbols t_i, and to all the object constant symbols o_i. The remaining temporal constants are left undefined.*

- *R, the history, is a mapping $R: \mathcal{T}_{\mathcal{B}} \rightarrow \mathcal{R}$. Thus it specifies the values of features at all times from zero until now.*

- *$\mathcal{A}$, the past action set, is a set of occurrences $\langle s, E, t\rangle$ where s and t are members of $\mathcal{B}$.*

- *$\mathcal{C}$, the current action set, is a set of tuples $\langle s, E\rangle$ where s is a member of $\mathcal{B}$, and E is a member of $\mathcal{E}$.*

Let $\mathcal{J}$ be a domain of finite developments.

Definition 8 (complete development) *A complete development (over images) is defined as a tuple $\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle$ as in the definition of finite development with the only difference that R is a mapping from $\mathcal{T}$ to $\mathcal{R}$ and that $\mathcal{B}$ may be either a finite or an infinite subset of $\mathcal{T}$.*

Let $\mathcal{J}^\star$ be a domain of complete developments.

Definition 9 (infinite development) *We define an infinite development as a complete development over the time domain $\mathcal{T} = [0, \infty]$.*

The ego and the world playing the game construct a development starting from an initial state of the world at time zero and ranging to plus infinity. When the ego makes a move he can specify that actions and activities are to be started and that activities are to be terminated at time *now*. When the world makes the respective move he can start event and terminate actions and events at time *now*. The world extends also the time interval and the definitions of feature values over the additional time-points.

Let us give the formal definitions of ego and world. We first introduce the auxiliary concept of correct revision.

Definition 10 (correct revision) *Let $J = \langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle$ and $J' = \langle \mathcal{B}', M', R', \mathcal{A}', \mathcal{C}'\rangle$ be two finite developments, and let n and n' be the largest members of $\mathcal{B}$ and $\mathcal{B}'$, respectively. J' is said to be a correct revision of J iff the following conditions hold. It is required that $\mathcal{B} \subseteq \mathcal{B}'$, $M \subseteq M'$, $\mathcal{A} \subseteq \mathcal{A}'$, and that the restriction of R' to the time interval $[0, n]$ shall equal R. Every $b \in \mathcal{B}' - \mathcal{B}$ must satisfy $b > n$. For every $\langle s, E, t\rangle \in (\mathcal{A}' - \mathcal{A})$, it must be the case that $t = n'$, and that $\langle s, E\rangle$ is a member of $\mathcal{C}$ but not of $\mathcal{C}'$. For every $\langle s, E\rangle \in (\mathcal{C} - \mathcal{C}')$, it must be the case that $s = n$ or $\langle s, E, n'\rangle \in \mathcal{A}'$. Also, for every $\langle s, E\rangle \in (\mathcal{C}' - \mathcal{C})$, it must be the case that $s = n$.*

The following changes can therefore be made to a finite development in a correct revision.

- To add members to $\mathcal{B}$ obtaining $\mathcal{B}'$. Any new member must be larger than the largest member of $\mathcal{B}$, and have the effect of extending the time interval of the finite development.
- To extend the time interval of the development, while defining the values of R' over the added interval but not changing it in the previous interval $[0, n]$.
- To add bindings to M, but not changing or removing previous bindings.
- To add elements of the form $\langle n, E\rangle$ to $\mathcal{C}$, meaning to start an action or activity.
- To remove an element of the form $\langle s, E\rangle$, where $s < n'$ from $\mathcal{C}$, and instead to add the element $\langle s, E, n'\rangle$ to $\mathcal{A}$, meaning to terminate the action or activity after a non-empty time interval.
- To remove a member of the form $\langle n, E\rangle$ from $\mathcal{C}$, without making a corresponding addition to $\mathcal{A}$. This represents that the world is unable to execute an action that has been invoked by the ego. The requested action is not recorded in $\mathcal{A}$.

We can now give the definitions of ego and world.

Definition 11 (ego) *An ego K is a mapping $\mathcal{J} \rightarrow \mathcal{J}$ where*

- *$K(J)$ is a correct revision of J, and*
- *if $K(\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle) = \langle \mathcal{B}', M', R', \mathcal{A}', \mathcal{C}'\rangle$, then $\mathcal{B} = \mathcal{B}'$.*

An ego therefore can:

- start an action or activity, by adding $\langle n, E\rangle$ to $\mathcal{C}$;
- end an activity, by removing $\langle s, E\rangle$ from $\mathcal{C}$ and making a corresponding addition to $\mathcal{A}$.

Definition 12 (world) *A world W is a transformation from $\mathcal{J}$ to $\mathcal{J} \cup \mathcal{J}^\star$, i.e. a subset of $\mathcal{J} \times (\mathcal{J} \cup \mathcal{J}^\star)$, such that $\forall J\, \exists J' \quad W(J, J')$, and if*

$$W(\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle, \langle \mathcal{B}', M', R', \mathcal{A}', \mathcal{C}'\rangle)$$

then $\mathcal{C}' \subseteq \mathcal{C}$, the second argument is a correct revision of the first argument, and if n is the largest member of $\mathcal{B}$, either of the following must hold:

1. *there is some $n' > n$ such that $\mathcal{B}' = \mathcal{B} \cup \{n'\}$, and $\langle \mathcal{B}', M', R', \mathcal{A}', \mathcal{C}' \rangle$ is a correct revision of $\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C} \rangle$.*

2. *$\mathcal{B} = \mathcal{B}'$, R' is a mapping from $\mathcal{T}$ to $\mathcal{R}$, but its restriction to $[0, n]$ equals R.*

The moves that the world can do are therefore:

- in one alternative it adds exactly one more member n' to $\mathcal{B}$, and constructs a new finite development as a correct revision up to n';
- in the other alternative the world leaves $\mathcal{B}$ unchanged, and extends the history R so that it becomes complete. This case can occur in two ways: for infinite time if the action continues over an infinite interval, and for finite time if the action continues to the end of time. In either case the "game" is concluded, and the ego is not allowed to make any more moves.

Definition 13 (finite game) *Let an ego K, a world W, an initial state r_0, and an initial valuation M_0 be given. A finite game for K, W, r_0, and M_0 is a finite sequence of developments $J_0, J_0', J_1, J_1', \ldots, J_k$ where $J_k \in \mathcal{J}^\star$ and*

$$\begin{array}{l} J_0 = \langle \{0\}, M_0, \{0 \mapsto r_0\}, \emptyset, \emptyset \rangle \\ J_i' = K(J_i) \textit{ for } 0 \le i < k \\ W(J_i', J_{i+1}) \textit{ for } 0 \le i < k \end{array}$$

If $J_k = \langle \mathcal{B}_k, M_k, R_k, \mathcal{A}_k, \mathcal{C}_k \rangle$ and $M_k \subseteq M$ where M is a complete valuation, then $\langle \mathcal{B}_k, M, R_k, \mathcal{A}_k, \mathcal{C}_k \rangle$ is a resulting development from the finite game.

Definition 14 (infinite game) *A infinite game for K, W, r_0 and M_0 (chosen as in the definition for a finite game) is an infinite sequence of developments $J_0, J_0', J_1, J_1', \ldots$ where each J_i and J_i' is a finite development, and*

$$\begin{array}{l} J_0 = \langle \{0\}, M_0, \{0 \mapsto r_0\}, \emptyset, \emptyset \rangle \\ J_i' = K(J_i) \textit{ for } i \ge 0 \\ W(J_i', J_{i+1}) \textit{ for } i \ge 0 \end{array}$$

If for every i, $J_i = \langle \mathcal{B}_i, M_i, R_i, \mathcal{A}_i, \mathcal{C}_i \rangle$, then a resulting development from the game is any development of the form

$$\langle \lim_{i \to \infty} \mathcal{B}_i, M, \lim_{i \to \infty} R_i, \lim_{i \to \infty} \mathcal{A}_i, \lim_{i \to \infty} \mathcal{C}_i \rangle$$

where for the purpose of taking the limit each R_i is considered as a set of maplets of the form $(t \mapsto r)$, M is a complete valuation, and $\lim_{i\to\infty} M_i \subseteq M$.

We can describe the game informally as follows. Given a state r_0 and an initial valuation M_0 we define the initial finite development $\langle\{0\}, M_0, \{0 \mapsto r_0\}, \emptyset, \emptyset\rangle$. The ego does the first move, deterministically transforming J_0 to $J_0' = K(J_0)$. Then the world makes its first move by transforming J_0' to any J_1 such that $W(J_0', J_1)$. Then the players alternate. After the game ends, arbitrary values are added to the remaining undefined temporal constant symbols.

Definition 15 (inhabited dynamical system (IDS)) *A dynamical system is inhabited iff it contains one or more agents which can influence the system's state at later times by performing actions or activities. An inhabited dynamical system is a pair (W, K) where W is a world and K is an ego, and where W and K use the same $\mathcal{R}$ and $\mathcal{E}$ domains.*

Definition 16 (scenario) *A scenario of a dynamical system is a set of possible developments in the system. Let W be an IDS world. A scenario for W is a set of games where W is the world player.*

2.5 The ontological taxonomy

In the structure of the world we can identify special cases with various restrictions. To distinguish clearly between the different cases we introduce the *ontological characteristics*. A summary of the ontological characteristics can be found in table 2.1.

The following definitions can help in the comprehension of the table. The starting (ending) state of an action is the state of the world when the action starts (ends). An action is *unconditional* iff the set of possible ending states for the action is independent of the starting state, and it is *conditional* otherwise. An action is *non-deterministic* iff the set of possible ending states for some starting states has more than one member. An action *lacks alternatives* iff it is deterministic and unconditional, and it *has alternatives* otherwise. In other words, if an action lacks alternatives then there is one single and the same ending state for the action, regardless of the starting state.

Definition 17 (ontological designator) *An ontological designator is a set of characteristics codes taken from the first column in table 2.1.*

We give in the following a brief description of the most important ontological designators.

The code letters **I** and **R** represent inertia in the case of discrete and continuous time, respectively. It is not necessary that the inertia is actual, it can also be just perceived.

The code letter **A** characterizes the case in which actions have different effects depending on the state of the world when the action started.

The ontological family **IA** is a combination of the characteristics of **I** and **A**, that is all the perceived features have constant values unless there is an action that changes them; each action only changes some specific features and in specific ways; there are no concurrent actions. This is an important case and most of Sandewall's work is done under these assumptions. It is also the classical case in A.I.

The code letters **U** and **D** represent dependences between features. The dependences can derive from dependences between phenomena in the material system that are reflected in the images. For example if we have three pipes A, B and C, and A and B give water to C, than the feature that specifies whether water is flowing in C is dependent on the values of the features that specify whether water is flowing in A and B. Another source of dependence is when there are features on several levels, and the value of a higher level feature is defined to be a function of the values of some lower level features. In local dependencies (**U**) the action can affect only features of objects that occur as arguments of the action. In structural dependences (**D**) the action can affect features of other objects, than those which occur as arguments of the action, or the effect of the action can depend on the starting-states features of objects that do not occur as arguments of the action.

We have a delayed effect (**L**) if an action has effects after the action itself is finished. For example the action of turning on the bathtub tap changes the feature for "tap open" from false to true, but also causes the feature "floor wet" to became true a while later.

2.6 Scenario description

In Knowledge Representation research within Artificial Intelligence several logic-methods have been proposed for reasoning about scenarios in commonsense domains and for obtaining the intended commonsense conclusions for each scenario. In this section we formalize the concept of scenario description and define the intended conclusions and developments of a given scenario description as ego/world interaction.

We give first an informal idea of the parts of a scenario description and than we give the formal definition. The parts of a scenario description are the following:

1. A description of the world where the scenario takes place. The description of the world comprehends the ontological family for the world and a number of *action laws* which characterize the effect of each action type. Therefore we describe a world as a tuple $\langle \psi, \Gamma_1, \Gamma_2, \ldots, \Gamma_k \rangle$ where ψ is an ontological descriptor and Γ_i are sets of action laws or other formal entities, structured according to the choice of ψ.

2. A statement about the *object domain*, i.e. the sets of objects which are involved in the scenario. The object domain contains only physical objects, not time points, numbers, names, and other similar abstract entities. We consider just *domain-specific* scenarios, i.e. scenarios where the object domain is finite, fixed for the duration of the scenario, and completely specified.

3. A statement about the *schedule* of the scenario, i.e. which actions take place and at what time.

4. A statement about the *observations* in the scenario, i.e. statements about features values at distinct points in time.

5. A *epistemological assumption* associated with the scenario description. The epistemological assumptions are similar to the ontological assumptions, but instead of characterizing the world, they characterize the whole scenario description. An example of epistemological assumption is the assumption of complete knowledge about the set of actions in the schedule.

The importance of the epistemological assumptions together with the ontological assumptions is that they make precise and they put on the meta-logical level the assumptions that, often implicitly, are made about developing a logic for reasoning about action and change. Stating explicitly the assumptions gives us a taxonomy of scenario descriptions and it is possible to identify, for each proposed logic its range of applicability, i.e. for which family of scenario descriptions in the taxonomy the logic obtains the correct conclusions.

An important assumption is that all actions are known with certainty. We call the scenarios that admit this assumption *chronicle*.

Definition 18 (epistemological characteristics) *An epistemological characteristic is one of the symbols (letter or pair of letters) listed in the first column of table 2.2.*

$\mathcal{K}$ is the standard case. $\mathcal{Q}$ allows actions which are only defined as possible for some starting states, whereas in $\mathcal{K}$ for each combination of action and starting state the effects of the action must be defined although possibly nondeterministically. $\mathcal{K}$ and $\mathcal{Q}$ allow just elementary actions, i.e. the schedule specifies each action in terms of action symbol, arguments, starting time, and ending time. $\mathcal{P}$ generalizes $\mathcal{Q}$ allowing composite actions which are formed by sequencing, conditional execution, and repetition.

$\mathcal{M}_o$ generalizes $\mathcal{K}$ allowing unreliable observations. We study this case in the following chapters.

Definition 19 (scenario description) *A scenario description is a tuple $\langle \mathcal{Z}, \Gamma_1, \Gamma_2, \ldots, \Gamma_k \rangle$ where $\mathcal{Z}$ is an epistemological characteristic that identifies the epistemological assumptions made about the scenario, and the following elements are sets of formulae or other formal objects. The number and types of the following elements are specific for each choice of $\mathcal{Z}$.*

An important case of scenario description is the scenario description in $\mathcal{K}$. The definition of scenario description in $\mathcal{K}$ is as follows.

Definition 20 (scenario description in $\mathcal{K}$) *A scenario description in K is a tuple $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS \rangle$ where*

- *$\mathcal{O}$ is a finite object domain,*
- *$Diw = \langle \mathbf{IA}, Law \rangle$ is a description of an IDS world where **IA** is the ontological designator that states strict inertia and alternative results for actions and Law is a set of action laws,*
- *SCD is a schedule specifying a number of actions and their temporal order,*
- *OBS is a set of observations of the values of features at some point in time.*

The definitions of scenario descriptions in $\mathcal{N}$ is as follows.

Definition 21 (scenario description in $\mathcal{N}$) *A scenario description in $\mathcal{N}$ is a tuple $\langle \mathcal{N}, \mathcal{O}, OBS \rangle$ where $\mathcal{O}$ is an object domain, and OBS is a set of observation.*

The designator $\mathcal{N}$ represents the absence of any epistemological assumptions, so any development where $\mathcal{O}$ is the object domain and all members on OBS are true is a model for $\langle \mathcal{N}, \mathcal{O}, OBS \rangle$. There is no schedule and no assumption of inertia.

What we would like to do is to conclude as much as possible about the values of fluents at different points in time. Such sets of conclusions will be referred to as "chronicle completion". Of course, our interest is in intended completions, i.e. those completions containing desirable conclusions constrained by an assumption of inertia.

2.6.1 Example of scenario description in $\mathcal{K}$

The example that we consider is the following:

$$\begin{array}{ll} obs1 & [0]on(glass, table). \\ obs2 & [4]on(glass, shelf). \\ law1 & [t,s]Move(glass, table, shelf) \Rrightarrow \\ & \quad [t]on(glass, table) \rightarrow [t,s]on(glass, shelf){:=}T \\ & \quad \wedge [t,s]on(glass, table){:=}F. \\ scd1 & [2,3]Move(glass, table, shelf). \end{array} \tag{2.1}$$

This represents a scenario description $\langle \mathcal{K},\ \mathcal{O},\ \langle \mathbf{IA}, Law \rangle,\ SCD,\ OBS \rangle$ where the the set of action laws Law contains the formula labeled $law1$. This formula specifies exactly the action of moving a glass from the table to the shelf. If the glass is on the table when the action is performed, then the glass will be on the shelf at the end of the action. If the glass is not on the table, than the action will have no effect. One important aspect of the $\mathcal{K}$ epistemological assumption is that the actions are exactly specified.

The observation $obs1$ says that at the time point 0 the glass is on the table. In the schedule there is the action of moving the glass from the table to the shelf and the observation $obs2$ confirms that the glass is on the shelf at the time point 4.

The action instances in the schedule are expanded accordingly with the action law, for example the action $[2,3]Move(glass, table, shelf)$ is expanded with:

$$[2]on(glass, table) \rightarrow [2,3]on(glass, shelf){:=}T \wedge [2,3]on(glass, table){:=}F.$$

$[t,s]on(glass, shelf){:=}T$ and $[t,s]on(glass, table){:=}F$ are reassignment formulae stating that the feature "on(glass,shelf)" will get the value "T" and the feature "on(glass,table)" will get the value "F" somewhere in the interval $(t,s]$.

In general the information that is contained in the scenario is incomplete, but we can deduce some of the missing information considering the ontological assumption of inertia $\mathbf{I}$. For example the action law $law1$

requires that the glass is on the table at 2 to get the effect that the glass is on the shelf after the moving action. It can be derived from $obs1$ and the ontological assumption **I** that this condition holds; in fact as the changes of feature-value are only due to the actions in the schedule and there are no actions in the schedule that can change the position of the glass from time point 0 to time point 2, it can be derived that the glass is on the table at time point 2.

The assumption **A** says that the actions can have alternative result. This is expressed in the action laws in this form:

$$[s,t]A \Rrightarrow ([s]\delta_1 \rightarrow [s,t]\alpha_1) \wedge ([s]\delta_2 \rightarrow [s,t]\alpha_2) \wedge ...$$

That is, if δ_1 holds the action has $[s,t]\alpha_1$ as effect, if δ_2 holds the action has $[s,t]\alpha_2$ as effect and so on. We call $[s]\delta_1$, $[s]\delta_2$,... the *conditions* of an action and $[s,t]\alpha_1$, $[s,t]\alpha_2$,... the *effects* of the action.

In the previous example the condition to get the effect of having the glass on the shelf after the moving action is that the glass was on the table at the time point of the starting of the action.

In the case that any of the conditions for an action in the schedule are satisfied the action it is considered done, but without having any effect.

2.6.2 Model sets for scenario descriptions

Definition 22 (Models for a scenario description) *For a given scenario description* $\boldsymbol{\Upsilon}$ *and time domain* $\mathcal{T}$*,* $Mod(\boldsymbol{\Upsilon})$ *denotes the set of complete developments which are correctly described by* $\boldsymbol{\Upsilon}$*. For each* $m \in Mod(\boldsymbol{\Upsilon})$*, m is a model of* $\boldsymbol{\Upsilon}$*.*

Let a chronicle $\langle \mathcal{O}, Diw, SCD, OBS \rangle$ be given, then the set of *intended models* of this chronicle is obtained as follows.

First it is selected an arbitrary world W which is exactly described by A, an arbitrary ego, an arbitrary initial state and an arbitrary initial valuation. Then are generated all possible developments which can be obtained in games between them. Finally the set of developments are restricted to those where all formulas in $Law(SCD) \cup OBS$ are satisfied, and where there is a one-to-one correspondence between actions in the $\mathcal{A}$ component of the development and action statements in SCD. $Law(SCD)$ represent the result of replacing each action statement in SCD by the effects specified by the laws in Law.

The set of intended models for the given chronicle are the set of interpretations $\langle M, R \rangle$ obtained extracting the M and the R components from the remaining developments.

The set of intended models for a chronicle Υ is denoted by

$$\sum_{IA}(\langle \mathcal{O}, A, SCD, OBS\rangle) = \{\langle M, R\rangle | \langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle \in Mod(\Upsilon)\}.$$

If the knowledge in the scenario description is incomplete the scenario has several models in which the knowledge is completed in all the possible ways that respect the principle of inertia.

If the observations conflict with the necessary conditions of the actions, the scenario description has still models; in these models the actions are performed, but without effects.

2.7 Entailment relations

The *entailment relations* give the possibility to express several common varieties of temporal reasoning. We need the *prediction* operator in the next chapter and this is the reason why we introduce here the concept of entailment relation.

Sandewall defines three entailment relations and shows how a number of reasoning operations and reasoning problems can be expressed in term of these entailment relations. We consider here just one entailment relation and how the prediction operator can be expressed in terms of it.

Definition 23 (entailment relation) *Let Υ and Υ' be scenario descriptions. The entailment relation is defined as follows:*

$$\Upsilon \Vdash \Upsilon' \Leftrightarrow Mod(\Upsilon) \subseteq Mod(\Upsilon')$$

Given a scenario description $\Upsilon = \langle \mathcal{K}, \mathcal{O}, \langle \mathbf{IA}, Law\rangle, SCD, OBS\rangle$, the prediction operator allows to infer if a proposition p is true, false or unknown at the ending time t of a sequence of actions in Υ. This can be formulated in terms of entailment relation as follows:

p is true if holds:

$\Upsilon \Vdash \langle \mathcal{N}, \mathcal{O}, \{[\mathbf{t}]p\}\rangle$

p is false if holds:

$\Upsilon \Vdash \langle \mathcal{N}, \mathcal{O}, \{[\mathbf{t}]\neg p\}\rangle$

and p is unknown if neither of the two holds.

2.8 Conclusion

In this chapter we have introduced the concept of inhabited dynamical system and in particular we have considered the case in which a single agent interacts with a material system. The agent receives informations from the material system by means of observations and acts in the material system by means of actions.

A scenario description represents a set of games between ego and world, and contains such things as actions and observations of the world state. The concept of scenario description is quite important in the rest of the work, in fact we use this concept as base to define the concept of hypothetic scenario description, that is a scenario description that can deal with incorrect observations.

Finally we have defined the models of a scenario description as complete developments, that is ego/world interaction in form of game.

Code	Ontological characteristics	Absence
I	Inertia without alternatives, surprises, concurrency, etc., and discrete "integer time". (None of the following arises).	Disorder
R	Like **I** but for continuous time.	Disorder
A	Alternative results (of actions). (*Nondeterministic or conditional result*).	Single result
U	Local ramification – dependencies between features of the objects directly involved in an action. (*Change in one feature implies possibility of immediate change in another feature*).	Local Independence
D	Structural ramification – dependencies with features of objects which are only indirectly in an action. (*Change in one feature implies possibility of immediate change in another feature*).	Structural independence
L	Delayed effect (on features). (*Changes due to an action may occur at a later time*).	Inertia
C	Concurrency (of actions). (*Concurrency can't be excluded*).	Sequentiality
S	Surprises (in features). (*Additional changes except those that are due to the previous characteristics are possible but occur only infrequently*).	Predictability
N	Normality (in features). (*Certain states or partial states of the world are stated to be more frequent than others. This may apply to the initial state or to the result of various actions or other occurrences*).	Equinormality
H	Hierarchical occurrences. (*Actions/subactions, event/subevent*	Flat occurrences
T	Timebound objects. (*Actions or events for the creation and/or the termination of objects occur in the system*).	Timeless objects
M	Memory. (*Result of action may depend on feature values at earlier times than the start of the action*).	No memory

Table 2.1: Ontological characteristics

code	Epistemological characteristics	Absence
$\mathcal{N}$	No epistemological assumptions.	
$\mathcal{K}$	Chronicle with explicit, correct and accurate knowledge. *No partiality, misperception, or illusion*	
$\mathcal{Q}$	Chronicle with qualified action laws. *For some actions there are preconditions such that the effect of the action is known when preconditions are satisfied, but unknown otherwise*	Unqualified
$\mathcal{P}$	Chronicle with qualified action laws and composite actions. *Action sequences, conditional actions, and iteration are allowed in the schedule.*	Non-composite
$\mathcal{U}_a$	Uninformed about actions. *Some of the instances of actions remain unspecified.*	Informed
$\mathcal{U}_e$	Uninformed about events. *Sometimes fails to observe or infer event.*	Informed
$\mathcal{U}_l$	Uninformed about actions laws. *Don't always know the effects of actions, or even what features they effect.*	Informed
$\mathcal{M}_o$	Misperception (observations). *Sometimes observes the incorrect feature value.*	Observant
$\mathcal{M}_e$	Misperception (events). *Sometimes vehicle reports a non-occurring event.*	Observant
$\mathcal{I}_a$	Illusion (of action or activities). *Sometimes vehicle believes that it performs an action that it actually does not.*	Illusion free

Table 2.2: Epistemological characteristics

Chapter 3

Introduction of unreliable observations in the features and fluents framework

3.1 Introduction

In this chapter we consider the case in which is possible to have misperception in observations, allowing for occasional incorrect observations of feature values. This case is designated, in the Sandewall taxonomy, with the epistemological characteristics $\mathcal{M}_o$.

The problem in the case of misperception is that we cannot detect in general if an observation is incorrect, otherwise the solution would be just to eliminate it. Only in case an incorrect observation contradicts some other observation or the rest of the scenario can we detect the presence of the incorrect observation, but we can not in general detect which observation is incorrect.

In particular, there are three cases in which we can detect if there are incorrect observations:

1. There is a conflict between the observations. This happens if there are contradictory observations at the same time point, or if observations at different time points conflict because of the principle of inertia. This last case happens when new observations give information about changes in the world that are not justified by any action in the schedule. As it is assumed that only actions can change the world, this conflict of observations is ascribed to the presence of incorrect observations, even if there is no possibility in general to know which of the observations is incorrect. An example of this first case is the scenario (3.1) where we have

two contradictory observations at the same time point.

$$\begin{array}{ll} obs1 & [0]on(glass, table). \\ obs2 & [0]\neg on(glass, table). \end{array} \tag{3.1}$$

Scenario (3.2) is the case in which two observations contradict each other because of the principle of inertia. The glass is on the table at time point 0 and it is not on the table at time point 2 without any action that can justify this change.

$$\begin{array}{ll} law1 & [t,s]Move(glass, table, shelf) \Rrightarrow \\ & \quad [t]on(glass, table) \rightarrow [t,s]on(glass, shelf) {:=} T \\ & \quad \wedge [t,s]on(glass, table) {:=} F. \\ obs1 & [0]on(glass, table). \\ obs2 & [2]\neg on(glass, table). \\ scd1 & [2,3]Move(glass, table, shelf). \end{array} \tag{3.2}$$

2. The conditions of the action are in accordance with the observations but the observations do not agree with the effects that the actions are supposed to have given these conditions. This means, as we assume that the actions perform as they are supposed to do, that there are incorrect observations either in the preconditions or in the postconditions. In the following example the condition $[2]on(glass, shelf)$ is satisfied. In fact, $obs1$ says that the glass is on the table at 0 and there are no actions that move it between 0 and 2, but the action does not seem to have the effect that it is supposed to have. Instead, $obs2$ says that the glass is not on the shelf.

$$\begin{array}{ll} law1 & [t,s]Move(glass, table, shelf) \Rrightarrow [t]on(glass, table) \rightarrow \\ & \quad [t,s]on(glass, shelf) {:=} T \wedge [t,s]on(glass, table) {:=} F. \\ obs1 & [0]on(glass, table). \\ scd1 & [2,3]Move(glass, table, shelf). \\ obs2 & [4]\neg on(glass, shelf). \end{array} \tag{3.3}$$

3. Given a scenario description $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS \rangle$ one observation $obsj = [t]f$ conflicts with the schedule if

$$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{\} \rangle \vdash \langle \mathcal{N}, \mathcal{O}, \{[t]\neg f\} \rangle.$$

The scenario description (3.4) is an example of an observation, $obs1$, which conflicts with the schedule. In fact, if the glass is on the table at time point 4, it was on the table also at time point 2 as a consequence of inertia and the action of moving it on the shelf should have been successful. However this implies that the glass should not be on the table anymore at time point 3 and, as there are no other actions, it should not be on the table at time point 4 either.

$$\begin{array}{ll} law1 & [t,s]Move(glass,table,shelf) \Rrightarrow [t]on(glass,table) \to \\ & \quad [t,s]on(glass,shelf){:=}T \wedge [t,s]on(glass,table){:=}F. \\ scd1 & [2,3]Move(glass,table,shelf). \\ obs1 & [4]on(glass,table). \end{array} \tag{3.4}$$

In case of observations that conflict with the schedule we can say that these observations are surely incorrect as we assume that the actions have always the effect that are supposed to have and we can eliminate them.

In the other two cases we know that there are incorrect observations, but we do not know which observations are incorrect. We then consider all the possible alternative complete developments in which as many as possible of the observations are true. This is the idea behind the definition of models in $\mathcal{M}_o$ that we give in section 3.2.

As a way of constructing models for $\mathcal{M}_o$ we introduce the concepts of hypothetic scenario description, narrative and revision.

The concept of *hypothetic scenario description* is introduced in section 3.3. This is a scenario description in which there is a number of sets of observations, instead of just one, representing alternative histories of the world.

In the hypothetic scenario description we have the assumptions present in $\mathcal{K}$ except the assumption that the observations are all reliable. As ontological assumptions we have **I** and **A**. The principle of inertia is therefore valid, and the action can have alternative results.

In Sandewall's work the schedule SCD contains all the actions that should be done and the set OBS contains all the observations from the first to the last time point. In case of incorrect observations, it is useful not to have all the observations from the first moment, but to consider the process of getting information time point after time point. We therefore introduce in section 3.3 the concept of *narrative*, that is, a mapping from time points to hypothetic scenario descriptions. For each time point there is a corresponding hypothetic scenario description that considers the observations until that time point.

A set of new observations is added at each time point to the old observations. If the hypothetic scenario description obtained by adding the new observations is correct, so that for each history of the world there is a sequence of actions and an assignment to the features that is coherent with the scenario, then we continue with the next time point.

The hypothetic scenario description is not correct if there are observations which make the scenario inconsistent. This happens in the three cases we have considered before.

If the hypothetic scenario description in a narrative that corresponds to the last time point, intuitively the *now*, is not *correct*, that is there are some incorrect observations in it, then we say that the narrative is *faulty* and in this case we do a *revision* of the narrative, as described in section 3.3, that transforms this narrative into a correct one.

In section 3.4 we define the models for hypothetic scenario descriptions and we give a constructive definition of models for $\mathcal{M}_o$.

In section 3.5 we consider the possibility of doing the revision not every time it is necessary, but to wait some time points. In theorem 3.5 it is proved that doing a *delayed revision* gives results similar to the ones that can be obtained doing the revision every time it is necessary.

As a consequence of this theorem we prove that the constructive and the declarative definition of models for $\mathcal{M}_o$ are equivalent.

In section 3.6 we examine several cases of preferential revision, that is, revisions in which certain states of the world are preferred over others and we give the definition of preferred models for $\mathcal{M}_o$.

In section 3.7, finally, a preference relation for models is defined and a link is established (theorem 3.7) between this model preference and one of the observation preferences that are described in section 3.6.

3.2 Syntax and models for scenario descriptions in $\mathcal{M}_o$

The definition of scenario descriptions in $\mathcal{M}_o$ is similar to the definition of scenario descriptions in $\mathcal{K}$.

Definition 24 (scenario descriptions in $\mathcal{M}_o$) *A scenario descriptions in $\mathcal{M}_o$ is a tuple:*

$$\langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$$

where $\mathcal{O}$ is an object domain, Diw is a description of an IDS world $\langle \mathbf{IA}, Law \rangle$, SCD is a schedule specifying a number of actions and their temporal order, and OBS is a set of observations of the values of features at specific points in time.

We define the intended models for $\langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$ as all the possible alternative complete developments that satisfy the rest of the scenario besides the observations and in which the maximal part of observations is true.

Definition 25 (Models for scenario descriptions in $\mathcal{M}_o$) (declarative definition) *Given a scenario description $\mathbf{\Upsilon} = \langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$, $Mod(\mathbf{\Upsilon})$ is the set of those members of $Mod(\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \emptyset \rangle)$ that maximize the subset of OBS that is satisfied.*

In other words $Mod(\mathbf{\Upsilon})$ is the union of the model sets of the scenario descriptions $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS_i \rangle$ for $i \in \{1, ..., n\}$, where $OBS_1, ..., OBS_n$ are the maximal subsets of OBS obtaining non-empty model sets.

This is a declarative definition of models. The aim of the following sections is to define a constructive way to obtain these models.

3.3 Narrative and revision

As we have anticipated in the previous section, a possible way of dealing with the problem of incorrect observations is to admit that we do not know which observations are incorrect, and therefore to consider several alternative representations of the world. Now we introduce some definitions that help us to formalize this idea.

First we give the definition of hypothetic scenario description. This definition is similar to the one of scenario description, but instead of just one set of observations, we have several sets of observations that represent different alternative histories of the world.

Definition 26 (hypothetic scenario description) *A hypothetic scenario description is a tuple:*

$$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, \ldots, OBS_n\} \rangle$$

where $\mathcal{O}$ is an object domain, Diw is a description of an IDS world $\langle \mathbf{IA}, Law \rangle$, SCD is a schedule specifying a number of actions and their temporal order, and every OBS_i is a set of hypothetic observations of the values of features at some points in time.

Now we can define the concept of narrative. The idea behind this definition is that to each time point corresponds a hypothetic scenario description in which we have the observations that we have got until that time point. At each time point t_i we get a set of new observations that we call $NewObs_i$ and that we add to our set of observations.

Definition 27 (narrative) *A narrative is a mapping from* $\{t_0, \ldots, t_k\}$ *to hypothetic scenario descriptions which have* $\mathcal{K}$, $\mathcal{O}$, Diw *and* SCD *in common and if* t_i *maps to* $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^i, \ldots, OBS_n^i\}\rangle$ *and* t_{i+1} *maps to* $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^{i+1}, \ldots, OBS_m^{i+1}\}\rangle$ *and* $NewObs_{i+1}$ *is the set of new observations at time* t_{i+1}*, then for all* $j \in \{1, \ldots m\}$ *there exists an* $s \in \{1, \ldots n\}$*:* $OBS_j^{i+1} \subseteq OBS_s^i \cup NewObs_{i+1}$*.*

For simplicity, we suppose that there are no observations at t_0. This is not a limitation. In fact, if we have a narrative in which there are observations at t_0, we can add at the beginning of the narrative a new mapping from a time point, say t_0', to a hypothetic scenario description equal to the one we had at t_0, but with no observations.

We denote by $N(t_p)$ the hypothetic scenario description that corresponds to the mapping to the time point t_p. The time point t_k is called the *last time point of the narrative*, and the hypothetic scenario description that corresponds to the last time point of the narrative is called *last step of the narrative*.

A hypothetic scenario description is correct when all the scenario descriptions that can be obtained substituting in the hypothetic scenario description the sets of observations with one set of observations have models.

Definition 28 (correct hypothetic scenario description) *A hypothetic scenario description* $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, ..., OBS_n\}\rangle$ *is correct if for all* OBS_k*,* $\sum_{IA}(\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS_k\rangle) \neq \emptyset$*.*

A set of observations OBS_k is *correct for a narrative* if $\sum_{IA}(\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS_k\rangle) \neq \emptyset$, where $\mathcal{K}$, $\mathcal{O}$, Diw, SCD is the part that is common to all the hypothetic scenario descriptions of the narrative.

We say that an observation $obsj = [t]f$ is *contradicts a narrative* if

$$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \emptyset\rangle \vdash \langle \mathcal{N}, \mathcal{O}, \{[t]\neg f\}\rangle.$$

where $\mathcal{K}$, $\mathcal{O}$, Diw, SCD is the part that is common to all the hypothetic scenario descriptions of the narrative.

Given the definition of correct hypothetic scenario description we define correct, completely correct and faulty narrative.

Definition 29 (correct narrative) *A correct narrative is a narrative in which the hypothetic scenario description that corresponds to the last time point* t_k *of the narrative is correct.*

Definition 30 (faulty narrative) *A faulty narrative is a narrative in which the hypothetic scenario description that corresponds to the last time point* t_k *of the narrative is not correct.*

Definition 31 (completely correct narrative) *A completely correct narrative is a narrative in which all the hypothetic scenario descriptions of the narrative are correct.*

The reason why the correct and the faulty narrative are defined in terms of thelast step of the narrative is that it represents the "now" and the present alternatives histories of the world.

Given a faulty narrative a revision is a function that transforms it into a correct narrative.

Definition 32 (maximal consistent subset) *A set of observations OBS' is a maximal consistent subset of a set of observations OBS relative to a scenario $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \emptyset \rangle$ if $OBS' \subseteq OBS$, $\Sigma_{IA}(\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS' \rangle) \neq \emptyset$, and for any OBS'' such that $OBS' \subset OBS'' \subseteq OBS$, $\Sigma_{IA}(\langle \mathcal{K}, \mathcal{O}, Diw, SCD, OBS'' \rangle) = \emptyset$.*

Definition 33 (revision) *Let $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, ..., OBS_n\} \rangle$ be the last step of a faulty narrative N. A revision is a function Rev(N) which maps the faulty narrative N into the narrative which is equal to N, except for the final scenario which has this form: $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1', \ldots, OBS_m'\} \rangle$ where $\{OBS_1', ..., OBS_m'\}$ contains for every element in $\{OBS_1, ..., OBS_n\}$ exactly the maximal consistent subsets of this element relative to $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \emptyset \rangle$.*

Theorem 1 *Rev(N) always returns a correct narrative and this narrative is unique.*

Proof. There always exist maximal correct subsets of each set of observations in the hypothetic scenario. In the worst case the only maximal correct subset is the empty set. When substituted in the last scenario description, these subsets give a correct narrative. Therefore, Rev(N) gives back a correct narrative.

The narrative that the revision gives back is unique. In fact, suppose that there exist two different narratives, N and M, that the revision can give back. These narratives can only differ, for definition of revision, in the last hypothetic scenario description, $N(t_k)$ and $M(t_k)$, and in particular in the sets of observations A of these scenarios. Consider a set of observations that is present in $M(t_k)$ and not in $N(t_k)$. As it is a set of observations for which the hypothetic scenario description is correct, it should be properly included in one of the sets of observations in $N(t_k)$. In fact, $N(t_k)$ contains all the maximal subsets of our original set of observations for which the hypothetic scenario description is correct. But also the set A of observation in $M(t_k)$ is a maximal subset of our original sets of observations for which the hypothetic scenario description is correct. A similar proof shows that if a set of observations is in $N(t_k)$ it is also in $M(t_k)$.

The set of maximal subsets is uniquely determined. Therefore, $M(t_k)$ and $N(t_k)$ are equal.

■

A narrative can be constructed as follows:

- First we have the narrative that only contains the mapping from time point t_0 to the hypothetic scenario description $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{\emptyset\} \rangle$.

- If we have a narrative whose last time point is t_k, in order to add another mapping element

$$t_{k+1} \rightarrow \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^{k+1}, \ldots, OBS_n^{k+1}\} \rangle$$

  we add the new observations that we get at t_{k+1} to every OBS_1^i,..., OBS_n^i and if we get a faulty narrative we do a revision.

We use the notation $Add(N, NewObs)$ for the modified narrative where $NewObs$ has been added (by set union) to every OBS_1^i,...,OBS_n^i of the last step of the narrative.

Constructing a narrative in this way, we obtain a completely correct narrative. In the section 3.5 we consider the case in which delayed revision is done. In this case, the narrative that is obtained is correct, but not completely correct.

3.4 Models of hypothetic scenario description and constructive definition of models in $\mathcal{M}_o$

The models of a scenario description are those complete developments which are compatible with the scenario. If instead of one set of observations, we have several sets of observations which models do we have?

We define as models of the hypothetic scenario description the union of the models that we get in the scenario descriptions which contain just one of the sets of observations of the hypothetic scenario description.

Definition 34 (Models of hypothetic scenario description) *If S = $\langle\mathcal{K},\mathcal{O},Diw,SCD,\{OBS_1,\ldots,OBS_n\}\rangle$ is a hypothetic scenario description and if*

$$Model_i = \sum_{IA}(\langle\mathcal{K},\mathcal{O},Diw,SCD,OBS_i\rangle)$$

for all $i \in \{1,\ldots n\}$, then the set of models of S, Mod(S), is the union of $Model_i$ for all $i \in \{1,\ldots n\}$.

Let us see intuitively what this definition means when applied to hypothetic scenario descriptions in a narrative with the help of the example in figure 3.1. The narrative that we consider is:

$t_0 \rightarrow \langle\mathcal{K},\mathcal{O},Diw,SCD,\{\emptyset\}\rangle$

$t_1 \rightarrow \langle\mathcal{K},\mathcal{O},Diw,SCD,\{A,B\}\rangle$

$t_2 \rightarrow \langle\mathcal{K},\mathcal{O},Diw,SCD,\{A,B,C\}\rangle$

$t_3 \rightarrow \langle\mathcal{K},\mathcal{O},Diw,SCD,\{\{A,B,C\},\{A,B,\neg C\}\}\rangle$

We indicate in the figure the model that satisfies some observations as a circle containing the observations that it satisfies.

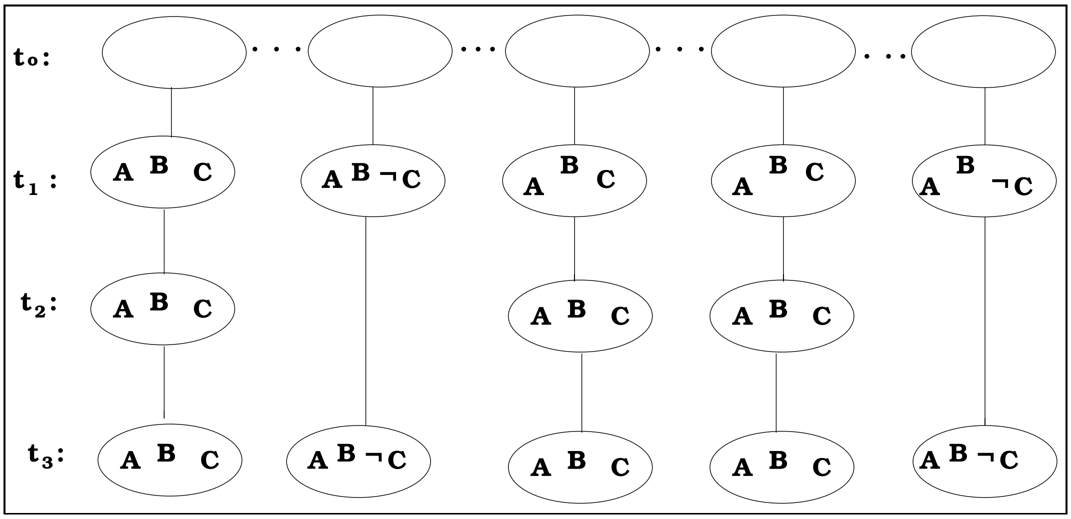

Figure 3.1: Example of models of hypothetic scenario description

In this example, at time t_0 we have the models that satisfy the schedule and the principle of inertia. The observations at t_1 are A and

B and in the figure we see the models that satisfy the schedule, the principle of inertia, and also these observations. Some of these models satisfy C as well and some do not. At t_2 we get the observation C and so the only models left are the ones for which C is true, assuming that there was not an action in the schedule requesting $\neg C$ at t_2. At t_3 we get the observation $\neg C$ and, as the set of observations obtained by adding $\neg C$ is inconsistent, we split the set into two sets: $\{A, B, C\}$ and $\{A, B, \neg C\}$. The models of the hypothetic scenario description that we get doing this splitting are the ones that satisfy $\{A, B, C\}$ and the ones that satisfy $\{A, B, \neg C\}$. In a sense we make the sets that we had thrown away at t_1 alive again.

In the next theorem we see that this intuitive view of models of hypothetic scenario descriptions in a narrative corresponds effectively to our formal definitions.

Theorem 2 *Let N be a narrative.*

(a) The models of the hypothetic scenario description at t_0 are all the models that satisfy the schedule and which preserve inertia.

(b) If at t_{i+1} we do not do revision, the models of the hypothetic scenario description at t_{i+1} are a subset of the models of the hypothetic scenario description at t_i.

(c) If at t_{i+1} we do revision, each of the models of the hypothetic scenario description that we obtain after the revision was also model of at least one previous scenario description.

Proof.

(a) As we have supposed that the set of observations at t_0 is empty the hypothetic scenario description at t_0 is really a scenario description in which the set of observations is empty. For the definition of models of scenario descriptions, the models of this scenario description are all the models that satisfy the schedule and the principle of inertia.

(b) As at t_{i+1} we do not do revision, the sets of observations at t_{i+1} are the sets of observations at t_i to which we have added the new observations $NewObs_{i+1}$. Let OBS_j^{i+1} be a set of observations at t_{i+1}. As $OBS_j^{i+1} = OBS_q^i \cup NewObs_{i+1}$ where OBS_q^i is a set of observations at t_i, the models that satisfy OBS_j^{i+1} satisfy also OBS_q^i and so they are also models of the hypothetic scenario description at t_i. Repeating this proof for all the sets of observations at t_{i+1}, we get the proof of our proposition.

(c) As all the models of the hypothetic scenario descriptions in the narrative satisfy the schedule and the principle of inertia, they are all also models of the hypothetic scenario description at t_0.

■

We can now give a constructive definition of models for scenario descriptions in $\mathcal{M}_o$.

Definition 35 (Models for scenario descriptions in $\mathcal{M}_o$) (constructive definition)

Given a scenario description $\boldsymbol{\Upsilon} = \langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$, *where* t_f *is the maximal time point in the observations in* OBS, *if* N *is the narrative obtained adding at each time point* t *the observations in* OBS *that have* t *as time point and doing the revision every time it is necessary, then* $Mod(\boldsymbol{\Upsilon})$ *are the models of the hypothetic scenario description corresponding to* t_f *in* N.

In section 3.5.1 we prove that this constructive definition of models with relation to $\mathcal{M}_o$ is equivalent to the declarative definition of models $\mathcal{M}_o$ that we gave in section 3.2.

3.5 Delayed revision

Since it can be expensive to check consistency and to do revision, we consider the possibility of delaying the revision for some time points. If instead of checking whether a narrative is faulty we would go on developing the narrative, just adding the new observations at each time point and doing the revision at time point k, does then the revision at time point k give the same result as we would have obtained at k if we had done the revision every time it was necessary? The following theorem proves that we obtain the same scenario description at time point k.

Theorem 3 *Let* N *be a faulty narrative and let NewObs1 and NewObs2 be two sets of observations. It is possible to skip the revision between two additions of observations and to get the same hypothetic scenario description as last step,* t_k *of the narrative:*

$$Rev(Add(Add(N, NewObs1), NewObs2))(t_k) =$$
$$Rev(Add(Rev(Add(N, NewObs1)), NewObs2))(t_k)$$

Proof.

For simplicity let:

- $T = Rev(Add(Add(N, NewObs1), NewObs2))$ and let

$$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^T, \ldots, OBS_n^T\} \rangle$$

  be its last step.

- $M = Rev(Add(Rev(Add(N, NewObs1)), NewObs2))$ and let

  $$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^M, \ldots, OBS_m^M\}\rangle$$

  be its last step.
- $S = Rev(Add(N, NewObs1))$ and let

  $$\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1^S, \ldots, OBS_s^S\}\rangle$$

  be its last step.

We want to prove that for each $t \in \{1, \ldots n\}$ there exists a $p \in \{1, \ldots m\}$ such that $OBS_t^T = OBS_p^M$ and for each $p \in \{1, \ldots m\}$ there exists a $t \in \{1, \ldots n\}$ such that $OBS_p^M = OBS_t^T$.

$\forall t \exists p\, (OBS_t^T = OBS_p^M)$: By definition of revision for each OBS_t^T there exists an OBS_i in the last step of $Add(N, NewObs1)$ such that OBS_t^T is one of the maximal subsets of $OBS_i \cup NewObs2$ which preserves the correctness of the narrative. Let OBS^S be a generic set that we get doing the revision from OBS_i.

If we prove that

$$OBS_t^T \subseteq OBS^S \cup NewObs2 \subseteq OBS_i \cup NewObs2$$

we prove also that OBS_t^T is one of the maximal subsets of $OBS^S \cup NewObs2$ which preserves the correctness of the narrative. In fact, there exists a maximal subset of $OBS_i \cup NewObs2$ which preserves the correctness of the narrative and it is included in $OBS^S \cup NewObs2$ which is a subset of $OBS_i \cup NewObs2$. But, as the OBS_p^M are the maximal subsets of $OBS^S \cup NewObs2$ for a generic OBS^S, which preserves the correctness of the narrative, OBS_t^T is one of them.

$OBS^S \cup NewObs2 \subseteq OBS_i \cup NewObs2$ by definition of revision.

We need only to prove that $OBS_t^T \subseteq OBS^S \cup NewObs2$.

$OBS_t^T \subseteq OBS_i \cup NewObs2$, so OBS_t^T can be partitioned into a part Q which is a subset of OBS_i and a part P which is a subset of *NewObs2*.

As OBS_t^T is correct for the narrative Q is also correct for the narrative, and as Q is included in OBS_i and correct it must be included in OBS^S which is a set that we get doing the revision from OBS_i.

OBS_t^T is included in $OBS^S \cup NewObs2$ since the Q part of it is included in OBS^S and the P part is included in $NewObs2$.

$\forall p \exists t\,(OBS_p^M = OBS_t^T)$: According to the definition of revision there exists an OBS_i^S such that $OBS_p^M \subseteq OBS_i^S \cup NewObs2$.

As we have proven before, for each OBS_t^T there exists a generic OBS^S such that $OBS_t^T \subseteq OBS^S \cup NewObs2$.

Let us consider those OBS^T which are included in $OBS_i^S \cup NewObs2$. OBS_p^M is one of them. In fact, OBS_p^M is a maximal subset of $OBS_i^S \cup NewObs2$, which preserves the correctness. And, the OBS^T are all the maximal subsets, included in $OBS_i^S \cup NewObs2$, which preserves the correctness.

■

Corollary 1 *Let N be a narrative. If the revision is skipped between n additions of observations in N the same hypothetic scenario description is obtained as a result of the revision after the n additions as if the revision would have been done every time it was necessary.*

3.5.1 Equivalence of declarative and constructive definition of models for scenario description in $\mathcal{M}_o$

In sections 3.2 and section 3.4 we have defined the models for $\mathcal{M}_o$ in a declarative and in a constructive way. As a consequence of the theorem just proven we can prove in the following theorem that the models obtained by the two definitions are identical.

Theorem 4 *The models for scenario description in $\mathcal{M}_o$ obtained by the declarative definition are the same as the models obtained by the constructive definition.*

Proof. Lets a scenario description $\mathbf{\Upsilon} = \langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$ be given. Following the declarative definition of models $Mod(\mathbf{\Upsilon})$ are the models of the hypothetic scenario description S = $\langle \mathcal{K}$, $\mathcal{O}$, Diw, SCD, $\{OBS_1, \ldots, OBS_n\}\rangle$. But S is identical, according to corollary 1, to the hypothetic scenario description at the last step of the narrative N in the constructive definition of models for scenario description in $\mathcal{M}_o$. In fact $\mathbf{\Upsilon}$ can be considered as the last step of a narrative obtained adding the observations without doing revision and S the hypothetic scenario description obtained doing the revision at the very last step.

■

3.6 Preferential revision

When we do a revision of a narrative we can get several sets of observations. Instead of taking all the sets, we can make a choice and take just some of them. This is the idea behind the definition of preferential revision.

Definition 36 (preferential revision) *Let N be a narrative, N' the narrative obtained doing the revision of N and $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, ..., OBS_n\}\rangle$ the last step of N'. A preferential revision returns a narrative equal to N' except for the last hypothetic scenario description which has the form $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS'_1,...,OBS'_m\}\rangle$, where $\{OBS'_1,...,OBS'_m\} \subseteq \{OBS_1,...,OBS_n\}$.*

Often $\{OBS'_1,...,OBS'_m\}$ is chosen as $Min(<, \{OBS_1, ..., OBS_n\})$, where $<$ is a preference relation on observation sets.

The preference relation should depends mostly upon the application we are dealing with. We can anyway give some general criteria of preference:

- we prefer the last observations;
- we prefer the first observations;
- we have a different belief in every observation and we prefer the ones in which we believe more;
- our belief in the observations is context-dependent.

3.6.1 Preference for the last or the first observations

Let us consider the first of our criteria of choice, i.e. the case in which we prefer the last observations.

A way to prefer the last observations is to take, when we do the revision, just the maximal subsets that contain all the new observations. This is expressed more formally in the following definitions:

Definition 37 (Last observation preference) *Let $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, \ldots, OBS_n\}\rangle$ be the last hypothetic scenario description in a narrative and let NewObs be the new observations that we have added at this step. A last observation preference is a function which maps S to the hypothetic scenario description $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS'_1, \ldots, OBS'_m\}\rangle$ where $\{OBS'_1,...,OBS'_m\}$ contains all the OBS_i in $\{OBS_1,...,OBS_n\}$ such that $NewObs \subseteq OBS_i$.*

Definition 38 (Last observation preferential revision) *Let* $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, \ldots, OBS_n\} \rangle$ *be the last hypothetic scenario description of a faulty narrative N. A last observation preferential revision is a function which maps the faulty narrative N in the unique correct narrative which is equal to N, except for the final scenario which is obtained first doing the revision and then doing the last observation preference.*

This preferential revision has the advantage that the choice of the preferred subset is very easy: one only needs to check a set inclusion.

A disadvantage is that we must assume that the set of the new observations is consistent. This can be reasonable, but of course it is a limitation.

Another thing that we can notice is that we do not consider the order in which the observations had arrived in the successive revisions. We will explain this better with the help of the figure 3.2.

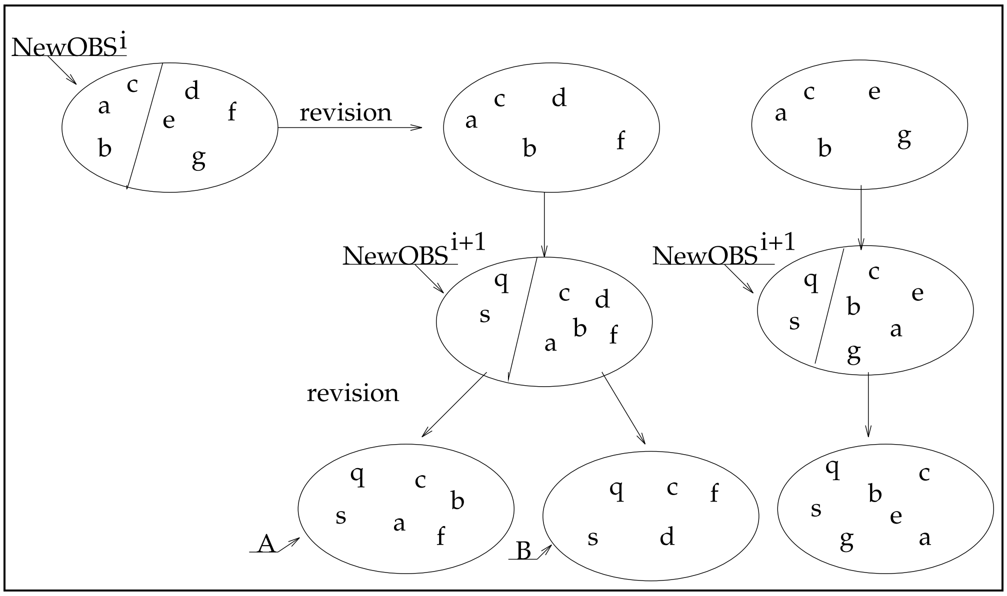

Figure 3.2: Example of a disadvantage of last observation preferential revision

As we see in the figure 3.2, doing the first revision we maintain the $NewObs_i$ in both the sets that we obtain. At this point we add $NewObs_{i+1}$ and we do again the revision. The $NewObs_{i+1}$ are of course present in all the revised sets, but in the set A we have more observations belonging to $NewObs_i$ than in the set B. It might be reasonable to prefer the set A to the set B, but our preference keeps both.

Let us now consider the preference of the first observations. In this case, we choose the maximal subsets that contain all the old observations. The formal definitions are quite similar to the previous case:

Definition 39 (First observation preference) *Let $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, \ldots, OBS_n\}\rangle$ be a hypothetic scenario description at t_i in a narrative and let NewObs be the new observations that we have added at this step. A last observation preference is a function which maps S to the hypothetic scenario description $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS'_1, \ldots, OBS_m{}'\}\rangle$ where $\{OBS'_1, \ldots, OBS'_m\}$ contains all the OBS_i in $\{OBS_1, \ldots, OBS_n\}$ such that there exists a set of observations OBS_j^{i-1} in the hypothetic scenario description at t_{i-1}: $OBS_j^{i-1} \subseteq OBS_i$.*

Definition 40 (First observation preferential revision) *Let $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS_1, \ldots, OBS_n\}\rangle$ be the last hypothetic scenario description of a faulty narrative N. A first observation preferential revision is a function which maps the faulty narrative N in the unique correct narrative which is equal to N, except for the final scenario which is obtained first doing the revision and then doing the first observation preference.*

In this case we include without checking its consistence a OBS_j^{i-1} in every set that we obtain doing the revision, but this does not give any problems because the OBS_j^{i-1} are all consistent. We still have the same considerations as before about the fact that we do not keep the order of arrival of the observations in the successive revisions.

3.6.2 Preference for observations in which we have different belief

To define this kind of preferential revision we first introduce the new concept of *belief observation* that is a couple (observation, number) where the number represents our belief in the observation.

Definition 41 (belief observation) *A belief observation is a couple (obs, n) where obs is one of our observations and n is a number that represents our belief in the observation.*

Definition 42 (belief scenario description) *A belief scenario description is an hypothetic scenario description where we have belief observations instead of observations.*

Definition 43 (belief narrative) *A belief narrative is a narrative where we have belief scenario description instead of hypothetic scenario description.*

The definitions of belief faulty and correct narrative are the same as faulty and correct narrative with just belief scenario description instead of hypothetic scenario description.

Definition 44 (get-val) *Get-val is a function that takes a belief observation (obs, n) and give back n.*

We can now define a *belief preferential revision*; in this case we prefer the sets of observations for which the sum of the beliefs of the observations is maximal.

Definition 45 (Belief preference) *Let* $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{BOBS_1, \ldots, BOBS_n\}\rangle$ *be a belief scenario description in a belief narrative and let NewObs be the new observations that we have added at this step. A belief preference is a function which maps S in the belief scenario description* $\langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{BOBS'_1, \ldots, BOBS'_m\}\rangle$ *where* $\{BOBS'_1,\ldots,BOBS'_m\}$ *contains all the* $BOBS_i$ *in* $\{BOBS_1,\ldots,BOBS_n\}$ *such that* $\sum$ *get-val*(obs_i, n) *is maximal (no other* $BOBS_j$ *gives a higher value) and where the* (obs_i, n) *are the belief observations of* $BOBS_i$*.*

Definition 46 (Belief preferential revision) *Let* $S = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{BOBS_1, \ldots, BOBS_n\}\rangle$ *be the last belief scenario description of a faulty belief narrative N. A belief preferential revision is a function which maps the faulty belief narrative N in the unique correct belief narrative which is equal to N, but for the final scenario which is obtained first doing the revision and then doing the belief preference.*

With the help of the belief preferential revision we can define different versions of preference of the last and the first observations. In the first case we can just put as belief value the time of arrival of the observations. In the second case we can put as belief value the time of arrival changed of sign or we can put the time of arrival as it is, but instead of preferring the sets of observations for which the sum of the belief of the observations is maximal, we prefer the sets for which it is minimal.

With this new definition we do not have the problems that we had with the previous definitions of first and last observation preference revision. In fact we do not need to assume that the new observations are consistent as the new observations are checked in the same way as the other observations and, most important, the order of arrival is preserved also in the successive revisions.

Another property that can be interesting is that if we are in the case of last observation preference and we have to choose between some observations that are newer and more observations that are older, but not so old, we can choose the second set of observations.

We explain this better using the example in figure 3.3. For simplicity we write the time point in which the observation arrives to denote the observation. So for example the two 1's in the picture mean two different observations that have arrived at time 1.

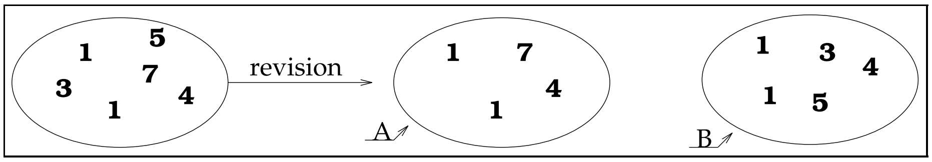

Figure 3.3: Example of choosing the older observations instead of the newer in last observation preference

In this example the belief preferential observation will choose the set B instead of the set A as the two observations 5 and 3 have together a greater value than the observation 7 even if the observation 7 is newer.

3.6.3 Context-dependent belief

In the previous section we have assumed that a belief level is directly associated with each observation. However in many realistic applications the belief in observations can be context-dependent. If there is not so much light we may be more prepared to accept that we make a mistake in an observation, than if it is clear daylight.

Also, the observations can be differently affected by context factors, for example the observations that concern the sight can be affected by the amount of light, while the observations that concern the hearing can be affected by the amount of noise in the environment.

To formalize context-dependent belief we can divide our observations into two categories: context observations, that is observations that give information about the context, for example: it is daylight, there is music, and non-context observation. Each context observation changes the value of belief of the subset of non-context observations which are affected by it.

Definition 47 (context observation) *A context observation is a belief observation which gives information about the context.*

Definition 48 (context function) *A context function is a function that takes a set of context observations and a non context dependent* *observation (obs, n) and gives back a pair (obs, n′), where n′ is the belief value of the observation modified according to the context observations.*

Definition 49 (context belief narrative) *A context belief narrative is a belief narrative where at each time point t_i, instead of adding $NewObs_i$ at each $BOBS_j$, the set $NewObs_i^{'}$ is added, which is obtained by applying the context function to all the (obs, n) in $NewObs_i$, with the context observations in $BOBS_j \cup NewObs_i$ as set of context observations.*

An algorithm for calculating the context function is the following:

- we decide the observations that will be context observations;
- we associate to each context observation the set of observations that are affected by it and how much the fact that this observation is present increases or decreases the belief value of the observations (we suppose that the context observation affects all the observations in the same way);
- the context function takes a set of context observations C and a pair (obs, n), checks if obs is in the set associated with a context observation in C and, if that is the case, changes the value n accordingly. The observation obs can be in the set associated with several context observations. The value that the context function gives back is the one after all the changes.

Definition 50 (context belief preferential revision) *The definition of context belief preferential revision is the same as the definition of belief preference, substituting context belief narrative for belief narrative.*

3.6.4 Preferred models for scenario descriptions in $\mathcal{M}_o$

In section 3.2 we have defined the models for scenario descriptions in $\mathcal{M}_o$. In this section we have then defined preferential revision. We now want to consider the case in which, constructing the models for scenario descriptions, we use preferential revision instead of revision. We call the models obtained in this case preferred models for scenario descriptions in $\mathcal{M}_o$.

Definition 51 (Preferred models for scenario descriptions in $\mathcal{M}_o$) *Given a scenario description $\mathbf{\Upsilon} = \langle \mathcal{M}_o, \mathcal{O}, Diw, SCD, OBS \rangle$, where t_f is the maximal time point in the observations in OBS, if N is the narrative obtained adding at each time point t the observations in OBS*

that have t as time point and doing the preferential revision P every time it is necessary, then the preferred models of $\mathbf{\Upsilon}$ with preference P, $Mod^P(\mathbf{\Upsilon})$, are the models of the hypothetic scenario description corresponding to t_f in N.

The preferred models for scenario descriptions in $\mathcal{M}_o$ can then be obtained constructing the corresponding narrative and constructing the models of the last hypothetic scenario description of the narrative.

3.7 Preference in models of hypothetic scenario description and link with preference in observations

As we have defined a preference with respect to observations we can also define a preference with respect to models. We first give the definition of resumed model.

Definition 52 (resumed model) *Let N be a narrative. A resumed model is a model of a hypothetic scenario description of N at t_{i+1} that was not a model of the hypothetic scenario description at t_i.*

We can now define the FO preference in models.

Definition 53 (FO preference in model) *A FO preference in model is a function FOpref that takes the set of models of a hypothetic scenario description and returns the models that are not resumed models.*

In the next theorem, we prove that there is a correspondence between the FO preference in model and the first observation preference, in the following sense. If we have a hypothetic scenario description S and, doing the first observations preference, we get a hypothetic scenario description S', than the models of S' are the same models that we could get doing the FO preference in the models of S. A graphic representation of this theorem is shown in figure 3.4.

Theorem 5 *Let $S = \langle \mathcal{K},\mathcal{O},Diw,SCD,\{OBS_1,\ldots,OBS_n\}\rangle$ be the hypothetic scenario description at t_{i+1} in a narrative, $T = \langle \mathcal{K}, \mathcal{O}, Diw, SCD, \{OBS'_1,\ldots,OBS'_m\}\rangle$ the hypothetic scenario description that we get doing the first observation preference and Mod(S) and Mod(T) respectively the models of S and T.*

If we apply the preference in models to Mod(S) we get

$$FOpref(Mod(S)) = Mod(T).$$

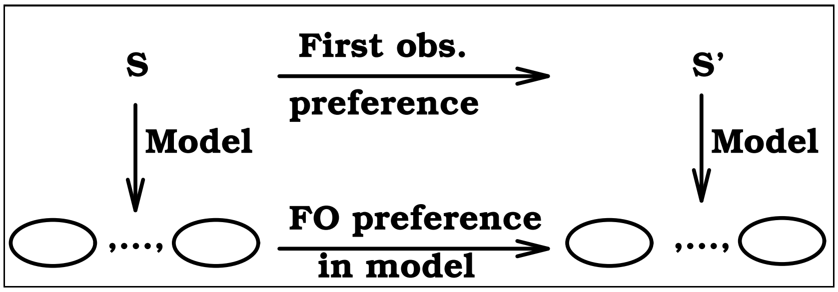

Figure 3.4: Relation between first observation preference and FO preference in model

Proof.

$Mod(T) \subseteq FOpref(Mod(S))$: By the definition of first observation preference for each OBS'_j there exists a set of observations in the hypothetic scenario description that we get at t_i that is a subset of OBS'_j. Then, all the models in Mod(T) are also models of the hypothetic scenario description at t_i, and so are not resumed models.

$FOpref(Mod(S)) \subseteq Mod(T)$: Let us prove that all the non-resumed models are in Mod(T). Let M be an arbitrary non-resumed model i. We want to prove that it is also a model for one of the OBS'_j. Suppose it is not a model of any OBS'_j. This means that it is a model of a set of observations in S of the form $(OBS^i_t \setminus C) \cup (NewObs_{i+1} \setminus D)$ where OBS^i_t is one set of observations of the hypothetic scenario description at t_i and $C \neq \emptyset$. As it is a non-resumed model, it is also a model of one set of observations of the hypothetic scenario description at t_i, say OBS^i_q. In general, it is a model of $OBS^i_q \cup (NewObs_{i+1} \setminus Q)$, but this set is possibly not in S, and therefore not in T, as it can be not maximal. M is a model of $(OBS^i_t \setminus C) \cup (NewObs_{i+1} \setminus D)$ and of OBS^i_q $(OBS^i_t \setminus C) \cup OBS^i_q$. We have two cases:

- if $(OBS^i_t \setminus C) \not\subseteq OBS^i_q$, then $OBS^i_q \subset (OBS^i_t \setminus C) \cup OBS^i_q$. $OBS^i_t \setminus C \cup OBS^i_q$ has model M, but OBS^i_q was a maximal subset of all the observations that we had at t_i for which the scenario had model, and this is in contradiction with the fact that a superset of OBS^i_q, $(OBS^i_t \setminus C) \cup OBS^i_q$, has model.

- if $(OBS^i_t \setminus C) \subseteq OBS^i_q$, then $OBS^i_q \cup (NewObs_{i+1} \setminus D)$ must not have any model otherwise we would have taken it instead of $(OBS^i_t \setminus C) \cup (NewObs_{i+1} \setminus D)$ doing the revision. However, M is a model for $OBS_q^i \cup NewObs_{i+1} \setminus D$ as it is a model of OBS_q^i and $NewObs_{i+1} \setminus D$.

■

3.7.1 Last observation preference and resumed models

We could expect that choosing the resumed model, we could get a model preference that corresponds to the last observation preference. Unfortunately the resumed models are models for the observations of the form:

$OBS_j \cup NewObs$
$OBS_j \cup (NewObs \setminus Q)$

and the not resumed models are models of observations of the form:

$(OBS_j \setminus T) \cup NewObs$
$(OBS_j \setminus T) \cup (NewObs \setminus Q)$

Therefore, we can not characterize the last observation preference in terms of resumed models.

3.8 Conclusion

In this chapter, we have introduced the concept of hypothetic scenario descriptions, that is, a scenario description in which we have a number of sets of observations instead of just one set. We have then defined the concept of narrative as a mapping between time points and hypothetic scenario descriptions, and the concepts of faulty and correct narrative. A revision function maps a faulty narrative into a correct narrative. We have defined the concept of revision in general and we have also introduced a delayed revision and proven that delayed revision gives results similar to generic revision.

We have finally defined four preferential revisions and a model preference and we have shown a correspondence between this preference and first observation preferential revision.

Chapter 4

Relations with belief revision

4.1 Introduction

In this section we consider Gärdenfors' approach to belief revision and, in particular, how Gärdenfors' postulates can be applied in our case.

In his book 'Knowledge in flux' [Gär88], Gärdenfors considers the case in which people are forced to revise their beliefs, that is, they have a new belief and if they want to accept this new belief they must give up some of their old beliefs.

The problem consists of deciding which beliefs to give up. Unfortunately, there are not in general logical reasons for choosing one belief other another. One can only consider a *criterion of informational economy* that demands that as few beliefs as possible should be given up in order to accept the new belief. In some sense the change should be a *minimal change*. Gärdenfors' postulates have the aim to delimit the meaning of "minimal change". He believes that with the postulates and with certain *non logical* factors it is possible to obtain a unique revision of a set of beliefs.

Our approach is in some ways similar: we revise a particular kind of beliefs, that is, the observations of the world, and we respect the principle of minimal change. In fact, we take the maximal sets of observations that are correct for our narrative, but we do not have general non logical factors giving us a unique revised set of observations.

In general, it is not possible, given a set of observations, to decide which ones are incorrect. Therefore, our revision gives back several alternative sets of observations. Non-logical factors can give a preference in the observations, and this is the case that we have examined with preferential revision, but in general preferential revision can also not give back a unique set of observations. It can just choose some of the sets.

In the next section, we describe Gärdenfors' postulates. In section 4.3 we examine how the Gärdenfors postulates apply to our revision in its general form, and in section 4.4 how they apply to preferential revision.

4.2 Gärdenfors' postulates

Gärdenfors considers the case where there is a consistent belief set K to which a new belief A is added. If the new belief does not contradict the beliefs already in K, K is expanded by means of an *expansion function* $+$. This function takes a set K and a belief A and gives back a new belief set K^+_A that contains the old and the new beliefs. If the new belief A contradicts the beliefs that are already in K, it is necessary in order to maintain the consistency to revise K. A *revision function* $*$ is a function that takes a belief set K and a sentence A and returns a new belief set K^*_A.

The belief set in which there are two contradictory beliefs and in which therefore everything is believed is indicated by $K_\perp$.

The postulates that Gärdenfors presents in his book are based on logical considerations about what can be expected from expansion and revision functions, and have the aim of circumscribing the set of rational solutions for these functions. We present here just the postulates for the revision function, as these are the ones that are more interesting for our work.

The first postulate requires that the output of the revision is still a belief set.

(K^*1) For any sentence A and any belief set K, K^*_A is a belief set.

The second postulate guarantees that the new belief A is accepted in K^*_A. This is a strong assumption, which in fact implies that the last belief is always the one that is kept.

(K^*2) $A \in K^*_A$.

The third and the fourth postulates give the relations between the revision and the expansion function. These postulates entail that in general revision is a special case of expansion except in the case in which $K^+_A = K_\perp$.

(K^*3) $K^*_A \subseteq K^+_A$.

(K^*4) If $\neg A \notin K$, then $K^+_A \subseteq K^*_A$.

The next postulate says that the only case where revision leads to the inconsistent belief set is when the new belief is in itself contradictory. In all other cases the new belief set is consistent.

(K^*5) $K^*_A = K_\perp$ if and only if $\vdash \neg A$.

The sixth postulate says that, if two formulae A and B are equivalent they lead to identical changes.

(K^*6) If $\vdash A \leftrightarrow B$, then $K^*_A = K^*_B$.

Finally, the last two postulates say that the minimal change to include two beliefs A and B in K at the same time, indicated by $K^*_{A\&B}$, corresponds to doing the expansion of K^*_A by B, as long as B does not contradict the beliefs in K^*_A.

(K^*7) $K^*_{A\&B} \subseteq (K^*_A)^+_B$.

(K^*8) If $\neg B \notin K^*_A$, then $(K^*_A)^+_B \subseteq K^*_{A\&B}$.

4.3 Applicability of Gärdenfors' postulates to incorrect observations revision

Let us now consider how Gärdenfors' postulates apply to our revision. Our revision maps a narrative to a narrative, and a narrative seems a concept quite far from the concept of belief set, but what our revision really does is to consider the last step of a narrative and in particular the sets of observations in this step. If a set is not correct for the narrative, it is mapped into a set of correct sets of observations. The concept of observation is quite similar to the concept of belief: they can both be considered "inputs" from the outside world. In a certain sense, revising the sets of observations can be seen as revising the beliefs about the world. In Gärdenfors' approach, anyway, the new belief can contradict just the other beliefs, while in general an observation does not contradict another observation in a classical logical sense. Instead, the observations contradict within a scenario description, that is, with respect to the actions, the principle of inertia, and in general the ontological and epistemological assumptions in the scenario. We therefore consider examining the postulates, not with respect to logical contradiction, but for incorrectness with respect to a scenario.

An important difference is that in Gärdenfors' work it is assumed that the revised set K^*_A is unique and that it should contain the new belief A. He does not compare the belief A with the old beliefs that should be thrown away to accept A, he just chooses to privilege A in every case.

In our case, the new observations are not in general privileged and in the definition of revision without preference we keep both the new and the old observations; we just put the contradictory observations in different sets and we consider these sets as alternative histories of the world.

In the rest of the section we consider the Gärdenfors postulates one by one and we see how they apply to our revision. To make the reading of this part easier we use a notation similar to the one of Gärdenfors and we express the postulates in this notation. We call O the set of sets of observations at the last step of a narrative N before adding the new observation, o the new observation added at this step, O^+_o the set of sets of observations after adding o and O^*_o the set of sets of observations in the last step of $Rev(N)$.

As the first six postulates consider the case in which just one new belief is added to the belief set, we consider, examining these postulates, the case in which only one new observation is added. The case in which more than one observation is added at the same time is considered in the last two postulates.

Postulate (K^*1) translated into our case says that if o is an observation and O is a set of sets of observations, O^*_o is a set of sets of observations. This is trivially true in our case.

Postulate (K^*2) says that $o \in O^*_o$. This is true if we consider the membership of o to O^*_o as membership of one of its sets of observations. In fact, our revision maintains in general all the observations, it just puts them into different sets.

An exception is represented by the case in which o is contradictory with the narrative. In this case, the revision eliminates o from all the sets. This is similar to the case that Gärdenfors considers in postulate (K^*5). The postulate says that if the formula to add is a contradiction then the revision returns an inconsistent set. In fact, in Gärdenfors' approach, the new belief is always added, and therefore if it is contradictory it makes the set inconsistent. In our case, the revision maintains the consistence at the expense of throwing away the new observation.

The postulates (K^*3) and (K^*4) indicate the relations between revision and expansion.

In particular the postulate (K^*3) says that O^*_o is included in O^+_o. This is true by definition of revision if we consider the inclusion in the sense that each set in O^*_o is included in one set in O^+_o.

The postulate (K^*4) says that if o is consistent with the observations in O, then $O^+_o \subseteq O^*_o$. This is not completely true in our case. In fact, o should be correct for the narrative, that is, if one adds it to the last hypothetic scenario description of the narrative the scenario should still be correct. It is not enough that o is consistent with the observations in O. This is quite natural since the basic structure that we have is the narrative, and therefore we refer to it for the correctness.

We have discussed the postulate (K^*5) before. We can just add that as our observations are atomic formulae, we do not have the case in which o is a contradiction, but just the case in which o is contradictory with the narrative.

The postulate (K^*6) is not very meaningful in our case. As the observations are atomic, they are logically equivalent just if they are equal.

The postulates (K^*7) and (K^*8) consider the case in which we add more observations at the same time.

The postulate (K^*7) says that if we add two observations $o1$ and $o2$ at the same time and we do revision, (we indicate this with $O^*_{o1\&o2}$), then, if $o2$ is correct with the revised narrative after the adding of $o1$, we obtain the same result as if we would do the expansion of O^*_{o1} by $o2$. This is true in our case. If $o2$ is correct with the revised narrative after the adding of $o1$, this means that it is not in contradiction with $o1$ and with the rest of the scenario. Therefore it is kept in all the sets of observations and it does not influence the revision for $o1$. But this has the same effect as we would have obtained adding $o2$ to all the sets of observations in the revised narrative after the addition of $o1$, that is, doing the expansion in the revised narrative.

The postulate (K^*8) says that if $o2$ is consistent with the observations in O^*_{o1} then $(O^*_{o1})^+_{o2} \subseteq O^*_{o1\&o2}$. This is not completely true in our case for the same reason that we have given for the postulate (K^*4). In fact, $o2$ should be correct for the narrative, that is adding it to the last hypothetic scenario description of the narrative the scenario should still be correct. Also in this case is not enough that $o2$ is consistent with the observations in O^*_{o1}.

4.4 Applicability of Gärdenfors' postulates to preferential revision

In case of preferential revision, not all the sets that the revision returns are kept, but just a choice of them. Let us consider how this affects the applicability of Gärdenfors' postulates.

The postulate (K^*1) is trivially true also in case of preferential revision.

The postulate (K^*2) is really the one that can not be satisfied in general in preferential revision. In fact, as some sets are discarded, the new observation is not always kept.

An interesting case is the last observation preferential revision. In last observation preferential revision the postulate (K^*2) is not only satisfied, but also in a stronger way respects the general revision.

In last observation preferential revision the new observations are always preferred, and this corresponds to Gärdenfors' assumption that the new belief is always accepted. Therefore, in last observation preferential revision, the new observation is included in all the sets of observations that the revision returns, and not only in at least one of them as in the case of general revision.

Also, the postulate (K^*5) is satisfied in a strong way in the case of last preferential revision. When we have studied the case of last preference revision, we have assumed that the new set of observations added at each step was consistent. Without this assumption we would have obtained inconsistent sets of observations from revision as in Gärdenfors' case.

The rest of the postulates are applicable in the same way to preferential revision as to revision.

4.5 Conclusion

Our work seems to have several points in common and analogies with Gärdenfors' work, and the parts that differ point out interesting properties of the revision function. In particular it is interesting to notice that our last observation preference corresponds quite well to Gärdenfors' idea of revision.

In the future we would like to develop this relation with Gärdenfors' work considering not just his postulates, but also the more recent part of his work.

Chapter 5

Conclusions, future work and acknowledgements

5.1 Conclusions

In this work, we have considered the problem of unreliable information, and we presented a solution that respects the principle of minimal change, that is, a solution that tries to maintain all the information that can possibly be maintained.

We can not in general decide which information is incorrect and which is right. We therefore give a number of different alternative views of the world and some preferences on these alternative views.

If we do not have any preference we have proven that we can postpone doing the revision of our information.

We have considered four preferences in our work, which are quite general but others could be considered. The preference is really application dependent and just knowing from which sources the observations come and how the agent uses them, we can define the better preference for each application.

We have also proven a correspondence between a preference in semantics, FO preference in model, and a syntactic preference, last observation preferential revision.

Finally we have considered the relations between our definition of revision and Gärdenfors postulates.

5.2 Future work

We see this as an initial work in a field not yet explored and quite promising. In particular we think that the theory about revision and preferential revision can be developed further. It can also be interesting to study how the actions and the schedule are affected by admitting incorrect observations. It seems reasonable to think that also the schedule should change in the development of the narrative.

5.2.1 Relation with Paraconsistent Logics

Paraconsistent Logics are logics in which contradictions are allowed and can be reasoned about. For an excellent survey of paraconsistency see [PRN89]. We would like to prove that the underlying logic of our theory is a paraconsistent logic.

5.2.2 Implementation

A natural way to implement the revision function and the preferential revision is with logic programming. This implementation should be inserted in the general project of encoding scenario description in logic programming that has been started in [Doh94].

The revision as it is now defined is quite costly. In the worst case all the maximal subsets of the sets of observations must be examined. It can be interesting to study in which cases and with which restrictions we can obtain good average performance.

5.2.3 Relation with belief revision

We think that belief revision is quite related with the problem of unreliable information and that the studies done in the field of belief revision can help in understanding better and can give impulses to the study of unreliable informations. In the present work we have considered the relation between Gärdenfors' postulates and our definition of revision. We hope in the future to extend this relation to other work by Gärdenfors and to consider also other authors in belief revision.

5.3 Acknowledgements

First of all, I would like to express my thanks to my advisor, Erik Sandewall, for guidance and encuragement. I would also like to thank my other advisor Giorgio Levi.

I would also like to thank Maria Simi and Patrick Lambrix for having read and commented the manuscript and Pavlos Peppas for his helpful suggestions.

Finally I would like to thank Lars Karlsson for his continuous support and for for having read and commented the draft versions of this work.

Bibliography

[BDP95] S. Benferhat, D. Dubois, and H. Prade. How to infer from inconsistent belief without revising? In *Proc. IJCAI'95*, pages 1449–1455. 1995.

[Doh94] P. Doherty. Notes on PMON circumscription. Technical Report 94-43, Department of Computer and Information Science, Linköping University, 1994. Available on WWW: `http://www.ida.liu.se/labs/rkllab/people/patdo/`.

[Gär88] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, 1988.

[Kar95] L. Karlsson. *Specification and Synthesis of Plans Using the Features and Fluents Framework.* Licentiate thesis, Department of Computer and Information Science, Linköping University, 1995. Available on WWW: `http://www.ida.liu.se/labs/rkllab/people/larka/`.

[PRN89] G. Priest, R. Routley, and J. Norman, editors. *Paraconsistent Logic: Essays on the Inconsistent.* Philosophia Verlag, 1989.

[RM70] N. Rescher and R. Mannor. On inference from inconsistent premises. In *Theory and Decision*, volume 1, pages 179–219. 1970.

[San94] E. Sandewall. *Features and Fluents.* Oxford Press, 1994.

[Yi95] Choong-Ho Yi. Towards assessment of logics for concurrent actions. Accepted to the AAAI 1995 Spring Symposium: Extended Theories of Action, Stanford University, 1995.