



**HAL**  
open science

# Modelling reasoning processes in natural agents: a partial-worlds-based logical framework for elemental non-monotonic inferences and learning

Christel Grimaud

► **To cite this version:**

Christel Grimaud. Modelling reasoning processes in natural agents: a partial-worlds-based logical framework for elemental non-monotonic inferences and learning. *Journal of Applied Non-Classical Logics*, 2016, 26 (4), pp.251 - 285. 10.1080/11663081.2016.1214805 . hal-01608194

**HAL Id: hal-01608194**

**<https://hal.science/hal-01608194>**

Submitted on 3 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

*Modelling reasoning processes in natural agents:  
a partial worlds based logical framework for elemental  
non-monotonic inferences and learning*

Christel Grimaud\*

*Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France*

In this paper we address the modelling of reasoning processes in natural agents. We focus on a very basic kind of non-monotonic inferences for which we identify a simple and plausible underlying process, and we develop a family of logical models that allow to match this process. *Partial worlds models*, as we call them, are a variant of Kraus, Lehmann and Magidor's cumulative models. We show that the inference relations they induce form a strict subclass of cumulative relations, and tackle the issue of providing sound and complete sets of rules to characterize them. Taking inspiration from Gabbay & Schlechta's work on preferential structures (Gabbay & Schlechta, 2008, 2009), we analyse the question in terms of definable sets of partial worlds, and conclude that completeness is probably unreachable using a standard propositional language. This brings us to enrich our language with an additional connective  $\parallel$ , allowing to distinguish between two kinds of disjunctions in partial worlds context. Within this renewed framework, we provide two representation theorems: one for inference relations induced by precisification-free *smooth* models, and the other for inference relations induced by precisification-free *ranked* models. Finally we give an aperçu of how partial worlds models lend themselves to the modelling of learning.

**Keywords:** partial worlds models; non-monotonic inferences; reasoning processes; learning

## 1. Introduction

Logical modelling of reasoning and learning is confronted with a handful of long-lasting problems. One of them is how the modelling of reasoning and that of learning should be articulated. Indeed, the acquisition of information by a cognitive agent triggers both inferential processes and learning, that is, a revision of the agent's dispositions to infer. An agent's dispositions to infer should therefore depend on the sequence of its previously acquired information, that is, on its past experience. Closely related with the previous one, another issue is that a suitable model of learning should afford iteration, since learning is essentially an ever repeating process. Yet at this time there is no consensus about how the revision of an agent's dispositions to infer should be conducted.

The present paper introduces a logical framework for the modelling of inferences that aims at opening the way for a solution to these problems. Its main thrust is that to address these issues one should take inspiration in the way natural agents (*i.e.*, humans and animals) actually proceed when they draw inferences and learn. Indeed,

---

\*Email: [ch.grimaud@numericable.fr](mailto:ch.grimaud@numericable.fr)

learning and reasoning are natively articulated in these agents, and a logic that would model their very mental processes would be naturally immune to the problem. In a first attempt to provide such a logic, we focus here on a very basic kind of inferences that, we shall argue, can be considered as the primary core of reasoning in natural agents. We suggest a plausible underlying process for these inferences, and develop a family of logical models that allow to simulate it. Then we tackle the issue of providing sets of rules to characterize the inference relations induced by the models. These rules are a by-product of the posited process, and should thus be taken as reasoning rules that arise from the functioning of a brain running such a process. In this view, the introduced framework gives an insight into how it is possible that intelligent reasoning and logic emerge from a brain at work. Finally we give an informal preview of how the suggested models naturally lend themselves to the modelling of the learning processes attached to these same inferences. A throughout formalization of the modelling of these learning processes shall be supplied in a further paper.

## **2. Motivation and intended interpretation of the framework**

Traditionally, the approach taken in logical modelling of reasoning is more or less tacitly a normative one: it considers how ideal agents should reason, not how real agents do. Moreover, it generally provides a purely external account, in the sense that it disregards the internal (mental/neural) processes by which reasoning is completed. By contrast, the approach that shall be taken here is descriptive and internal, which means that it shall seek to model what really happens in natural agents as they reason, and to match their reasoning processes. The first task is thus to identify these processes. This is actually a matter of cognitive sciences (in the very wide sense of the term) rather than a logical issue, yet a brief discussion of the topic may help the reader to grasp the intent and aims of the logical framework to be introduced below. This is why the first part of the present section is devoted to this question. Once a plausible and tractable process is identified, the second part of the section gives a first overview of the envisioned logical modelling and provides a number of clarifications.

### **2.1 *The modelled mental process***

When looking at reasoning in natural agents, the first thing that one may notice is that it is not a single and homogeneous process, but that it rather incorporates a wide range of cognitive abilities. Some of these can presumably be found in human beings only, as for instance reflexivity and superior language abilities, while others are more widely spread and appear at varying extents in the so-called ‘higher species’ (as for instance the ability to handle relations) and finally others are pervasive throughout animal kingdom. One can safely assume that each of these abilities is supported at the neural level by some corresponding neural process. It is also patent that the abilities and processes involved in human reasoning are manifold and intricate, and for many of them, still poorly understood. It would thus be unreasonable to hope to model the whole of human reasoning all at once. A more reasonable plan is to search for some simple, elemental and plausible process to model.

#### *2.1.1 Automatic inferences*

It happens that among the cognitive abilities that are commonly displayed by natural agents, there is one that appears to be at the same time very basic, pervasive and essential to reasoning. It is the ability to draw non-monotonic inferences out of the

available information.

In humans, the most elemental form in which it manifests itself is through those ‘quick and dirty’ inferences that we routinely draw based on our intuitive knowledge of things and situations, as for example when observing a bird we expect it to fly away if we attempt to come closer. An essential characteristic of these inferences is that they are not deliberate, but occur in an automatic manner. In fact, they do not even require consciousness for their processing, and they generally remain unconscious unless we bring them to consciousness by a reflexive move on ourselves. As far as introspection tells us, they do not seem to rely on verbal representation either, as is suggested by the fact that most of the time we simply ‘feel’ the premisses and the conclusion, without bothering to put them into words. And finally, they seem almost effortless, which makes it probable that they run on a very simple and ‘cheap’ process. They certainly do not constitute the whole of humans’ inferential capacities — obviously, reflexivity and verbal representation support much more sophisticated reasoning — but they take part in most of the decisions and judgements we make in our everyday life. In particular, we often use their conclusions as premisses in more complex reasoning processes such as, precisely, those involving reflexivity or language. Therefore, they appear as a simple, autonomous (*i.e.* not relying on any other ability) and fundamental (*i.e.* on which other abilities rely) piece in human reasoning.

However, the ability to draw such inferences is far from being specific to humans. The drawing of inferences comes very close to the making of decisions, all the more when it comes to automatic, unconscious and non-verbal inferences. But all animals do make decisions, and the question is whether these are the result of a rigid reflex or of a more flexible process. If the decision process is flexible, that is, if the animal’s decisions non-monotonically depend on the available information, then it can be seen as running a non-monotonic inferential process. It turns out that flexible decision making is the norm rather than the exception all across animal kingdom. Indeed, even minute-brained animals such as insects or spiders, the reasoning capacities of which are supposedly very limited, have been observed to make non-monotonic decisions depending on what they perceive from the situation at hand<sup>1</sup>. Thus the ability to draw non verbally-supported, unconscious non-monotonic inferences in an automatic manner is probably the most widespread of all reasoning abilities.

I shall call this kind of inferences *automatic inferences*. Due to the profound similarities that, despite evident superficial differences, prevail across species in the general functioning of brains, there is no reason to suppose that the drawing of these inferences is realized in a very different way from one species to another. One may thus imagine that the ability to draw automatic inferences relies on an autonomous process, simple enough to fit in smallest brains, and that it forms the primary core of reasoning in natural agents, while other cognitive abilities come as additional modules that plug into it to enhance reasoning in higher species.

### 2.1.2 A plausible process for automatic inferences

To identify the process that underlies the drawing of automatic inferences in natural agents, one may rely on the intuition one has of its own reasoning processes (*i.e.*, appeal to introspection), or look at the neural machinery that supports them. These two methods are not exclusive, and even they usefully supplement each other.

Introspection for its part suggests that to draw automatic inferences, we rely on the mental representations we have of the objects and situations that we know. For example,

---

<sup>1</sup> See for example Jackson & Cross (2011) pp.130-131, and Bartos (2008), both on spiders.

if we have to decide whether a bird that we are observing will fly away if we attempt to come closer, we appeal to our experience of birds. This means that we search our memories for similar situations, and check if in these situations the bird did fly away or not. If it did fly away in all the similar situations we can recall, then we conclude that it will do the same in the present case.

It should be noted that these memories need not be individual memories for each and any situation we faced in our life. On the contrary, in most cases the memorization process sums up similar experiences into a unique memory, leaving aside insignificant details. For example, we know about sparrows without remembering each and any situation in which we faced one. So the memories we have of things, and therefore those on which we rely to draw automatic inferences, are in fact mental representations of some archetypal cases, that is, they are some kind of concepts.

But they are not any kind of concepts. As is well known, concepts can be more or less general, with more general concepts subsuming more specific ones, thus forming clustering upward chains of more and more general concepts. A key point is that to draw automatic inferences, we rely only on the most specific of our concepts. The reason is that these are the most detailed and precise representations we have of things, and thus the most suited to base our inferences upon. We search these most precise representations, looking for some that match what we perceive of the present situation, that is, that satisfy all the features we are able to grasp of it, and we check whether the ones we have found also satisfy the feature ‘flies away’. If all of them do, then we infer that the bird will fly away<sup>1</sup>.

But in doing so we do not check each and any of our most specific concepts that match what we perceive of the present situation, which would be far too long and mentally demanding. Rather, we simply check the first ones that come to our mind, that is, the ones that are the most vivid in our memory. If all of these satisfy the feature ‘flies away’, then we conclude that the bird will fly away. It is this partial recollection of memories that causes automatic inferences to be non-monotonic, as a more complete information about the current situation would probably reactivate a different set of memories, and so bring us to different conclusions. However this incomplete retrieval does not significantly affect the relevance of our conclusions, because our most vivid memories are also likely to be the most significant and important for us, and thus the best able to provide us with useful conclusions.

It may also happen that we fail to find a memory of similar objects/situations, and that therefore the present situation appears new to us. In such cases the above process cannot be completed, which in turn triggers learning processes. These essentially consist in the creation of a new memory, or the supplementation of existent ones.

Brain sciences for their part provide details on how the above described inferential process could plausibly be realized in brains. At first glance, brains show noticeable disparities across animal kingdom, notably in size or anatomical plan. But beyond these obvious differences they all share the same general functional schema. In particular, in all brained species perception is achieved through a series of parallel sensory channels, one for each sensory modality (vision, olfaction, etc.) occurring in the considered species. Each of these channels collects cues from the outside world by the means of a set of specific neural receptors (light receptors in the retina for the visual channel, chemico-receptors in the olfactory channel, and so on) and further processes them to extract relevant features. For instance, edges, colours or directional motion are features

---

<sup>1</sup>In practice it is likely that more general concepts do nevertheless play a technical (*i.e.* not computational) role in the drawing of automatic inferences, notably by providing shortcuts enabling faster responses when possible. But since we are only interested here in computational issues, we can ignore this point.

typically extracted in visual channels, while particular odours are features extracted in olfactory channels. In all cases, the extraction of a given feature is achieved by a set of dedicated neurons that specifically respond to this feature. As one progresses along a sensory pathway, features may combine to form higher-level features, such as a particular angular arrangement between edges or a particular texture<sup>1</sup>. Which features are extracted by a given channel depends on its particular neural organization, so the type, number and complexity of features that are eventually perceived varies across species, but the general rule for all brained species from arthropods to humans is that sensory information is analysed into features.

And in all brained species too, sensory channels finally output to some central brain areas. These are diversely organized — obviously an insect’s brain does not look like that of a mammal — but they all have two main characteristics in common. First, they contain neurons that receive inputs from distinct sensory channels, so that they respond to the co-occurrence of features from different sensory modalities; and second, they are involved in learning and memories. In mammals, where these areas have been extensively investigated, these same neurons have been showed to be the physical support of concepts, in the sense that they fire specifically to the presentation of various stimuli corresponding to the considered concept, regardless to the stimuli’s other qualities. For example, neurons responding specifically to nests have been found in mice’s central brain areas<sup>2</sup>, while neurons responding specifically to particular persons, independently of whether the presented stimuli is a photo, the person’s name written or pronounced or even the simple thinking of her, have been found in humans’ same areas<sup>3</sup>. This remarkable behaviour led the researchers who study these neurons to dub them *concept-neurons*.

All taken together, these findings not only corroborate the former intuition that mental representations encode our knowledge of things and situations as sets of features, but also they allow to imagine a plausible neural process for automatic inferences: since a mental representation is evoked as soon as the particular neurons<sup>4</sup> that support it fire, its vividness in the mind depends on the ability of the considered neurons to respond to an appropriate input. At the neural level, the inferential process would then simply correspond to the fact that not all the neurons that receive a given input fire, but only those for which the sum of all the received inputs exceeds a certain threshold. Differences in strength of excitative connections and mutual excitation/inhibition between concept-neurons would then bring them to respond differently to a common input, causing the difference in vividness of our memories.

---

<sup>1</sup>The curious reader will find examples of such complex features extracted in macaques’ visual pathway in Tsunoda & al. (2001) and Tanaka (2003).

<sup>2</sup>Lin & al. (2007). The idea of animals having mental representations has long been dismissed, as animals were mostly regarded as mere stimulus-response machines. Yet over the past decades this view has been seriously challenged by animal cognition studies, and nowadays it is widely acknowledged by researchers in the field that most animals, including small brained ones such as arthropods, do have mental representations (see for instance Chittka & Niven (2009) and Jackson & Cross (2011) on mental representations and concepts in arthropods). The reluctance that the non-specialist may feel to regard animals, and particularly tiny brained ones, as having mental representations, presumably comes from the fact that (s)he tends to associate to the meaning of these words the intuition (s)he has of his/her own human subjective experience, which (s)he needs not do. From the cognitive standpoint — and as for instance Jackson & Cross put it — a mental representation is much more simply ‘*something more like an internal state that carries information and is then put to use during decision making. A key idea is that representations are used for processing that happens several steps removed from simple stimulus-response chains*’ (Jackson & Cross (2011), pp. 120). It is in such a dispassionate sense that we take the term.

<sup>3</sup>Quiroga (2012).

<sup>4</sup>For robustness reasons, redundancy is the general rule in brains, and the representation of concepts is no exception. So each concept is supported not by a single concept-neuron but by a set — technically referred to as an *assembly* — of such neurons.

## 2.2 Informal presentation of the framework

In the previous subsection, we argued that automatic inferences form the primary core of reasoning in natural agents, and we described a simple process that plausibly supports them, first at the mental level of description and then at the neural level. In light of this it appears that the most convenient approach for the modelling of reasoning in natural agents is to proceed by steps and to model first this core process, and leave the modelling of additional cognitive abilities to further work. This is what we shall do in the sequel of this paper.

In practice, we shall consider an unspecified cognitive agent (hereafter ‘the agent’) which shall be assumed to draw automatic inferences according to the above described process, and we shall set up a logical framework that intends to match this process. It should thus be kept in mind that the to-be-introduced framework is meant to account for automatic inferences alone, to the exclusion of any other cognitive ability the agent may enjoy. In particular, it does not intend to model more sophisticated kinds of inferences, notably those that rely in a decisive manner on consciousness, reflexivity or verbal representation. But in cases where automatic inferences occur as a subprocess of these, it shall of course account for this precise subprocess.

We shall call *inferential system* the part of the agent’s brain that, according to the above account, processes automatic inferences. It is this inferential system, or more exactly, its operations, that we shall intend to model.

To represent the information that is processed by the agent’s inferential system, we shall use a propositional language the variables of which shall stand for the features the agent is physiologically able to perceive — that is, the features that its perceptual system is able to extract from sensory information. Accordingly, features shall not be regarded as properties of external world’s objects, but as information (in the computational sense of the term) flowing within the agent’s brain. However we shall keep in mind that, from the agent’s point of view (*i.e.*, as it experiences them), features indeed appear as properties of external world objects. This is why we shall sometimes write such things as ‘*in the agent’s view, the object satisfies the feature  $f$* ’. It should be noted that the finite number of neurons in the agent’s brain only allows it to discriminate, and thus to perceive, a finite number of features, hence the logical language shall be finite. For simplicity, we shall assume that the set of the features that the agent is physiologically able to perceive does not change over time, so the logical language shall be supposed to be fixed. As is the custom, propositional variables shall be written in lower case italic latin letters  $p, q, r$ , etc., while formulas shall be denoted by lower case italic greek letters  $\alpha, \beta, \gamma$ , etc.

One may notice that the neural organization of sensory channels as sketched above only allows agents to perceive ‘positive’ features. For example, an agent may perceive something as being red, but not as not being black. Yet ‘red’ and ‘black’ are mutually exclusive features, in the sense where recognizing something as red automatically brings an agent to reject the idea that it is black. In other words, it brings the agent to conceive the considered object as not-black. At the neural plane, this is most likely achieved through mutual inhibition between concept-neurons, but from a computational standpoint, it comes down to having negative information such as ‘not-black’ being processed by the agent’s inferential system. For this reason we shall also consider negative features, and quite naturally represent them by negative literals of the language. Features themselves shall be denoted by slanted lower case latin letters  $f, g$ , etc. We shall sometimes write ‘*not- $f$* ’ to denote the negative counterpart of the feature  $f$ .

The mental representation of the elements of a given class of objects or situations<sup>1</sup> in the agent’s mind shall be construed as the set of positive and negative features that, in its view, are satisfied by the corresponding objects/situations. We shall make the simplifying assumption that all the features in this set equally contribute to the mental representation, that is, that this set is *not* structured. Under this assumption, a mental representation can be simply seen as the conjunction of all the features that compose it. We shall moreover assume that the agent’s mental representations are consistent, which means that a feature and its negation cannot take part at the same time in a same representation. This will allow us to figure mental representations by partial worlds. More specifically, if  $r$  is some mental representation in the agent’s mind, we shall represent  $r$  by the partial world  $w$  such that for any literal  $\lambda$  in the language,  $w$  satisfies  $\lambda$  if and only if  $\lambda$  stands for a feature that pertains to  $r$ . Mental representations shall be denoted by means of slanted lower case sans-serif latin letters  $a$ ,  $b$ ,  $c$ ,  $r$ , etc. If  $r$  is a mental representation in the agent’s mind, we shall sometimes denote  $w_r$  the partial world that stands for  $r$  in the model.

We shall say that a mental representation  $r$  *satisfies* a feature  $f$  if and only if  $f$  is in the set of features that compose  $r$ , that is, if and only if in the agent’s opinion, the corresponding object(s)/situation(s) satisfy/ies  $f$ . For example, the agent’s mental representation of sparrows (supposing it has one) will satisfy the feature ‘has a beak’ if and only if in the agent’s opinion, sparrows have a beak<sup>2</sup>. The fact that a partial world standing for a mental representation satisfies a literal  $\lambda$  shall thus represent the fact that, in the agent’s opinion, the corresponding object(s)/situation(s) satisfy/ies the feature  $\lambda$  stands for. Similarly, we shall say that a mental representation  $r$  satisfies the content of information represented the formula  $\alpha$  if and only if the partial world that represents  $r$  satisfies  $\alpha$ .

To know something is nothing but to have a certain mental representation of that thing. Therefore, for any class of objects/situations that the agent knows, there is a corresponding mental representation in its mind. As explained above, from a computational standpoint only the most specific (precise) of them play an effective role in the drawing of automatic inferences, so we can restrict our interest to these. The set of all these most precise mental representations can straightforwardly be represented by the set  $\mathcal{U}$  of partial worlds  $w$  such that  $w$  represents one of the agent’s most precise mental representations. Furthermore, since mental representations are supported in brains by assemblies of dedicated neurons,  $\mathcal{U}$  can also be seen as representing a set of such assemblies: namely, the set of all the assemblies of concept-neurons that support the agent’s most specific mental representations.

In the previous subsection, we remarked that mental representations can be more or less vivid in the memory of an agent, in the sense that they can be more or less prone to come to its mind in response to some appropriate informational input — that is, in

---

<sup>1</sup> As we saw, memorization process generally sums up multiple experiences into a unique memory, so that a mental representation is in fact the common representation of the elements of a class of objects or situations (in cases where a mental representation is built up from a single experience, this class will simply contain one single element). Furthermore, since mental representations boil down to sets of features, they do not distinguish between objects and the situations in which these appear. For example, a representation of birds may contain some information about the context in which the birds in question were observed. For this reason, a mental representation can be seen at the same time as the representation of some object(s), and as the representation of some situation(s) involving these/this object(s). The ability to separate objects and situations is most likely realized at a much more higher level — provided, of course, that such a higher level is implemented in the agent’s brain.

<sup>2</sup> For the ease of reading, in examples we shall regard such things as ‘has a beak’, ‘flies away’ and so on, as features. This is quite a simplification, as actual features are in fact more likely to be of the kind of those identified by Tsunoda & al. (2001) and Tanaka (2003), while ‘has a beak’, ‘flies away’ and so on, are more likely to be conjunctions of a great number of such features. However this simplification is harmless, since we regard mental representations as conjunctions of features and conjunction is associative.

response to its mentally considering the corresponding content of information. We further suggested that such an inequality in vividness might stem from a difference in the ability of the concept-neurons that support these representations to fire in response to a neural input. To account for this difference in vividness, we shall endow  $\mathcal{U}$  with a binary relation  $<$ . More specifically, if  $w$  and  $w'$  are partial worlds in  $\mathcal{U}$ , the fact that  $w < w'$  shall represent the fact that the mental representation figured by  $w$  is *more* vivid in the agent's memory than the one figured by  $w'$ . We shall call the structure  $\mathcal{M} = (\mathcal{U}, <)$  a *partial world model*. If one looks at  $\mathcal{U}$  as the set of the agent's most precise mental representations, then  $\mathcal{M}$  can be seen as representing the agent's worldview — its *Umwelt*, so to say<sup>1</sup>. But if one looks at  $\mathcal{U}$  as the set of the assemblies of concept-neurons that support its most specific concepts, then  $\mathcal{M}$  can be seen as representing the agent's inferential system.

We shall say that the agent is *disposed to infer  $\beta$  from  $\alpha$*  if and only if it is in such a condition that the following holds: if it were to consider the content of information  $\alpha$  (where  $\alpha$  represents the totality of the information that the agent considers at this time), then it would come, after a few instants, to consider, among others, the content of information  $\beta$ <sup>2</sup>. According to the inferential process suggested in section 2.1.2 above, the agent is disposed to infer  $\beta$  from  $\alpha$  if and only if all its most vivid mental representations among those that satisfy  $\alpha$  also satisfy  $\beta$ . At the neural level, a disposition to infer can be seen as a particular arrangement of the neural connections in the agent's inferential system such that, if a suitable excitative input was delivered within its inferential system specifically to its neurons that support concepts that satisfy  $\alpha$ , then after a few instants only those supporting concepts that also satisfy  $\beta$  would be active. In cases where the input originates from the agent's perceptual system,  $\alpha$  will always be a literal or a conjunction of literals, since natural agents can never perceive anything but features and conjunctions of features. But in cases where automatic inferences occur as a subprocess in a more complex process the input may come from other brain areas, and then  $\alpha$  may in principle be any formula. For this reason we shall work on the full language, and allow any formula to be a premiss or a conclusion of automatic inferences.

The agent's disposition to infer  $\beta$  from  $\alpha$  is naturally represented in the framework by the fact that all the  $<$ -minimal elements in  $\mathcal{U}$  among those that satisfy  $\alpha$  also satisfy  $\beta$ . The reader acquainted with preferential logics will have recognized here a familiar pattern:  $\mathcal{M}$  induces an inference relation on the language exactly in the same way as Kraus, Lehmann and Magidor's cumulative models do<sup>3</sup>. In our framework, the inference relation induced by  $\mathcal{M} = (\mathcal{U}, <)$  shall stand for the set of all the agent's dispositions to infer, that is, for its general knowledge about things. On the logical front, our main task shall be to characterize the inference relations induced by partial worlds models, that is, to identify the logical rules that, according to our account, structure the agent's knowledge about things.

It may be useful to clarify the status of the logical language in the framework. As previously mentioned, the logical language is the language in which we describe the information that flows within the agent's inferential system, where information takes the form of conjunctions of positive and negative features, and more generally, of boolean compositions of features. It follows that the logical language we shall use shall not intend to describe the agent's beliefs: indeed, contrary to propositions features are not subject to be believed nor disbelieved but only to be mentally represented and computed upon

---

<sup>1</sup> See Jackson & Cross (2011) pp. 154–155.

<sup>2</sup> This definition is freely inspired from Leitgeb's book (Leitgeb, 2004).

<sup>3</sup>Kraus et al. (1990), hereafter KLM.

in mental processes. For example, a proposition such as ‘all swans are white’ can be believed or disbelieved by an agent, but features such as ‘white’ or ‘swan’ cannot, nor can their conjunction ‘white swan’, nor either any boolean composition of features. Certainly, some of the mental processes that involve features may result in the agent having some beliefs, but this need not be the case for all. For instance, the mental process by which an agent imagines or recalls a white swan only consists in the reactivation of the corresponding mental representation, and does not, by itself, yield any belief. The same holds for automatic inferences, which are noting but transitions between activation states of mental representations<sup>1</sup>.

Nor should the logical language be taken as the agent’s language, even in the sense of some Fodorian ‘language of thought’. The fact is that the posited computation does not occur at some linguistic level and does not involve any language, even a computational one. Indeed, even though assemblies of concept-neurons can be seen as symbolic representatives of the concepts they support, there are no symbolic representatives of the logical operators: logical operations are directly performed on the assemblies of neurons, as simultaneous activation or lack of activation. It follows that the computation does not involve any language in the agent, for which the logical language could be taken to stand. Rather, the logical language is the language that we use to describe the processing of information in the agent’s brain, quite in the same way that a physicist may use some suitable mathematical language to describe some physical phenomenon.

Therefore, the logical rules that shall be shown to characterize the induced inference relations shall not be interpreted as reasoning rules that the agent applies (which would require some inner language), but as rules it obeys, just as physical phenomena obey physical rules without knowing anything of them.

### 3. The logical setup

We now turn to the formal introduction of the above drafted logical framework. For generality’s sake and easier comparison with other author’s works, throughout this sec-

---

<sup>1</sup>This is not to say, however, that beliefs cannot be represented at all in our framework. If one accepts a rather liberal definition of beliefs according to which beliefs need not be able to be consciously entertained, then dispositions to infer can be seen as some kind of beliefs. For instance, a disposition to infer ‘can fly’ from ‘bird’ can be seen as the belief that birds can fly. More precisely, such a disposition can be seen as the belief that birds *always* can fly if all the agent’s mental representations of birds satisfy the feature ‘can fly’ (that is, if it is a disposition to monotonically infer ‘can fly’ from ‘bird’), and as the belief that birds *generally* can fly, if some of the agent’s less vivid mental representations of birds fail to satisfy the ‘can fly’ feature (that is, if it is a disposition to non-monotonically infer ‘can fly’ from ‘bird’). Similarly, a disposition to infer  $\perp$  (‘false’) from ‘bird & can fly’ can be seen as the belief that birds able to fly do not exist.

If, on the contrary, one does not accept such a liberal definition of beliefs and sticks to the idea that only contents that are, at least in principle, accessible to consciousness can be properly called beliefs, then of course dispositions to infer cannot be regarded as beliefs since they do not satisfy this condition. As mental representations are presumably the means by which processed mental contents can reach awareness, it is mental representations of dispositions to infer rather than dispositions to infer themselves, that should then be called beliefs. By contrast, dispositions to infer should only be regarded as knowledge, in agreement with the view widely held in cognitive psychology that knowledge need not be accessible to consciousness.

This second position is the one we adopt in this paper. Now it should be stressed that the ability to mentally represent one’s own dispositions to infer is clearly additional relative to the ability to draw automatic inferences, and that most natural agents probably lack it. It is unlikely, for instance, that animals such as arthropods can ever be aware of their own dispositions to infer. Therefore, beliefs in this stricter sense are not modelled in the here-presented framework. If one wanted to model them, one would have to extend the framework with a modelling of this additional ability. Namely, one would have to define a meta-language of which the logical expressions denoting dispositions to infer would be formulas, and to endow this meta-language with some modal operator accounting for the fact that the agent has some mental representation of the dispositions to infer denoted by modalised formulas — that is, that it believes (this time in the full sense of the term) the corresponding propositions (for instance, ‘birds generally can fly’, ‘flying birds do not exist’, and so on). This is one of the many possible developments of the present framework.

tion and the following section 4, we shall *not* ask the logical language to be necessarily finite. We shall focus down on finite languages from section 5 on.

The plan of this section is as follows: in subsection 3.1, we define a notion of partial worlds, along with some useful notions of monotonic consequence and equivalence; in subsection 3.2 we provide a brief reminder of KLM's framework, and in subsection 3.3 we take inspiration from their work to define partial worlds models. Then we show that inference relations induced by *smooth* partial worlds models form a strict subclass of KLM's cumulative relations. We identify some supplementary valid rules, but these are not sufficient to reach completeness.

### 3.1 Partial worlds, $\mathcal{U}$ -consequence and $\mathcal{U}$ -equivalence

Let  $\mathcal{L}$  be a (not necessarily finite) propositional language, and  $Var(\mathcal{L})$  the set of propositional variables of  $\mathcal{L}$ . A *partial truth assignment* on  $Var(\mathcal{L})$  is a partial function  $\mathbf{w} : Var(\mathcal{L}) \rightarrow \{0, 1\}$ . A *complete truth assignment* on  $Var(\mathcal{L})$  is a total function  $\mathbf{w}' : Var(\mathcal{L}) \rightarrow \{0, 1\}$ . In line with usual definitions, we regard total functions as partial functions that happen to be defined everywhere, and thus complete truth assignments as partial truth assignments that happen to be defined for all propositional variables.

Each partial truth assignment  $\mathbf{w}$  generates a *partial world*  $w$ , and each total truth assignment  $\mathbf{w}'$  generates a *complete world*  $w'$  (except for the empty truth assignment, for which we state that it doesn't generate any world). We denote  $\mathcal{W}_{\mathcal{L}}$  the set of all complete worlds for  $\mathcal{L}$ , and  $\mathcal{W}_{\mathcal{L}}^p$  the set of all partial worlds for  $\mathcal{L}$  ( $\mathcal{W}_{\mathcal{L}} \subseteq \mathcal{W}_{\mathcal{L}}^p$ ). For each  $w \in \mathcal{W}_{\mathcal{L}}^p$ , we define:

- .  $Var^+(w)$  is the set of propositional variables  $p$  from  $\mathcal{L}$  such that  $\mathbf{w}(p) = 1$ ;
- .  $Var^-(w)$  is the set of propositional variables  $p$  from  $\mathcal{L}$  such that  $\mathbf{w}(p) = 0$ ;
- .  $Lit(w) = Var^+(w) \cup \{\neg p_i / p_i \in Var^-(w)\}$  is the set of literals from  $\mathcal{L}$ .

In the particular case where  $\mathcal{L}$  is finite we use the pair  $(Var^+(w), Var^-(w))$  to denote  $w$ . For example, if  $Var(\mathcal{L}) = \{p, q, r\}$ , the pair  $(\{p\}, \{q\})$  denotes the partial world  $w$  such that  $\mathbf{w}(p) = 1$ ,  $\mathbf{w}(q) = 0$  and  $\mathbf{w}(r)$  is undefined. Moreover (and still in the case where  $\mathcal{L}$  is finite), for each  $w$  in  $\mathcal{W}_{\mathcal{L}}^p$ , we define the  $\mathcal{L}$ -formula  $\delta(w)$ :

$$\delta(w) = \bigwedge \lambda / \lambda \in Lit(w) \text{ } ^1.$$

We call  $\delta(w)$  the *description* of the partial world  $w$ .

*Satisfaction of  $\mathcal{L}$ -formulas by a complete world*  $w$  is denoted  $\models$ , and defined as usual:

- . if  $\alpha$  is a propositional variable, then  $w \models \alpha$  iff  $\mathbf{w}(\alpha) = 1$ ;
- . if  $\alpha = \neg\beta$ , then  $w \models \alpha$  iff  $w \not\models \beta$ ;
- . if  $\alpha = \beta \wedge \gamma$ , then  $w \models \alpha$  iff  $w \models \beta$  and  $w \models \gamma$ ,

other connectives being defined as usual in terms of negation and conjunction. We denote  $\vdash$  the *classical consequence* relation on  $\mathcal{L}$ , and  $\equiv$  the *classical equivalence*. If  $T$  is a set of  $\mathcal{L}$ -formulas, we say that  $w$  satisfies  $T$  and write  $w \models T$  if and only if for any formula  $\alpha \in T$ ,  $w \models \alpha$ .

*Satisfaction of  $\mathcal{L}$ -formulas by a partial world*  $w$  is denoted  $\models\!\!\!\equiv$ , and defined with the help of supervaluations:

For any  $\mathcal{L}$ -formula  $\alpha$  and any partial world  $w$ ,  $w \models\!\!\!\equiv \alpha$  iff for every complete world  $w'$  such that  $w' \models Lit(w)$ ,  $w' \models \alpha$ .

---

<sup>1</sup> This formulation is freely inspired from (Dubois, 2007, pp. 202-203).

In other words, a partial world  $w$  satisfies a  $\mathcal{L}$ -formula  $\alpha$  iff any way one may ‘complete’  $w$  to get a complete world  $w'$ , the latter satisfies  $\alpha$ . One can easily check that if  $w$  itself is a complete world, then  $w \models \alpha$  iff  $w \models \alpha$ . In the particular case where  $\mathcal{L}$  is finite, the above definition comes down to:

$$w \models \alpha \text{ iff for every complete world } w' \text{ such that } w' \models \delta(w), w' \models \alpha, \text{ i.e.} \\ \text{iff } \delta(w) \vdash \alpha.$$

Note that:

- .  $w \models \alpha \wedge \beta$  iff ( $w \models \alpha$  and  $w \models \beta$ );
  - . ( $w \models \alpha$  or  $w \models \beta$ ) entails  $w \models \alpha \vee \beta$ , but
  - .  $w \models \alpha \vee \beta$  *does not* entail ( $w \models \alpha$  or  $w \models \beta$ )
- For example, let  $w$  be such that  $w(p)$  is undefined: then  $w \models p \vee \neg p$ , but  $w \not\models p$  and  $w \not\models \neg p$ .

Let  $\mathcal{U} \subseteq \mathcal{W}_{\mathcal{L}}^p$  be a set of partial worlds. For any set  $T$  of  $\mathcal{L}$ -formulas, we denote  $\mathcal{W}_u(T)$  the (possibly empty) set of partial worlds  $w \in \mathcal{U}$  such that for any formula  $\alpha \in T$ ,  $w \models \alpha$ . When  $T = \{\alpha\}$ , we omit the brackets and simply write  $\mathcal{W}_u(\alpha)$ .

We define on  $\mathcal{L}$  a  $\mathcal{U}$ -consequence relation  $\models_{\mathcal{U}, \mathcal{L}}$ :

$$\alpha \models_{\mathcal{U}, \mathcal{L}} \beta \text{ iff } \mathcal{W}_u(\alpha) \subseteq \mathcal{W}_u(\beta).$$

It is immediate that  $\models_{\mathcal{U}, \mathcal{L}}$  is transitive. It is also supra-classical (i.e.  $\alpha \models_{\mathcal{U}, \mathcal{L}} \beta$  whenever  $\alpha \vdash \beta$ ): indeed, suppose that  $\alpha \vdash \beta$  and let  $w \in \mathcal{W}_u(\alpha)$ : every complete world  $w'$  that satisfies  $\text{Lit}(w)$  also satisfies  $\alpha$  and thus also satisfies  $\beta$ , so  $w \models \beta$ , i.e.  $w \in \mathcal{W}_u(\beta)$ . On the other hand,  $\alpha \models_{\mathcal{U}, \mathcal{L}} \beta$  does *not* imply  $\alpha \vdash \beta$ : for instance, if  $w = (\{p, q\}, \emptyset)$  and  $\mathcal{U} = \{w\}$ , we have  $p \models_{\mathcal{U}, \mathcal{L}} q$ , though  $p \not\vdash q$ . Moreover,  $\models_{\mathcal{U}, \mathcal{L}}$  doesn’t meet contraposition (i.e.  $\alpha \models_{\mathcal{U}, \mathcal{L}} \beta$  doesn’t entail  $\neg\beta \models_{\mathcal{U}, \mathcal{L}} \neg\alpha$ ): for instance, if  $w = (\{p, q\}, \emptyset)$ ,  $w' = (\emptyset, \{q\})$  and  $\mathcal{U} = \{w, w'\}$ , we have  $p \models_{\mathcal{U}, \mathcal{L}} q$ , but  $\neg q \not\models_{\mathcal{U}, \mathcal{L}} \neg p$ , since  $w' \models \neg q$  and  $w' \not\models \neg p$ . Finally, for any  $\mathcal{L}$ -formula  $\alpha$ , we have  $\alpha \models_{\mathcal{U}, \mathcal{L}} \perp$  iff  $\mathcal{W}_u(\alpha) = \emptyset$ .

We also define on  $\mathcal{L}$  a  $\mathcal{U}$ -equivalence relation  $\cong_{\mathcal{U}, \mathcal{L}}$ :

$$\alpha \cong_{\mathcal{U}, \mathcal{L}} \beta \text{ iff } \alpha \models_{\mathcal{U}, \mathcal{L}} \beta \text{ and } \beta \models_{\mathcal{U}, \mathcal{L}} \alpha, \text{ that is,} \\ \text{iff } \mathcal{W}_u(\alpha) = \mathcal{W}_u(\beta).$$

Supra-classicality of  $\models_{\mathcal{U}, \mathcal{L}}$  entails that  $\alpha \cong_{\mathcal{U}, \mathcal{L}} \beta$  whenever  $\alpha \equiv \beta$ .

### 3.2 A brief reminder of Kraus, Lehmann & Magidor’s framework.

We remind the definitions of cumulative and preferential models and their corresponding induced inference relations, as stated in Kraus, Lehmann and Magidor (1990). As usual, if  $E' \subseteq E$  are sets and  $<$  is a binary relation on  $E$ , we say that  $x$  is *<-minimal in  $E'$*  iff  $x \in E'$  and for any  $y$  in  $E'$ ,  $y \not< x$ . We use  $\mathcal{P}$  to denote the power set operator.

Given a propositional language  $\mathcal{L}$ , a *cumulative model* is a triple  $\mathcal{M} = \langle \mathcal{S}, l, \prec \rangle$  where  $\mathcal{S}$  is a set (intuitively interpreted as a set of states),  $l : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{W}_{\mathcal{L}}) - \{\emptyset\}$  is a function that labels every state with a (non-empty)<sup>1</sup> set of complete worlds, and  $\prec$  is a

---

<sup>1</sup> The definition of cumulative models in (Kraus et al., 1990) (def. 5 p. 16) explicitly excludes the possibility of ‘empty labels’, that is, of states  $s$  such that  $l(s) = \emptyset$ . Yet the completeness’ proof (p. 19, between lemmas 10 and 11) makes use of a state  $s$  that is labelled with the empty set ( $s$  is the equivalence class of  $\perp$ ). It appears that the restriction to non-empty labels is useless, as every cumulative model containing such labels can be turned into an equivalent model without empty labels (we shall not prove this here). So we could modify KLM’s definition 5 to allow empty labels. But one can alternatively keep definition 5 as it is and modify instead the completeness’ proof p. 19, taking  $\mathcal{S} = (\mathcal{L}/\sim) - \perp$ ’s equivalence class in place of  $\mathcal{S} = (\mathcal{L}/\sim)$ . Since the construction suggested by KLM is such that for any state  $s \in (\mathcal{L}/\sim) - \perp$ ’s equivalence class,  $s \prec \perp$ ’s equivalence class, one easily checks that suppressing  $\perp$ ’s equivalence class in the model leaves the induced inference relation unaffected. For convenience reasons, it is this second solution that we adopt here.

binary relation (dubbed ‘preference relation’) over states, that satisfies the smoothness condition (defined below).

*Satisfaction of a formula  $\alpha$  by a state  $s$*  is denoted  $s \models \alpha$  and defined by:

$$s \models \alpha \text{ iff for any } w \text{ in } l(s), w \models \alpha.$$

The *smoothness* condition boils down to the following : for any  $s$  in  $\mathcal{S}$  and any formula  $\alpha$ , if  $s \models \alpha$  and  $s$  isn’t  $\prec$ -minimal in  $\{s \in \mathcal{S} / s \models \alpha\}$ , then there is an  $s'$  in  $\mathcal{S}$  such that  $s' \prec s$  and  $s'$  is  $\prec$ -minimal in  $\{s \in \mathcal{S} / s \models \alpha\}$ . For briefness’ sake, when  $s$  is  $\prec$ -minimal in  $\{s \in \mathcal{S} / s \models \alpha\}$ , we say that  $s$  is  $\prec$ -*minimal for  $\alpha$* .

The *inference relation*  $\vdash_{\mathcal{M}}$  induced on  $\mathcal{L}$  by  $\mathcal{M}$  is defined by:

$$\alpha \vdash_{\mathcal{M}} \beta \text{ iff any } s \text{ } \prec\text{-minimal for } \alpha \text{ satisfies } \beta.$$

KLM show that any inference relation  $\vdash_{\mathcal{M}}$  induced by a cumulative model  $\mathcal{M}$  satisfies the rules:

<b>Reflexivity</b>	$\alpha \vdash_{\mathcal{M}} \alpha$
<b>Left Equivalence</b>	if $\alpha \equiv \beta$ and $\alpha \vdash_{\mathcal{M}} \gamma$ , then $\beta \vdash_{\mathcal{M}} \gamma$
<b>Right Weakening</b>	if $\alpha \vdash \beta$ and $\gamma \vdash_{\mathcal{M}} \alpha$ , then $\gamma \vdash_{\mathcal{M}} \beta$
<b>Cut</b>	if $\alpha \wedge \beta \vdash_{\mathcal{M}} \gamma$ and $\alpha \vdash_{\mathcal{M}} \beta$ , then $\alpha \vdash_{\mathcal{M}} \gamma$
<b>Cautious Monotony</b>	if $\alpha \vdash_{\mathcal{M}} \beta$ and $\alpha \vdash_{\mathcal{M}} \gamma$ , then $\alpha \wedge \beta \vdash_{\mathcal{M}} \gamma$

and that conversely, any inference relation  $\vdash$  that satisfies these rules admits a cumulative model. The above set of rules is named *system C*, and relations on the language that satisfy all the rules in C are called *cumulative inference relations*. The authors also give a number of rules that can be derived from C, among which :

<b>And</b>	if $\alpha \vdash_{\mathcal{M}} \beta$ and $\alpha \vdash_{\mathcal{M}} \gamma$ , then $\alpha \vdash_{\mathcal{M}} \beta \wedge \gamma$
<b>Equivalence</b>	if $\alpha \vdash_{\mathcal{M}} \beta$ , $\beta \vdash_{\mathcal{M}} \alpha$ and $\alpha \vdash_{\mathcal{M}} \gamma$ , then $\beta \vdash_{\mathcal{M}} \gamma$

Moreover, one readily gets

<b>Supra-classicality</b>	if $\alpha \vdash \beta$ , then $\alpha \vdash_{\mathcal{M}} \beta$
---------------------------	---

from *Reflexivity* and *Right Weakening*.

A *preferential model* is a cumulative model  $\mathcal{M} = \langle S, l, \prec \rangle$  such that states are labelled by singletons and  $\prec$  is a strict partial order. KLM show that in addition to the rules from C, the following rule *Or* is valid in preferential models:

<b>Or</b>	if $\alpha \vdash_{\mathcal{M}} \gamma$ and $\beta \vdash_{\mathcal{M}} \gamma$ , then $\alpha \vee \beta \vdash_{\mathcal{M}} \gamma$ ,
-----------	--

and that conversely, any inference relation  $\vdash$  satisfying the rules in  $C \cup \{Or\}$  admits a preferential model. The set of rules  $C \cup \{Or\}$  is called *system P*, and the inference relations satisfying all the rules in P are called *preferential relations*.

### 3.3 Partial worlds models and induced inference relations

Let  $\mathcal{L}$  be as above,  $\mathcal{U} \subseteq \mathcal{W}_{\mathcal{L}}^p$  and  $<$  a binary relation on  $\mathcal{U}$ . We call  $\mathcal{M} = (\mathcal{U}, <)$  a *partial worlds model*.

In the particular case where  $\mathcal{L}$  is finite,  $\mathcal{W}_{\mathcal{L}}^p$  is finite and so are the partial worlds in  $\mathcal{W}_{\mathcal{L}}^p$ . So in this case  $\mathcal{U}$  is a finite set of finite partial worlds. We call partial worlds models such that  $\mathcal{U}$  satisfies this condition *finite partial worlds models*.

We say that a partial world  $w \in \mathcal{U}$  is  $<$ -*minimal for  $\alpha$*  if  $w$  is  $<$ -minimal in  $\mathcal{W}_w(\alpha)$ . We define the inference relation  $\vdash_{\mathcal{M}}$  induced on  $\mathcal{L}$  by  $\mathcal{M}$  by:

$$\alpha \vdash_{\mathcal{M}} \beta \text{ iff every } w \text{ } <\text{-minimal for } \alpha \text{ satisfies } \beta.$$

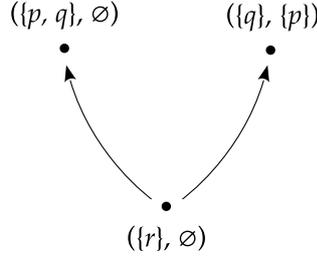


Figure 1.  $\mathcal{M}_1 = (\mathcal{U}_1, <_1)$

We say that  $<$  is  $\mathcal{L}$ -smooth iff for any  $w \in \mathcal{U}$  and any  $\mathcal{L}$ -formula  $\alpha$ , the following holds:  
 if  $w \models \alpha$  and  $w$  isn't  $<$ -minimal for  $\alpha$ , then there is a  $w' \in \mathcal{U}$  such that  $w' < w$  and  $w'$  is  $<$ -minimal for  $\alpha$ .

$\mathcal{L}$ -smoothness is the transposition of usual smoothness in partial worlds' context. If  $\mathcal{M} = (\mathcal{U}, <)$  is a partial worlds model such that  $<$  is  $\mathcal{L}$ -smooth, we call  $\mathcal{M}$  a  $\mathcal{L}$ -smooth partial worlds model.

It turns out that any  $\mathcal{L}$ -smooth partial worlds model can be seen as a cumulative model. Indeed, let  $\mathcal{M} = (\mathcal{U}, <)$  be a  $\mathcal{L}$ -smooth partial worlds model, and  $\vdash_{\mathcal{M}}$  the inference relation induced by  $\mathcal{M}$ . We may build a cumulative model  $\mathcal{M}' = \langle \mathcal{S}, l, \prec \rangle$  such that  $\mathcal{S} = \mathcal{U}$ ,  $\prec = <$ , and  $\vdash_{\mathcal{M}'} = \vdash_{\mathcal{M}}$ . For this we just have to define  $l$  by: for any  $w \in \mathcal{S}$ ,  $l(w) = \{w' \in \mathcal{W}_{\mathcal{L}} / w' \models Lit(w)\}$ . Note that  $l$  is injective since  $\mathcal{U}$  may contain at most one copy of each partial world. We check that for any  $w \in \mathcal{S}$  and any formula  $\alpha$ , we have  $w \models \alpha$  iff  $w \models \alpha$ :

$w \models \alpha$  iff (by  $\models$ 's def.) for any  $w' \in \mathcal{W}_{\mathcal{L}}$  such that  $w' \models Lit(w)$ ,  $w' \models \alpha$ , i.e.  
 iff for any  $w' \in l(w)$ ,  $w' \models \alpha$ , i.e.  
 iff  $w \models \alpha$ .

Consequently, since  $\prec = <$ ,  $w$  is  $<$ -minimal for  $\alpha$  iff it is  $\prec$ -minimal for  $\alpha$ . So, since  $<$  is  $\mathcal{L}$ -smooth,  $\prec$  satisfies smoothness, thus  $\mathcal{M}'$  is a cumulative model. Moreover, since for any  $\alpha$ ,  $\{w \in \mathcal{U} / w \text{ is } <\text{-minimal for } \alpha\} = \{w \in \mathcal{S} / w \text{ is } \prec\text{-minimal for } \alpha\}$ , we have  $\vdash_{\mathcal{M}'} = \vdash_{\mathcal{M}}$ .

It follows that if  $\mathcal{M} = (\mathcal{U}, <)$  is a  $\mathcal{L}$ -smooth partial worlds model, then  $\vdash_{\mathcal{M}}$  is a cumulative inference relation, so KLM's system C is sound for  $\mathcal{L}$ -smooth partial worlds models.

However C is not complete for these same models. For example, the following rule (\*) is valid in  $\mathcal{L}$ -smooth partial worlds models but can't be derived from C:

(\*) For any  $p$  and  $q$  from  $Var(\mathcal{L})$ ,  $(p \vdash_{\mathcal{M}} \perp$  and  $q \vdash_{\mathcal{M}} \perp)$  iff  $p \vee q \vdash_{\mathcal{M}} \perp$

Indeed, for any  $\mathcal{L}$ -smooth partial worlds model  $\mathcal{M} = (\mathcal{U}, <)$  and propositional variables  $p$  and  $q$ , we have  $p \vdash_{\mathcal{M}} \perp$  and  $q \vdash_{\mathcal{M}} \perp$  iff  $\mathcal{W}_u(p) = \mathcal{W}_u(q) = \emptyset$ , iff  $\mathcal{W}_u(p \vee q) = \emptyset$ , iff  $p \vee q \vdash_{\mathcal{M}} \perp$ . That (\*) can't be derived from C follows from the fact that it is not valid in cumulative models. For example the cumulative model  $\mathcal{M}' = \langle \mathcal{S}, l, \prec \rangle$  such that  $\mathcal{S} = \{s\}$ ,  $l(s) = \{(\{p\}, \{q\}), (\{q\}, \{p\})\}$ , and  $\prec = \emptyset$  induces an inference relation  $\vdash_{\mathcal{M}'}$  such that  $p \vdash_{\mathcal{M}'} \perp$ ,  $q \vdash_{\mathcal{M}'} \perp$  and  $p \vee q \not\vdash_{\mathcal{M}'} \perp$  (since  $s \not\models p$  and  $s \not\models q$ , but  $s \models p \vee q$ ).

Yet obviously the rule Or is not valid in  $\mathcal{L}$ -smooth partial worlds models. For example, consider the model  $\mathcal{M}_1 = (\mathcal{U}_1, <_1)$ , where  $\mathcal{U}_1 = \{(\{p, q\}, \emptyset), (\{q\}, \{p\}), (\{r\}, \emptyset)\}$ , and  $<_1 = \{((\{r\}, \emptyset), (\{p, q\}, \emptyset)), ((\{r\}, \emptyset), (\{q\}, \{p\}))\}$ , which may be graphically represented as in Figure 1: we have  $p \vdash_{\mathcal{M}_1} q$  and  $\neg p \vdash_{\mathcal{M}_1} q$ , but  $p \vee \neg p \not\vdash_{\mathcal{M}_1} q$ , since  $(\{r\}, \emptyset)$  is  $<$ -minimal for  $p \vee \neg p$ .

The question arises of the rules we would need to add to  $\mathbf{C}$  to get a sound and complete system for  $\mathcal{L}$ -smooth partial worlds models. Unfortunately, no complete set of rules could be found so far, and in fact there are reasons to believe that reaching completeness is impossible, at least working within a standard propositional language. These reasons shall be discussed in next section. Before that we mention a few rules that even though they do not provide us with completeness, are useful in this context. The following rules are valid in  $\mathcal{L}$ -smooth partial worlds models:

$$\begin{array}{ll}
\mathbf{U-Consistence} & \text{if } \alpha \Vdash_{\mathcal{M}} \perp, \text{ then } \alpha \Vdash_{u, \mathcal{L}} \perp \\
\mathbf{U-Left Equivalence} & \text{if } \alpha \cong_{u, \mathcal{L}} \beta \text{ and } \alpha \Vdash_{\mathcal{M}} \gamma, \text{ then } \beta \Vdash_{\mathcal{M}} \gamma \\
\mathbf{U-Right Weakening} & \text{if } \alpha \Vdash_{u, \mathcal{L}} \beta \text{ and } \gamma \Vdash_{\mathcal{M}} \alpha, \text{ then } \gamma \Vdash_{\mathcal{M}} \beta
\end{array}$$

The proof for *U-Consistence* is immediate using  $\mathcal{L}$ -smoothness and the definitions of  $\Vdash_{\mathcal{M}}$  and  $\Vdash_{u, \mathcal{L}}$ ; *U-Left Equivalence* follows from the fact that  $\alpha \cong_{u, \mathcal{L}} \beta$  iff  $\mathcal{W}_u(\alpha) = \mathcal{W}_u(\beta)$ , and *U-Right Weakening* follows from the fact that  $\alpha \Vdash_{u, \mathcal{L}} \beta$  iff  $\mathcal{W}_u(\alpha) \subseteq \mathcal{W}_u(\beta)$ . Since  $\Vdash_{u, \mathcal{L}}$  is supra-classical, *U-Left Equivalence* and *U-Right Weakening* are slightly stronger than the original *Left Equivalence* and *Right Weakening*, and can be used in place of these. Similarly, the partial-worlds version of *Supra-classicality*,

$$\mathbf{Supra-U-Consequence} \quad \text{if } \alpha \Vdash_{u, \mathcal{L}} \beta, \text{ then } \alpha \Vdash_{\mathcal{M}} \beta$$

that can be derived from *Reflexivity* and *U-Right Weakening*, is slightly stronger than the original *Supra-classicality*.

#### 4. The issue of completeness in partial worlds models' context

To investigate the issue of completeness in partial worlds models' context we shall take inspiration from two papers from Gabbay and Schlechta, namely (Gabbay & Schlechta, 2008, 2009). Although the following discussion does not properly speaking make a proof, it nevertheless gives reasons to believe that no sound and complete set of rules can be found for inference relations induced on  $\mathcal{L}$  by partial worlds models.

In these papers, the authors use a (not necessarily finite) propositional language  $\mathcal{L}$ .  $\mathcal{M}_{\mathcal{L}}$  is the set of classical models of  $\mathcal{L}$ , *i.e.*, in our terminology, of complete worlds for  $\mathcal{L}$  (thus Gabbay and Schlechta's  $\mathcal{M}_{\mathcal{L}}$  is our  $\mathcal{W}_{\mathcal{L}}$ ). For any set  $T$  of  $\mathcal{L}$ -formulas,  $M_{(T)} \subseteq \mathcal{M}_{\mathcal{L}}$  is the set of models of  $T$ .  $D_{\mathcal{L}}$  is the set of definable subsets of  $\mathcal{M}_{\mathcal{L}}$ , that is  $D_{\mathcal{L}} = \{M_{(T)} / T \text{ is a set of } \mathcal{L}\text{-formulas}\}$ . A *preferential structure* is a pair  $\mathcal{M} = \langle U, \prec \rangle$ , where  $U$  is a set and  $\prec$  is a binary relation on  $U$ . Given a set  $\mathcal{Y}$  of sets, a function  $\mu$  of domain  $\mathcal{Y}$  is defined. In the general case where  $U$  may contain several 'copies' of some elements (*i.e.*,  $U$  is in fact a set of pairs  $\langle x, i \rangle$ , with  $i$  an index),  $\mu$  is defined by: for any  $X \in \mathcal{Y}$ ,  $\mu(X) = \{x \in X / \exists \langle x, i \rangle \in U \text{ and } \neg \exists \langle x', j \rangle \in U \text{ s.t. } (x' \in X \text{ and } \langle x', j \rangle \prec \langle x, i \rangle)\}$ . In the particular case where  $U$  contains at most one copy of each element so that we may drop the indexes,  $\mu(X)$  is simply defined as  $\{x / x \text{ is } \prec\text{-minimal in } X \cap U\}$ . The intended interpretation of  $\mu$  is that of a minimization function (note however that  $\mu(X)$  may be empty, notably in the case where  $X \cap \{x / \exists i \text{ s.t. } \langle x, i \rangle \in U\}$  is). When  $\mathcal{Y} \subseteq \mathcal{P}(\mathcal{M}_{\mathcal{L}})$ ,  $\mu$  gives rise to an inference relation  $\sim$  in the usual way: for any set  $T$  of  $\mathcal{L}$ -formulas and any  $\mathcal{L}$ -formula  $\alpha$ ,  $T \sim \alpha$  if and only if  $\mu(M_{(T)}) \subseteq M_{\{\alpha\}}$ . Therefore, preferential structures are a generalization of KLM's preferential models. The authors study the algebraic properties of preferential structures and, in the cases where  $\mathcal{Y} \subseteq \mathcal{P}(\mathcal{M}_{\mathcal{L}})$ , the correspondence between these properties and the usual rules for inference relations (for instance, the rule *Reflexivity* corresponds to the fact that for any  $X \in (\mathcal{Y} \cap D_{\mathcal{L}})$ ,  $\mu(X) \subseteq X$ ). Among these, the closure properties of the domain of  $\mu$  reveal themselves of special importance. In particular, in (Gabbay & Schlechta, 2008) the authors show that when smoothness

is not satisfied and the domain of  $\mu$  is not closed under finite unions, the algebraic counterpart of cumulativity breaks up into an infinity of non-equivalent properties, which jeopardizes the chances of getting a representation theorem.

Of course, we can't straightforwardly use their results for our own purposes, since the models we consider satisfy smoothness and moreover they are built out of partial worlds. But we can take inspiration from their approach, and in particular we can check for definability and closure properties issues. This requires that we first accommodate their tools to partial worlds models, and first of all that we set forth a suitable notion of definability, which we shall do as follows:

If  $\mathcal{L}$  is a (not necessarily finite) propositional language,  $\mathcal{U} \subseteq \mathcal{W}_{\mathcal{L}}^{\mathcal{P}}$  and  $A \subseteq \mathcal{U}$ , we shall say that  $A$  is  $\mathcal{L}$ -definable in  $\mathcal{U}$  if and only if there is a  $\mathcal{L}$ -formula  $\alpha$  such that  $A = \mathcal{W}_u(\alpha)$ . For instance,  $\{(\{p, q\}, \emptyset), (\{p\}, \{q\})\}$  is  $\mathcal{L}$ -definable in  $\mathcal{U} = \{(\{p, q\}, \emptyset), (\{p\}, \{q\}), (\{r\}, \emptyset)\}$ , but not in  $\mathcal{U}' = \{(\{p, q\}, \emptyset), (\{p\}, \{q\}), (\{p, r\}, \emptyset)\}$ , since for any  $\mathcal{L}$ -formula  $\alpha$  such that  $(\{p, q\}, \emptyset) \models \alpha$  and  $(\{p\}, \{q\}) \models \alpha$ , we also have  $(\{p, r\}, \emptyset) \models \alpha$ .

We shall denote  $D_{\mathcal{U}, \mathcal{L}}$  the set of  $\mathcal{L}$ -definable (in  $\mathcal{U}$ ) subsets of  $\mathcal{U}$ . So, formally,  $D_{\mathcal{U}, \mathcal{L}} = \{A \subseteq \mathcal{U} / \exists \alpha \in \mathcal{L} \text{ such that } A = \mathcal{W}_u(\alpha)\}$ . As we shall always consider the definability of a given subset of  $\mathcal{U}$  in  $\mathcal{U}$  itself, we shall, for short, speak of  $\mathcal{L}$ -definable subsets of  $\mathcal{U}$ .  $D_{\mathcal{U}, \mathcal{L}}$  is closed under finite intersections, since for any  $\mathcal{L}$ -formulas  $\alpha$  and  $\beta$ ,  $\mathcal{W}_u(\alpha) \cap \mathcal{W}_u(\beta) = \mathcal{W}_u(\alpha \wedge \beta)$ . But  $D_{\mathcal{U}, \mathcal{L}}$  is generally *not* closed under finite unions. For instance, if  $\mathcal{U} = \{(\{p, q\}, \emptyset), (\{p\}, \{q\}), (\{p, r\}, \emptyset)\}$ , then  $\mathcal{W}_u(p \wedge q) \cup \mathcal{W}_u(p \wedge \neg q) = \{(\{p, q\}, \emptyset), (\{p\}, \{q\})\} \notin D_{\mathcal{U}, \mathcal{L}}$ .

It is well known that a preferential model  $\langle S, l, \prec \rangle$  can be seen as a pair  $(U, \prec')$ , where  $U \subseteq (\mathcal{W}_{\mathcal{L}} \times I)$  (with  $I$  an index set) and  $\prec'$  is a preference relation on  $U$ : one just has to replace each state  $s$  in  $S$  with a suitably indexed copy  $w_i$  of the complete world  $w$  such that  $w = l(s)$ , and let  $\prec' = \prec$ . Similarly, a cumulative model  $\mathcal{M} = \langle S, l, \prec \rangle$  can be seen as a pair  $(V, \prec'')$ , where  $V \subseteq \mathcal{P}(\mathcal{W}_{\mathcal{L}})$  and  $\prec''$  is a preference relation on  $V$ . One just has to consider the cumulative relation  $\vdash_{\mathcal{M}}$  induced by  $\mathcal{M}$  and to follow the instructions given by KLM in their representation theorem<sup>1</sup> to build the corresponding cumulative model  $\mathcal{M}' = \langle S', l', \prec' \rangle$  of  $\vdash_{\mathcal{M}}$ . Then, since by construction  $l'$  is injective, one just has to replace each state  $s$  in  $S'$  with  $l'(s)$ , and to let  $\prec'' = \prec'$  to get  $(V, \prec'')$ .

It should also be recalled that KLM's demand that  $\prec$  be a strict partial order in preferential models only intends to make things 'nicer' but adds logically nothing, so that in preferential models as in cumulative ones, all that is finally required for  $\prec$  is that it satisfies smoothness. It follows that what essentially distinguishes preferential from cumulative models is that in the former the preference relation can be seen as defined over a set of (indexed) complete worlds, while in the latter it can be seen as defined over a set of *sets of* complete worlds.

Finally, it is also well known that on the logical side this difference corresponds to the addition of the rule *Or* to *C*. *Or* can therefore be seen as allowing the existence of models such that  $\prec$  is defined over a set of worlds.

Now, an essential characteristic of partial worlds models is precisely that the relation  $\prec$  holds between partial worlds rather than between sets of (partial) worlds. A complete set of rules for the inference relations they induce should thus contain some rule(s) corresponding to this property. But as we saw, *Or* is not valid in partial worlds models.

Gabbay and Schlechta's algebraic approach sheds some light on how the rule *Or*

---

<sup>1</sup>(Kraus et al., 1990), section 3.5.

produces its effect in the context of preferential relations. In (Gabbay & Schlechta, 2009), they show that the algebraic counterpart of the rule  $Or$  is the property  $(\mu Or)$ :

$$(\mu Or) \quad \text{For any } X \text{ and } Y \text{ in the domain, } \mu(X \cup Y) \subseteq \mu(X) \cup \mu(Y)$$

where  $\mu$  is the minimalization function. More precisely, they show that for any relation  $\sim$  on  $\mathcal{L}$  satisfying *Left Equivalence* and *Right Weakening*, the function  $\mu$  of domain  $D_{\mathcal{L}}$  defined by:  $\mu(M(T)) = M(\bar{T})$  (where  $\bar{T} = \{\mathcal{L}\text{-formulas } \alpha / T \sim \alpha\}$ ) satisfies  $(\mu Or)$  if and only if  $\sim$  satisfies  $Or^1$ .

Furthermore, one easily shows that for any set  $Z$  of pairs  $\langle x, i \rangle$ , any binary relation  $\prec$  on  $Z$  and any set  $\mathcal{Y}$  of sets, if  $\mu$  is a function of domain  $\mathcal{Y}$  defined by: for any  $X \in \mathcal{Y}$ ,  $\mu(X) = \{x \in X / \exists \langle x, i \rangle \in Z \text{ and } \neg \exists \langle x', j \rangle \in Z \text{ s.t. } (x' \in X \text{ and } \langle x', j \rangle \prec \langle x, i \rangle)\}$ , then  $\mu$  satisfies  $(\mu Or)$  provided that  $\mathcal{Y}$  is closed under finite unions. Indeed let  $X$  and  $Y$  be in  $\mathcal{Y}$ , and suppose that there is some  $x \in \mu(X \cup Y)$  such that  $x \notin \mu(X) \cup \mu(Y)$ . By  $\mu$ 's definition,  $\mu(X \cup Y) \subseteq X \cup Y$ , thus  $x \in X$  or  $x \in Y$ . If  $x \in X$ , then since  $x \notin \mu(X)$ , by  $\mu$ 's definition again there is a pair  $\langle x', j \rangle$  in  $Z$  such that  $x' \in X$  and  $\langle x', j \rangle \prec \langle x, i \rangle$ .  $x' \in X \cup Y$ , thus  $x \notin \mu(X \cup Y)$ : a contradiction. And similarly if  $x \in Y$ . It follows that for any set of pairs  $Z$ , any set  $\mathcal{Y}$  of sets and any function  $\mu$  of domain  $\mathcal{Y}$ , if  $\mu$  falsifies  $(\mu Or)$ , then there can be no binary relation  $\prec$  on  $Z$  such that for any  $X \in \mathcal{Y}$ ,  $\mu(X) = \{x \in X / \exists \langle x, i \rangle \in Z \text{ and } \neg \exists \langle x', j \rangle \in Z \text{ s.t. } (x' \in X \text{ and } \langle x', j \rangle \prec \langle x, i \rangle)\}$ . Simply put, this means that if  $\mu$  falsifies  $(\mu Or)$ , there can be no relation  $\prec$  on  $Z$  such that  $\mu$  is the minimalization function induced by  $\prec$ . Note that in all cases the closure of  $\mathcal{Y}$  under finite unions is required because otherwise  $\mu(X \cup Y)$  may not be defined, and then the definition of  $(\mu Or)$  does not make sense. These results can be easily transposed to the particular case where  $Z$  contains at most one copy of each element, and thus to the even more particular case where  $Z$  is a set of partial worlds.

In the particular case of partial worlds models,  $Z$  is  $\mathcal{U}$  and  $\prec$  is  $<$ , thus by the above it indeed holds that for any  $\mathcal{L}$ -formulas  $\alpha$  and  $\beta$ ,  $\{w/w \text{ is } <\text{-minimal in } \mathcal{W}_u(\alpha) \cup \mathcal{W}_u(\beta)\} \subseteq \{w/w \text{ is } <\text{-minimal in } \mathcal{W}_u(\alpha)\} \cup \{w/w \text{ is } <\text{-minimal in } \mathcal{W}_u(\beta)\}$ . But because  $\sim_m$  can only grasp  $\mathcal{L}$ -definable subsets of  $\mathcal{U}$ , it seems that the relation  $\sim_m$  is unable to express this property. Indeed, any function  $\mu$  defined by the means of  $\sim_m$  (using  $\mu(\mathcal{W}_u(\alpha)) = \mathcal{W}_u(\{\bar{\alpha}\})$  with  $\{\bar{\alpha}\} = \{\beta/\alpha \sim_m \beta\}$ , or any definition we would like) will always have  $D_{\mathcal{U}, \mathcal{L}}$  for domain. But  $D_{\mathcal{U}, \mathcal{L}}$  need not be closed under finite unions, and in cases it is not  $(\mu Or)$  will not be defined for the whole  $D_{\mathcal{U}, \mathcal{L}}$  but only for the subsets of  $D_{\mathcal{U}, \mathcal{L}}$  that are closed under finite unions. It can be assumed that a suitably defined function  $\mu$  will indeed satisfy  $(\mu Or)$  within each of these subsets, but still there will be several ways in which to extend  $\mu$  so as to close  $D_{\mathcal{U}, \mathcal{L}}$  under finite unions, some of which do not satisfy  $(\mu Or)$ . In other words,  $\mu$  will not ‘say enough’ to rule out the possibility that no relation  $<$  exists on  $\mathcal{U}$  such that  $\mu$  is a fragment of the minimalization function induced by  $<$ . It thus seems that whatever the valid rules we may find for  $\sim_m$ , these will not, in the general case, be enough to ensure the existence of a suitable relation  $<$  on  $\mathcal{U}$ , hence to get a representation theorem. The existence of a complete set of rules for the inference relations induced on  $\mathcal{L}$  by partial worlds models therefore appears very doubtful.

## 5. Augmenting the language

The absence of closure of  $D_{\mathcal{U}, \mathcal{L}}$  under finite unions can be seen as the consequence of a lack of expressivity of  $\mathcal{L}$  with regard to partial worlds. Indeed, using  $\mathcal{L}$  one can express

---

<sup>1</sup>(Gabbay & Schlechta, 2009), proposition 3.8.

the fact that a partial world  $w$  satisfies  $\alpha \vee \beta$ , but not that: it satisfies  $\alpha$  or it satisfies  $\beta$ . In the context of complete worlds, the two are equivalent, but not in that of partial worlds. A simple way to close  $D_{\mathcal{U}, \mathcal{L}}$  under finite unions is thus to add a suitable new connective to the language, so as to allow it to express the fact that a partial world  $w$  satisfies  $\alpha$  or satisfies  $\beta$ . This is what we shall do in what follows. For simplicity, and because for our purposes we only need finite languages, in doing so we shall only consider finite languages.

Using a finite propositional language  $\mathbf{L}$  as a basis (so  $\mathbf{L}$  is a particular case of the language  $\mathcal{L}$  we used so far), in subsection 5.1 we supplement it with two new binary connectives<sup>1</sup>  $\parallel$  and  $\wedge$  (it shall be shown in further subsections that  $\wedge$  brings in fact nothing to the expressivity of the augmented language, but is added for convenience). In 5.2 we transpose the previously defined notions to the resulting language  $\mathbf{L}^\parallel$ , and in 5.3 we provide a number of valid rules for inference relations induced on  $\mathbf{L}^\parallel$  by  $\mathbf{L}^\parallel$ -smooth partial worlds models. Finally in subsection 5.4 we introduce a couple of useful additional definitions.

### 5.1 The language $\mathbf{L}^\parallel$

Let  $\mathbf{L}$  be a *finite* propositional language. We define the language  $\mathbf{L}^\parallel$  by:

- . If  $\alpha$  is a  $\mathbf{L}$ -formula, then  $\alpha$  is a  $\mathbf{L}^\parallel$ -formula (*i.e.*  $\mathbf{L} \subseteq \mathbf{L}^\parallel$ );
- . If  $\alpha$  and  $\beta$  are  $\mathbf{L}^\parallel$ -formulas, then  $\alpha \wedge \beta$  and  $\alpha \parallel \beta$  are  $\mathbf{L}^\parallel$ -formulas;
- . Nothing else is a  $\mathbf{L}^\parallel$ -formula.

Satisfaction of a  $\mathbf{L}^\parallel$ -formula  $\alpha$  by a partial world  $w$  is denoted  $w \models \alpha$  as before. It is defined by the following clauses:

- . If  $\alpha$  is a  $\mathbf{L}$ -formula, then  $w \models \alpha$  iff  $\delta(w) \vdash \alpha$  (as before).
- . If  $\alpha = \beta \parallel \gamma$ , then  $w \models \alpha$  iff ( $w \models \beta$  or  $w \models \gamma$ );
- . If  $\alpha = \beta \wedge \gamma$ , then  $w \models \alpha$  iff ( $w \models \beta$  and  $w \models \gamma$ ).

As before, we note  $\mathcal{W}_u(\alpha)$  the set of partial worlds  $w \in \mathcal{U}$  such that  $w \models \alpha$ . In the particular case where  $w$  is a complete world, one easily checks that  $w \models \beta \parallel \gamma$  iff  $w \models \beta \vee \gamma$  iff  $w \models \beta \vee \gamma$ . In the general case,  $w \models \beta \parallel \gamma$  implies  $w \models \beta \vee \gamma$  but not conversely. Furthermore,  $\parallel$  is associative: for any  $\mathbf{L}^\parallel$ -formulas  $\alpha$ ,  $\beta$  and  $\gamma$ , and any partial world  $w$ ,  $w \models \alpha \parallel (\beta \parallel \gamma)$  iff  $w \models (\alpha \parallel \beta) \parallel \gamma$ . Consequently in the sequel we may drop the brackets, and simply write  $\alpha \parallel \beta \parallel \gamma$ . If  $X = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$  is a set of  $\mathbf{L}^\parallel$ -formulas, we may write  $\parallel \alpha_i / (\alpha_i \in X)$  as an abbreviation for the  $\mathbf{L}^\parallel$ -formula  $\alpha_1 \parallel \alpha_2 \dots \parallel \alpha_n$ . Finally, it directly follows from the definition of  $\parallel$  that  $\mathcal{W}_u(\alpha \parallel \beta) = \mathcal{W}_u(\alpha) \cup \mathcal{W}_u(\beta)$ .

As for the new connective  $\wedge$ , for now we may just remark that it operates on  $\mathbf{L}^\parallel$ -formulas exactly as the usual  $\wedge$  does on  $\mathbf{L}$ -formulas, since for any  $\mathbf{L}$ -formulas  $\alpha$  and  $\beta$  and any partial world  $w$ ,  $w \models \alpha \wedge \beta$  iff  $w \models \alpha$  and  $w \models \beta$ .  $\wedge$  can thus be seen as an extension of the usual  $\wedge$  to formulas containing  $\parallel$ .

One may wonder whether the introduction of the connective  $\parallel$  justifies itself by more than purely technical reasons, that is if  $\parallel$  has some natural intuitive interpretation in the framework. It turns out that  $\parallel$  is the syntactical counterpart of a mental operation that is commonly performed on information by natural agents. Namely,  $\parallel$ -disjunctions appear to be the primary form in which disjunction takes place in living beings.

The fact is that each time an agent considers different cases or options (as, for instance,

---

<sup>1</sup>A symbol  $\parallel$  is also used by Gabbay and Schlechta, with an algebraic meaning. The use we shall make here of this symbol is different, and should not be confused with the one made by these authors (anyway the context of use suffices to disambiguate, since here we use it as a connective).

to decide which case is the most plausible, or which action is to be taken), it forms a disjunction of this kind. Indeed, to be uncertain whether a given object/situation satisfies the conjunction of features  $\alpha$  or the conjunction of features  $\beta$  is nothing but to regard it as satisfying at least one of these two, that is, as satisfying  $\alpha \parallel \beta$ . To decide which of these two possibilities is the more likely, the agent feeds its inferential system with the information  $\gamma \wedge (\alpha \parallel \beta)$  (where  $\gamma$  is the conjunction of all the features it currently holds the object to satisfy), and checks for the outcome. Obviously this mental operation cannot be rendered using the usual disjunction  $\vee$ , since  $\gamma \wedge (p \vee \neg p) \equiv \gamma \wedge (q \vee \neg q) \equiv \gamma$ , while being uncertain whether a given object satisfies the feature represented by  $p$  rather than its negation  $\neg p$  is not the same as being uncertain whether it satisfies the feature represented by  $q$  rather than its negation  $\neg q$ , nor is it the same as not having such hesitations.

It should be stressed that realizing this kind of disjunctions does not require any particular cognitive ability from the agent, since having the idea that a given object satisfies  $\alpha \parallel \beta$  is simply not to have decided between two competing mental representations of the object, the one according to which it satisfies  $\alpha$  and the other according to which it satisfies  $\beta$ . According to the neural model suggested in section 2.1.2,  $\parallel$ -disjunctions are straightforwardly implemented in the agent's inferential system as unions of sets (assemblies) of neurons. More specifically, if  $A$  is the set of neurons that represents the information  $\alpha$  in the agent's inferential system (that is, if  $A$  is the set of concept-neurons that support its mental representations of objects that, in its opinion, satisfy  $\alpha$ ) and if similarly  $B$  is the set of neurons that represents the information  $\beta$ , then  $\alpha \parallel \beta$  is simply represented by  $A \cup B$ . By contrast,  $\vee$ -disjunctions cannot be represented by set-theoretical operations over concept-neurons assemblies, which makes it likely that they can only be implemented at the language level, provided of course that the considered agent has language abilities. It is thus probable that  $\parallel$ -disjunctions are the primary form in which disjunction occurs in natural agents, and that  $\vee$ -disjunctions are on the contrary a secondary construct relying on higher cognitive abilities. This renders the use of the connective  $\parallel$  very natural in our context.

Interestingly, it turns out that our satisfaction clauses for  $L^\parallel$ -formulas are similar to the satisfaction clauses for  $L$ -formulas given by Kripke for intuitionistic models<sup>1</sup>. More precisely, a partial world  $w \in \mathcal{W}_L^p$  can be mimicked using a Kripke model  $\Phi: (Var(L) \times \mathbf{K}) \rightarrow \{T, F\}$  defined on a model structure  $(\mathbf{G}, \mathbf{K}, \mathbf{R})$  such that  $\mathbf{G} = w$ ,  $\mathbf{K} = \{w\} \cup \{w' \in \mathcal{W}_L / \delta(w') \vdash \delta(w)\}$ ,  $\mathbf{R}$  is the reflexive closure of  $\{(w, w') / w' \in \mathbf{K}\}$  and for any  $p \in Var(L)$  and any  $w' \in \mathbf{K}$ ,  $\Phi(p, w') = T$  iff  $w'(p) = 1$ , and  $\Phi(p, w') = F$  otherwise. Indeed it is immediate that for any  $p \in Var(L)$ ,  $w \models p$  iff  $\Phi(p, w) = T$ , and using Kripke's clauses to extend  $\Phi$  to formulas one gets that if  $\alpha$  is a  $L$ -formula that contain no other connectives than  $\neg$  and  $\wedge$ , then  $w \models \alpha$  iff  $\Phi(\alpha, w) = T$  (the proof is by induction on the construction of formulas). Then using our satisfaction clause for  $\parallel$  on the one side and Kripke's clause for  $\vee$  on the other, one gets that if  $\beta$  is a  $L$ -formula that contain no other connectives than  $\neg$  and  $\wedge$ , then  $w \models \alpha \parallel \beta$  iff  $\Phi(\alpha \vee \beta, w) = T$ . Thus the disjunction denoted by  $\parallel$  can be seen in a certain sense as 'intuitionistic'. However, and contrary to what is the general case in Kripke models, here it holds that  $\Phi(\alpha, w) = T$  iff  $\Phi(\neg\neg\alpha, w) = T$ , i.e., that  $w \models \alpha$  iff  $w \models \neg\neg\alpha$ .

## 5.2 Transposition of the previous definitions in $L^\parallel$ -context

We shall now transpose the previously defined notions —  $\mathcal{U}$ -consequence, definability,

---

<sup>1</sup> See (Kripke, 1965, pp. 94 – 100). This similarity was suggested by K. Schlechta in a personal communication.

smoothness, inference relations and so on — to  $\mathbf{L}^\parallel$ -context.

Let  $\mathcal{U} \subseteq \mathcal{W}_L^\parallel$ . Since  $\mathbf{L}$  is finite,  $\mathcal{U}$  is a finite set of finite partial worlds.

We extend the relation of  $\mathcal{U}$ -consequence to  $\mathbf{L}^\parallel$ :

For any  $\mathbf{L}^\parallel$ -formulas  $\alpha$  and  $\beta$ ,  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \beta$  iff  $\mathcal{W}_u(\alpha) \subseteq \mathcal{W}_u(\beta)$ .

It is immediate that  $\Vdash_{\mathcal{U}, \mathbf{L}^\parallel}$  is transitive, and that  $\Vdash_{\mathcal{U}, \mathbf{L}} \subseteq \Vdash_{\mathcal{U}, \mathbf{L}^\parallel}$ . Note that:

- .  $\Vdash_{\mathcal{U}, \mathbf{L}^\parallel}$  is supra-classical, since  $\Vdash_{\mathcal{U}, \mathbf{L}}$  is.
- . For any  $\mathbf{L}^\parallel$  formula  $\alpha$ ,  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \perp$  iff  $\mathcal{W}_u(\alpha) = \emptyset$ .
- . For any  $\mathbf{L}^\parallel$  formulas  $\alpha, \beta$  and  $\gamma$ ,  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \gamma$  and  $\beta \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \gamma$  iff  $\alpha \parallel \beta \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \gamma$ .
- . For any  $\mathbf{L}^\parallel$  formulas  $\alpha, \beta$  and  $\gamma$ ,  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \beta$  and  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \gamma$  iff  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \beta \wedge \gamma$ .
- . For any  $w \in \mathcal{U}$  and any  $\mathbf{L}^\parallel$  formula  $\alpha$ ,  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \alpha$  iff  $w \models \alpha$  (see proof in Appendix A). Thus, in the particular case where  $\alpha$  is a  $\mathbf{L}$ -formula,  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \alpha$  iff  $w \models \alpha$  iff  $\delta(w) \vdash \alpha$ .

Similarly, we extend the relation of  $\mathcal{U}$ -equivalence to  $\mathbf{L}^\parallel$ :

For any  $\mathbf{L}^\parallel$ -formulas  $\alpha$  and  $\beta$ ,  $\alpha \cong_{\mathcal{U}, \mathbf{L}^\parallel} \beta$  iff  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \beta$  and  $\beta \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \alpha$ , that is, iff  $\mathcal{W}_u(\alpha) = \mathcal{W}_u(\beta)$ .

As before for  $\cong_{\mathcal{U}, \mathbf{L}}$ , supra-classicality of  $\Vdash_{\mathcal{U}, \mathbf{L}^\parallel}$  entails that  $\alpha \cong_{\mathcal{U}, \mathbf{L}^\parallel} \beta$  whenever  $\alpha \equiv \beta$ . Furthermore, one easily checks that:

- .  $(\alpha \parallel \beta) \parallel (\beta \parallel \gamma) \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha \parallel \beta \parallel \gamma$ .
- .  $(\alpha \parallel \beta \parallel \gamma) \wedge (\beta \parallel \gamma) \cong_{\mathcal{U}, \mathbf{L}^\parallel} \beta \parallel \gamma$ .
- .  $\perp \parallel \alpha \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha$ .
- . If  $\alpha \Vdash_{\mathcal{U}, \mathbf{L}^\parallel} \perp$ , then  $\alpha \cong_{\mathcal{U}, \mathbf{L}^\parallel} \|\delta(w)/(w \in \mathcal{W}_u(\alpha))\|$ .
- . If  $\alpha \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha'$ , then  $\alpha \parallel \beta \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha' \parallel \beta$  ('substitution of  $\mathcal{U}$ -equivalent disjuncts').
- . If  $\alpha$  and  $\beta$  are  $\mathbf{L}$ -formulas, then  $\alpha \wedge \beta \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha \wedge \beta$ .
- .  $(\alpha_1 \parallel \alpha_2) \wedge \beta \cong_{\mathcal{U}, \mathbf{L}^\parallel} (\alpha_1 \wedge \beta) \parallel (\alpha_2 \wedge \beta)$ .

Note that the last three facts taken together entail that any  $\mathbf{L}^\parallel$ -formula  $\alpha$  containing  $\wedge$  is  $\mathcal{U}$ -equivalent to a  $\mathbf{L}^\parallel$ -formula  $\alpha'$  that doesn't contain  $\wedge$ , but contains the usual  $\wedge$  instead.

Given  $A \subseteq \mathcal{U}$ , we shall say that  $A$  is  $\mathbf{L}^\parallel$ -definable (in  $\mathcal{U}$ ) if there is a  $\mathbf{L}^\parallel$ -formula  $\alpha$  such that  $A = \mathcal{W}_u(\alpha)$ . We denote  $D_{\mathcal{U}, \mathbf{L}^\parallel}$  the set of  $\mathbf{L}^\parallel$ -definable (in  $\mathcal{U}$ ) subsets of  $\mathcal{U}$ . As we shall always consider the  $\mathbf{L}^\parallel$ -definability of a given subset of  $\mathcal{U}$  in  $\mathcal{U}$  itself, we shall, for short, speak of  $\mathbf{L}^\parallel$ -definable subsets of  $\mathcal{U}$ . Since  $\mathbf{L} \subseteq \mathbf{L}^\parallel$ , we have  $D_{\mathcal{U}, \mathbf{L}} \subseteq D_{\mathcal{U}, \mathbf{L}^\parallel}$ . Note that even though  $\mathcal{U}$  is finite, generally  $D_{\mathcal{U}, \mathbf{L}^\parallel} \neq \mathcal{P}(\mathcal{U})$ . For instance, let  $\mathcal{U} = \{(\{p, q\}, \emptyset), (\{p\}, \emptyset)\}$ ; then  $\{(\{p\}, \emptyset)\} \notin D_{\mathcal{U}, \mathbf{L}^\parallel}$ , since any  $\mathbf{L}^\parallel$  formula that is satisfied by  $(\{p\}, \emptyset)$  is also satisfied by  $(\{p, q\}, \emptyset)$ . But contrary to  $D_{\mathcal{U}, \mathbf{L}}$ ,  $D_{\mathcal{U}, \mathbf{L}^\parallel}$  is closed under finite unions: if  $A$  and  $B \in D_{\mathcal{U}, \mathbf{L}^\parallel}$ , that is if there are some  $\mathbf{L}^\parallel$ -formulas  $\alpha$  and  $\beta$  such that  $A = \mathcal{W}_u(\alpha)$  and  $B = \mathcal{W}_u(\beta)$ , then  $A \cup B = \mathcal{W}_u(\alpha) \cup \mathcal{W}_u(\beta) = \mathcal{W}_u(\alpha \parallel \beta)$ , so  $A \cup B \in D_{\mathcal{U}, \mathbf{L}^\parallel}$ .

Moreover, since any  $\mathbf{L}^\parallel$ -formula  $\alpha$  containing  $\wedge$  is  $\mathcal{U}$ -equivalent to a  $\mathbf{L}^\parallel$ -formula  $\alpha'$  that does not contain  $\wedge$ , any subset of  $\mathcal{U}$  which is definable by a  $\mathbf{L}^\parallel$ -formula containing  $\wedge$  is also definable by a  $\mathbf{L}^\parallel$ -formula not containing  $\wedge$ . Thus  $\wedge$  adds nothing to the expressivity of  $\mathbf{L}^\parallel$ , and from a logical point of view it would have been strictly equivalent to define  $\mathbf{L}^\parallel$  by the sole addition of  $\parallel$ . We did not for convenience reasons, since this will allow us to adapt to  $\mathbf{L}^\parallel$  the usual rules for inference relations. Now, as  $\wedge$  is but an extension of the usual  $\wedge$  to formulas containing  $\parallel$ , in the sequel we shall drop the special writing  $\wedge$  to denote conjunctions of formulas containing  $\parallel$ , and use indifferently  $\wedge$  for all conjunctions of  $\mathbf{L}^\parallel$  formulas.

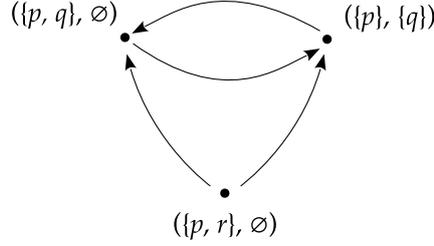


Figure 2.  $\mathcal{M}_2 = (\mathcal{U}_2, <_2)$

Let  $<$  be a binary relation on  $\mathcal{U}$ . We denote  $\|\sim_{\mathcal{M}}$  the inference relation induced on  $\mathbf{L}^{\parallel}$  by  $\mathcal{M} = (\mathcal{U}, <)$  in the usual manner, that is:

For any  $\mathbf{L}^{\parallel}$ -formulas  $\alpha$  and  $\beta$ ,  $\alpha \|\sim_{\mathcal{M}} \beta$  iff every  $w$   $<$ -minimal in  $\mathcal{W}_u(\alpha)$  satisfies  $\beta$ .

We say that  $<$  is  $\mathbf{L}^{\parallel}$ -smooth if and only if for any  $w$  in  $\mathcal{U}$  and any  $\mathbf{L}^{\parallel}$ -formula  $\alpha$ , the following holds:

If  $w \models \alpha$  and  $w$  isn't  $<$ -minimal in  $\mathcal{W}_u(\alpha)$ , then there is a  $w'$  in  $\mathcal{U}$  such that  $w' < w$  and  $w'$  is  $<$ -minimal in  $\mathcal{W}_u(\alpha)$ .

Since  $\mathbf{L} \subseteq \mathbf{L}^{\parallel}$ , any  $\mathbf{L}^{\parallel}$ -smooth relation is also  $\mathbf{L}$ -smooth, but not conversely. For instance, consider the model  $\mathcal{M}_2 = (\mathcal{U}_2, <_2)$  showed in figure 2:  $<$  is  $\mathbf{L}$ -smooth, since any  $\mathbf{L}$ -formula satisfied by both  $(\{p, q\}, \emptyset)$  and  $(\{p\}, \{q\})$  is also satisfied by  $(\{p, r\}, \emptyset)$ . But  $<$  is not  $\mathbf{L}^{\parallel}$ -smooth, since there is no  $<$ -minimal world for  $(p \wedge q) \parallel (p \wedge \neg q)$ .

If  $\mathcal{M} = (\mathcal{U}, <)$  is a partial worlds model such that  $<$  is  $\mathbf{L}^{\parallel}$ -smooth, we call  $\mathcal{M}$  a  $\mathbf{L}^{\parallel}$ -smooth partial worlds model. Any  $\mathbf{L}^{\parallel}$ -smooth model is a  $\mathbf{L}$ -smooth model, but not conversely.

### 5.3 Inference relations induced on $\mathbf{L}^{\parallel}$ by $\mathbf{L}^{\parallel}$ -smooth partial worlds models

We say that a  $\mathbf{L}^{\parallel}$ -formula  $\alpha$  is *maximal-consistent in  $\mathcal{U}$*  if

- i)  $\alpha \in \mathbf{L}$ ,
- ii)  $\alpha \not\|\sim_{\mathcal{U}, \mathbf{L}^{\parallel}} \perp$ , and
- iii) For any  $\mathbf{L}$ -formula  $\beta$  such that  $\alpha \not\vdash \beta$ ,  $\alpha \wedge \beta \not\|\sim_{\mathcal{U}, \mathbf{L}^{\parallel}} \perp$ .

For any formula  $\alpha$  that is maximal-consistent in  $\mathcal{U}$ , there is a partial world  $w \in \mathcal{U}$  such that  $\delta(w) \equiv \alpha$ . Indeed, if  $\alpha$  is maximal-consistent in  $\mathcal{U}$ , then by the clause ii) there is a  $w \in \mathcal{U}$  such that  $w \models \alpha$ , iff (by the first clause)  $\delta(w) \vdash \alpha$ . So  $\alpha \wedge \delta(w) \equiv \delta(w) \not\|\sim_{\mathcal{U}, \mathbf{L}^{\parallel}} \perp$ , which by clause iii) entails that  $\alpha \vdash \delta(w)$ .

Let  $\mathcal{M} = (\mathcal{U}, <)$  be a  $\mathbf{L}^{\parallel}$ -smooth partial worlds model.  $\|\sim_{\mathcal{M}}$  satisfies the rules:

<b>Reflexivity</b>	$\alpha \ \sim_{\mathcal{M}} \alpha$
<b>(<math>\mathbf{L}^{\parallel}</math>) <math>\mathcal{U}</math>-Left Equivalence</b>	if $\alpha \cong_{\mathcal{U}, \mathbf{L}^{\parallel}} \beta$ and $\alpha \ \sim_{\mathcal{M}} \gamma$ , then $\beta \ \sim_{\mathcal{M}} \gamma$
<b>(<math>\mathbf{L}^{\parallel}</math>) <math>\mathcal{U}</math>-Right Weakening</b>	if $\alpha \not\ \sim_{\mathcal{U}, \mathbf{L}^{\parallel}} \perp$ and $\gamma \ \sim_{\mathcal{M}} \alpha$ , then $\gamma \ \sim_{\mathcal{M}} \beta$
<b>Cut</b>	if $\alpha \wedge \beta \ \sim_{\mathcal{M}} \gamma$ and $\alpha \ \sim_{\mathcal{M}} \beta$ , then $\alpha \ \sim_{\mathcal{M}} \gamma$
<b>Cautious Monotony</b>	if $\alpha \ \sim_{\mathcal{M}} \beta$ and $\alpha \ \sim_{\mathcal{M}} \gamma$ , then $\alpha \wedge \beta \ \sim_{\mathcal{M}} \gamma$
<b>(<math>\mathbf{L}^{\parallel}</math>) <math>\mathcal{U}</math>-Consistence</b>	if $\alpha \ \sim_{\mathcal{M}} \perp$ , then $\alpha \not\ \sim_{\mathcal{U}, \mathbf{L}^{\parallel}} \perp$
<b><math>\parallel</math>-Or</b>	if $\alpha \ \sim_{\mathcal{M}} \gamma$ and $\beta \ \sim_{\mathcal{M}} \gamma$ , then $\alpha \parallel \beta \ \sim_{\mathcal{M}} \gamma$
<b>Injectivity</b>	for any $\alpha$ maximal-consistent in $\mathcal{U}$ , if $\alpha \parallel \beta \parallel \gamma \ \sim_{\mathcal{M}} \beta \parallel \gamma$ and $\alpha \parallel \beta \not\ \sim_{\mathcal{M}} \beta$ , then $\alpha \parallel \gamma \ \sim_{\mathcal{M}} \gamma$

The proof for *Reflexivity* is trivial. Those for  *$\mathcal{U}$ -Left Equivalence* and  *$\mathcal{U}$ -Right Weakening* are similar to those given in section 3.3 above for  $\vdash_{\mathcal{M}}$  (one just needs to replace  $\mathcal{L}$

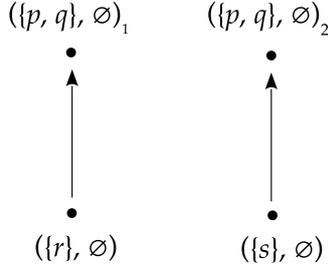


Figure 3.  $\mathcal{S} = (\mathcal{V}, <)$

with  $\mathbf{L}^\parallel$ , and generally  $\mathcal{L}$ -notions by the corresponding  $\mathbf{L}^\parallel$ -notions). Those for *Cut* and *Cautious Monotony* are similar to those given in (Kraus et al., 1990)<sup>1</sup>. The proof for *U-Consistence* is immediate using  $\mathbf{L}^\parallel$ -smoothness and the definitions of  $\|\sim_{\mathcal{M}}$  and  $\|\sim_{\mathcal{U}, \mathcal{L}^\parallel}$ . For *U-Or*, let  $w$  be  $<$ -minimal for  $\alpha \parallel \beta$ : either  $w \models \alpha$ , or  $w \models \beta$ . If  $w \models \alpha$ , then it is  $<$ -minimal for  $\alpha$  (for suppose not, then there is  $w' < w$  such that  $w' \models \alpha$ , so  $w$  is not  $<$ -minimal for  $\alpha \parallel \beta$ : contradiction). Since by hypothesis  $\alpha \|\sim_{\mathcal{M}} \gamma$ ,  $w \models \gamma$ . Similarly, if  $w \models \beta$ , then  $w \models \gamma$ . The proof for *Injectivity* is given in Appendix B.

*Injectivity* is thus named because it is not valid in structures  $\mathcal{S} = (\mathcal{V}, <)$  such that  $\mathcal{V}$  contains several ‘copies’ of a same partial world, which corresponds in KLM’s framework to the fact that the labelling function is not injective. For example, let  $\mathcal{V} = \{(\{p, q\}, \emptyset)_1, (\{p, q\}, \emptyset)_2, (\{r\}, \emptyset), (\{s\}, \emptyset)\}$ . Consider the structure  $\mathcal{S} = (\mathcal{V}, <)$  in Figure 3, and define  $\|\sim_{\mathcal{V}, \mathcal{L}}$ ,  $\|\sim_{\mathcal{S}}$ , ... etc. as expected.  $p \wedge q$  is maximal-consistent in  $\mathcal{V}$ , yet we have  $(p \wedge q) \parallel r \parallel s \|\sim_{\mathcal{S}} r \parallel s$  and  $(p \wedge q) \parallel r \not\|\sim_{\mathcal{S}} r$ , but  $(p \wedge q) \parallel s \|\sim_{\mathcal{S}} s$ .

The meaning of *Injectivity* is best seen considering the equivalent writing of the rule:

For any formula  $\alpha$  maximal-consistent in  $\mathcal{U}$ ,  
 if  $\alpha \parallel \beta \not\|\sim_{\mathcal{M}} \beta$  and  $\alpha \parallel \gamma \not\|\sim_{\mathcal{M}} \gamma$ , then  $\alpha \parallel \beta \parallel \gamma \not\|\sim_{\mathcal{M}} \beta \parallel \gamma$

Since  $\alpha$  is maximal-consistent in  $\mathcal{U}$ , there is some partial world  $w$  in  $\mathcal{U}$  such that  $\alpha \equiv \delta(w)$ . According to  $\mathcal{M}$ ’s intended interpretation,  $w$  stands for some mental representation  $a$  in the agent’s mind, so  $w$  can be denoted  $w_a$ .  $\delta(w_a) = \delta(w)$  stands for the conjunction of the features that, in the agent’s opinion, are satisfied by the objects/situations of which  $a$  is the mental representation. Similarly, let  $b_1, \dots, b_n$  be the agent’s mental representations such that  $\mathcal{W}_u(\beta) = \{w_{b1}, \dots, w_{bn}\}$ , and  $c_1, \dots, c_m$  its mental representations such that  $\mathcal{W}_u(\gamma) = \{w_{c1}, \dots, w_{cm}\}$ . If  $\beta \not\cong_{\mathcal{U}, \mathcal{L}^\parallel} \perp$  and  $\gamma \not\cong_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ , then  $\beta \cong_{\mathcal{U}, \mathcal{L}^\parallel} \|\delta(w_{bi}) / (w_{bi} \in \mathcal{W}_u(\beta))$  and  $\gamma \cong_{\mathcal{U}, \mathcal{L}^\parallel} \|\delta(w_{cj}) / (w_{cj} \in \mathcal{W}_u(\gamma))$ , so using substitution of  $\mathcal{U}$ -equivalent  $\parallel$ -disjuncts (see p. 19), *U-Left Equivalence* and *U-Right Weakening*, one gets that

if  $\delta(w_a) \parallel \delta(w_{b1}) \parallel \dots \parallel \delta(w_{bn}) \not\|\sim_{\mathcal{M}} \delta(w_{b1}) \parallel \dots \parallel \delta(w_{bn})$   
 and  $\delta(w_a) \parallel \delta(w_{c1}) \parallel \dots \parallel \delta(w_{cm}) \not\|\sim_{\mathcal{M}} \delta(w_{c1}) \parallel \dots \parallel \delta(w_{cm})$ ,  
 then  $\delta(w_a) \parallel \delta(w_{b1}) \parallel \dots \parallel \delta(w_{bn}) \parallel \delta(w_{c1}) \parallel \dots \parallel \delta(w_{cm}) \not\|\sim_{\mathcal{M}} \delta(w_{b1}) \parallel \dots$   
 $\dots \parallel \delta(w_{bn}) \parallel \delta(w_{c1}) \parallel \dots \parallel \delta(w_{cm})$

which can be interpreted as follows: if, when the information that the agent considers is compatible with its mental representations  $a, b_1, \dots, b_n$ , it does not overlook  $a$  as a possibility, and if when the information that it considers is compatible with its mental representations  $a, c_1, \dots, c_m$  it does not overlook  $a$  as a possibility, then, when the information it considers is compatible with  $a, b_1, \dots, b_n, c_1, \dots, c_m$ , it does not overlook  $a$  as a possibility either. If  $\beta \cong_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ , then  $\{w_{b1}, \dots, w_{bn}\} = \emptyset$  and by substitution of

<sup>1</sup>(Kraus et al., 1990), p. 18.

$\mathcal{U}$ -equivalent  $\parallel$ -disjuncts,  $\mathcal{U}$ -Left Equivalence and  $\mathcal{U}$ -Right Weakening, the meaning of *Injectivity* gets trivial. Similarly if  $\gamma \cong_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ .

We shall call PW (for ‘partial worlds’) the above set of rules. Here are some additional rules that can be usefully derived from PW:

**( $\mathbf{L}^\parallel$ )Supra- $\mathcal{U}$ -Consequence:**

If  $\alpha \parallel_{\mathcal{U}, \mathcal{L}^\parallel} \beta$ , then  $\alpha \parallel_{\mathcal{M}} \beta$   
(from *Reflexivity* and  *$\mathcal{U}$ -Right Weakening*).

**Equivalence:**

If  $\alpha \parallel_{\mathcal{M}} \beta$ ,  $\beta \parallel_{\mathcal{M}} \alpha$  and  $\alpha \parallel_{\mathcal{M}} \gamma$ , then  $\beta \parallel_{\mathcal{M}} \gamma$   
(from *Cautious Monotony*,  *$\mathcal{U}$ -Left Equivalence* and *Cut*).

**And:**

If  $\alpha \parallel_{\mathcal{M}} \beta$  and  $\alpha \parallel_{\mathcal{M}} \gamma$ , then  $\alpha \parallel_{\mathcal{M}} \beta \wedge \gamma$   
(from *Reflexivity*,  *$\mathcal{U}$ -Right Weakening*, *Cautious Monotony* and *Cut*).

**$\parallel$ -Disjunct Equivalence:**

If  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ ,  $\alpha_2 \parallel_{\mathcal{M}} \alpha_1$  and  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \gamma$ , then  $\alpha_2 \parallel \beta \parallel_{\mathcal{M}} \gamma$   
(from *Reflexivity*,  *$\mathcal{U}$ -Right Weakening*, *Cautious Monotony*,  *$\mathcal{U}$ -Left Equivalence* and *Equivalence*).

**$\parallel$ -Transitivity:**

If  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ , ...,  $\alpha_{n-1} \parallel_{\mathcal{M}} \alpha_n$ , then  $\alpha_1 \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_n$   
(from *Reflexivity*,  *$\parallel$ -Or*,  *$\mathcal{U}$ -Right Weakening*, *Cautious Monotony*,  *$\mathcal{U}$ -Left Equivalence* and  *$\parallel$ -Disjunct Equivalence*).

The proof for *Supra- $\mathcal{U}$ -Consequence* is immediate. Those for *Equivalence* and *And* are similar to those given in (Kraus et al., 1990) p.14. Those for  *$\parallel$ -Disjunct Equivalence* and  *$\parallel$ -Transitivity* are given in Appendices C and D.

#### 5.4 A few additional definitions

The formula  $C_{\parallel}(\alpha)$

Let  $\mathbf{L}^\parallel / \cong_{\mathcal{U}, \mathcal{L}^\parallel}$  be the set of all equivalence classes of  $\mathbf{L}^\parallel$  under  $\cong_{\mathcal{U}, \mathcal{L}^\parallel}$ , and  $g$  an arbitrary choice function on  $\mathbf{L}^\parallel / \cong_{\mathcal{U}, \mathcal{L}^\parallel}$  that selects a representative for each equivalence class. Since  $\mathbf{L}^\parallel$  is finite, so is  $\mathbf{L}^\parallel / \cong_{\mathcal{U}, \mathcal{L}^\parallel}$ . For any  $\mathbf{L}^\parallel$ -formula  $\alpha$ , we denote  $C_{\parallel}(\alpha)$  the conjunction of all the  $\mathbf{L}^\parallel$ -formulas  $\beta$  such that:  $\alpha \parallel \beta$  and  $\beta$  is the representative of its equivalence class under  $\cong_{\mathcal{U}, \mathcal{L}^\parallel}$ . By a series of applications of the derived rule *And*, one readily shows that for any  $\mathbf{L}^\parallel$ -formula  $\alpha$ ,  $\alpha \parallel C_{\parallel}(\alpha)$ . Moreover, for any  $\mathbf{L}^\parallel$ -formulas  $\alpha$  and  $\beta$ ,  $\alpha \parallel \beta$  iff  $C_{\parallel}(\alpha) \parallel_{\mathcal{U}, \mathcal{L}^\parallel} \beta$ : indeed, if  $\alpha \parallel \beta$ , then  $\beta$  is  $\mathcal{U}$ -equivalent to some conjunct in  $C_{\parallel}(\alpha)$ , so any partial world in  $\mathcal{U}$  that satisfies  $C_{\parallel}(\alpha)$  also satisfies  $\beta$ , i.e.  $C_{\parallel}(\alpha) \parallel_{\mathcal{U}, \mathcal{L}^\parallel} \beta$ . Conversely, suppose that  $C_{\parallel}(\alpha) \parallel_{\mathcal{U}, \mathcal{L}^\parallel} \beta$ . Since  $\alpha \parallel C_{\parallel}(\alpha)$ , by  *$\mathcal{U}$ -Right Weakening* one gets  $\alpha \parallel \beta$ .

#### Precisifications and precisification-free partial worlds models

Given  $w \in \mathcal{W}_L^p$ , a *precisification* of  $w$  is a partial world  $w' \in \mathcal{W}_L^p$  such that  $w \subset w'$  (where  $w$  and  $w'$  are the truth assignments which generate respectively  $w$  and  $w'$ , and  $\subset$  is the strict inclusion). In other words,  $w'$  is a precisification of  $w$  iff  $\delta(w') \vdash \delta(w)$  and  $\delta(w) \not\vdash \delta(w')$ , or equivalently, iff  $\delta(w') \parallel_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w)$  and  $\delta(w) \not\parallel_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w')$  (since  $\delta(w)$  and  $\delta(w')$  are  $\mathbf{L}$ -formulas, see p. 19 above).

We shall say that a set of partial worlds  $\mathcal{U}$  is *precisification-free* if and only if for any  $w$  and  $w'$  in  $\mathcal{U}$ ,  $w'$  is *not* a precisification of  $w$ .

If  $\mathcal{U}$  is precisification-free, then it is immediate that for any  $w$  in  $\mathcal{U}$ ,  $\delta(w)$  is maximal-consistent in  $\mathcal{U}$ . As (*modulo* classical equivalence)  $\{\alpha/\alpha \text{ is maximal-consistent in } \mathcal{U}\} \subseteq$

$\{\delta(w)/w \in \mathcal{U}\}$  (cf. p. 20), it follows that, if  $\mathcal{U}$  is precisification-free, then (*modulo* classical equivalence)  $\{\alpha/\alpha \text{ is maximal-consistent in } \mathcal{U}\} = \{\delta(w)/w \in \mathcal{U}\}$ .

Moreover, if  $\mathcal{U}$  is precisification-free then  $D_{\mathcal{U}, \mathbf{L}} = \mathcal{P}(\mathcal{U})$ . Indeed, suppose that there is a set  $A = \{w_1, \dots, w_n\} \subseteq \mathcal{U}$  which is not  $\mathbf{L}$ -definable. Then, for any  $\mathbf{L}$ -formula  $\alpha$  such that any  $w_i$  in  $A$  satisfies  $\alpha$ , there is a  $w'$  in  $\mathcal{U}$  such that  $w' \notin A$  and  $w' \models \alpha$ . In particular, there is a  $w'$  such that  $w' \models \|\delta(w_i)/(w_i \in A)\|$  and  $w' \notin A$ . This implies that there is a  $w_i \in A$  such that  $w' \models \delta(w_i)$ , that is, such that  $\delta(w') \vdash \delta(w_i)$ . Since  $w' \notin A$ ,  $w' \neq w_i$ , thus  $w'$  is a precisification of  $w_i$ .

We shall say that a partial worlds model  $\mathcal{M} = (\mathcal{U}, <)$  is *precisification-free* if and only if  $\mathcal{U}$  is precisification-free.

## 6. Two representation theorems for finite precisification-free partial worlds models

We shall now give two representation theorems. The first one states that PW is sound and complete for inference relations induced on  $\mathbf{L}$  by finite  $\mathbf{L}$ -smooth precisification-free models. For the second one, we first introduce a notion of *ranked* partial worlds models, inspired from Lehmann and Magidor's *ranked models*<sup>1</sup>, and an additional rule *Rankedness*. Then we show that  $\text{PW} \cup \{\text{Rankedness}\}$  is sound and complete for inference relations induced on  $\mathbf{L}$  by finite ranked precisification-free partial worlds models<sup>2</sup>.

### 6.1 Representation theorem for finite $\mathbf{L}$ -smooth precisification-free partial worlds models

In section 5.3, we showed that PW is sound for inference relations induced on  $\mathbf{L}$  by finite  $\mathbf{L}$ -smooth partial worlds models. *A fortiori* it is sound for those induced by precisification-free finite  $\mathbf{L}$ -smooth models. We shall see now that PW is also complete for the latter. For this we shall show that any relation  $\|\sim$  satisfying PW and such that  $\|\sim_{\mathcal{U}, \mathbf{L}}$  is induced by a precisification-free subset  $\mathcal{U}$  of  $\mathcal{W}_{\mathbf{L}}^p$  admits a precisification-free finite  $\mathbf{L}$ -smooth partial worlds model.

Let  $\mathbf{L}, \mathbf{L}^\parallel$  and  $\mathcal{W}_{\mathbf{L}}^p$  be as above, and let  $\mathcal{U}$  be a precisification-free subset of  $\mathcal{W}_{\mathbf{L}}^p$ ,  $\|\sim_{\mathcal{U}, \mathbf{L}}$  the corresponding  $\mathcal{U}$ -consequence relation on  $\mathbf{L}^\parallel$ , and  $\|\sim$  a binary relation on  $\mathbf{L}^\parallel$  satisfying PW. Since  $\mathbf{L}$  is finite,  $\mathcal{U}$  is a finite set of finite partial worlds. We define a relation  $<$  on  $\mathcal{U}$  by: for any  $w$  and  $w'$  in  $\mathcal{U}$ ,  $w < w'$  iff  $w \neq w'$  and  $\delta(w') \|\sim \delta(w)$ . It is immediate that  $\mathcal{M} = (\mathcal{U}, <)$  is a finite precisification-free partial worlds model. We show that  $<$  is  $\mathbf{L}$ -smooth:

**Lemma 1.** *For any  $w \in \mathcal{U}$ ,  $w \models C_{\|\sim}(\delta(w))$*

*Proof:*

Let  $w \in \mathcal{U}$ :

<sup>1</sup> (Lehmann & Magidor, 1992).

<sup>2</sup> It has been remarked that both these results would hold just as well if  $\mathcal{U}$  was more generally defined as a subset of  $\mathcal{P}(\mathcal{W}_{\mathbf{L}})$  — we would then have to define a precisification of an element  $s \in \mathcal{U}$  as a strict subset of  $s$ ,  $\delta(s)$  as  $\bigvee \delta(w)/w \in s$ , and the other notions as expected. Indeed, PW and  $\text{PW} \cup \{\text{Rankedness}\}$  would respectively be sound and complete for inference relations induced on  $\mathbf{L}^\parallel$  by the resulting finite precisification-free  $\mathbf{L}$ -smooth and ranked models (what would change are the properties of  $\|\sim_{\mathcal{U}, \mathbf{L}}$ ). Yet such models would not be usable for our purposes, since elements in  $\mathcal{U}$  are meant to stand for mental representations of objects and situations, which should be represented by partial worlds.

- i)  $\delta(w) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ , so by  $\mathcal{U}$ -Consistence,  
 $\delta(w) \Vdash \perp$ , iff  
 $C_{\Vdash}(\delta(w)) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ , iff  
 $\exists w' \in \mathcal{U}$  such that  $w' \models C_{\Vdash}(\delta(w))$ .
- ii) By *Reflexivity*,  $\delta(w) \Vdash \delta(w)$ , iff  
 $C_{\Vdash}(\delta(w)) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w)$ , thus by i),  
 $w' \models \delta(w)$ .  
Thus  $w' \models \delta(w) \wedge \delta(w')$ , thus  
 $\delta(w) \wedge \delta(w') \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \perp$ .
- iii) By hypothesis  $\mathcal{U}$  is precisification-free, so  $\delta(w)$  and  $\delta(w')$  both are maximal-consistent. By definition of maximal-consistency, ii) implies that  $\delta(w) \equiv \delta(w')$ , iff  $w = w'$ . By i), this entails that  $w \models C_{\Vdash}(\delta(w))$ . □

**Lemma 2.**  $<$  is  $\mathbf{L}^\parallel$ -smooth.

*Proof:*

- i) By its definition,  $<$  is irreflexive.
- ii)  $<$  is transitive: indeed, suppose that there are  $w_1, w_2$  and  $w_3$  in  $\mathcal{U}$  such that  $w_3 < w_2 < w_1$ . Then:
  1. By definition of  $<$ ,  $\delta(w_1) \Vdash \delta(w_2) \Vdash \delta(w_2)$ , and  $\delta(w_2) \Vdash \delta(w_3) \Vdash \delta(w_3)$ .
  2. By *Reflexivity*,  $\delta(w_1) \Vdash \delta(w_1)$ , so by  $\mathcal{U}$ -Right Weakening  
 $\delta(w_1) \Vdash \delta(w_1) \Vdash \delta(w_2)$ .  
Similarly,  $\delta(w_2) \Vdash \delta(w_2) \Vdash \delta(w_3)$ .
  3. By  $\Vdash$ -Transitivity on 1 and 2, we get  $\delta(w_1) \Vdash \delta(w_3) \Vdash \delta(w_3)$ .
  4. Furthermore,  $w_1 \neq w_3$ . Indeed suppose that  $w_1 = w_3$ :
    - a) Then  $\delta(w_1) = \delta(w_3)$ , so by 1,  $\delta(w_3) \Vdash \delta(w_2) \Vdash \delta(w_2)$ .  
So by 2 + *Equivalence*,  $C_{\Vdash}(\delta(w_2)) = C_{\Vdash}(\delta(w_2) \Vdash \delta(w_3))$ .
    - b) By *Reflexivity*,  $\delta(w_3) \Vdash \delta(w_3)$ , so by  $\mathcal{U}$ -Right Weakening,  
 $\delta(w_3) \Vdash \delta(w_2) \Vdash \delta(w_3)$ .  
By 1 + *Equivalence*,  $C_{\Vdash}(\delta(w_3)) = C_{\Vdash}(\delta(w_2) \Vdash \delta(w_3))$ .  
So by a),  $C_{\Vdash}(\delta(w_2)) = C_{\Vdash}(\delta(w_3))$ .
    - c) By *Reflexivity*,  $C_{\Vdash}(\delta(w_2)) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w_2)$  and  $C_{\Vdash}(\delta(w_3)) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w_3)$ .
    - d) By Lemma 1,  $\delta(w_2) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} C_{\Vdash}(\delta(w_2))$ , so by b) + c) + transitivity of  
 $\Vdash_{\mathcal{U}, \mathcal{L}^\parallel}$ ,  $\delta(w_2) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w_3)$ , iff (since  $\delta(w_3)$  is a  $\mathbf{L}$ -formula)  $\delta(w_2) \vdash \delta(w_3)$ .
    - e) Similarly,  $\delta(w_3) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} C_{\Vdash}(\delta(w_3)) = C_{\Vdash}(\delta(w_2) \Vdash \delta(w_2)) \Vdash_{\mathcal{U}, \mathcal{L}^\parallel} \delta(w_2)$ , so  
 $\delta(w_3) \vdash \delta(w_2)$ .  
By d),  $\delta(w_2) \equiv \delta(w_3)$ , iff  $w_2 = w_3$ .
    - f) By hyp.  $w_3 < w_2$ , thus by def. of  $<$ ,  $w_2 \neq w_3$ , which contradicts e).
  5. By 3 and 4,  $w_3 < w_1$ .
- iii) By i) and ii),  $<$  is acyclic.
- iv) Since  $\mathcal{U}$  is finite, it follows from iii) that for any subset  $A \subseteq \mathcal{U}$ , there is at least one  $w \in A$  such that  $w$  is  $<$ -minimal in  $A$ . Since  $<$  is transitive, for any  $w'$  in  $A$  that is not  $<$ -minimal in  $A$ , there is a  $w \in A$  such that  $w$  is  $<$ -minimal in  $A$  and  $w < w'$ . This applies in particular to subsets  $A \subseteq \mathcal{U}$  such that  $A$  is  $\mathbf{L}^\parallel$ -definable. □

It follows from Lemma 2 that  $\mathcal{M}$  is a finite  $\mathbf{L}^\parallel$ -smooth precisification-free partial worlds model. We now show that  $\mathcal{M}$  is a model of  $\Vdash$ :

Let  $\Vdash_{\mathcal{M}}$  be the inference relation induced by  $\mathcal{M}$  on  $\mathbf{L}^\parallel$ . We shall show that  $\Vdash_{\mathcal{M}} = \Vdash$ . As seen in section 5.3,  $\Vdash_{\mathcal{M}}$  satisfies PW and its derived rules. Let  $\alpha$  be a  $\mathbf{L}^\parallel$ -formula.

**Lemma 3.** *If  $\mathcal{W}_u(\alpha) = \emptyset$ , then, for any  $\mathbf{L}^\parallel$ -formula  $\beta$ , we have both  $\alpha \Vdash_{\mathcal{M}} \beta$  and  $\alpha \Vdash \beta$ .*

*Proof:*

Immediate, using the fact that  $\mathcal{W}_u(\alpha) = \emptyset$  iff  $\alpha \Vdash_{\mathcal{U}, \mathcal{L}} \perp$ , *Supra- $\mathcal{U}$ -Consequence* and  *$\mathcal{U}$ -Right Weakening*.  $\square$

For the remainder of the proof, we shall assume that  $\mathcal{W}_u(\alpha) \neq \emptyset$ . Let  $X = \{w \in \mathcal{W}_u(\alpha) / \forall w' \in \mathcal{W}_u(\alpha) \text{ s.t. } w' \neq w, \delta(w) \parallel \delta(w') \not\sim \delta(w')\}$ , and  $Y = \mathcal{W}_u(\alpha) - X$ . By the definition of  $<$ ,  $X$  is the set of  $<$ -minimal elements in  $\mathcal{W}_u(\alpha)$ . By Lemma 2,  $X \neq \emptyset$ . Say  $X = \{w_1, \dots, w_n\}$ .

For each  $w_i \in X$ , let  $y_i = \{w \in Y / \delta(w) \parallel \delta(w_i) \sim \delta(w_i)\}$ . By the definition of  $<$ ,  $y_i$  is the set of elements of  $\mathcal{W}_u(\alpha)$  that are minimized by  $w_i$  ( $y_i$  may be empty). By Lemma 2,  $\bigcup y_i / (w_i \in X) = Y$ .

**Lemma 4.**  $C_{\perp}(\alpha) \Vdash_{\mathcal{U}, \mathcal{L}} \parallel \delta(w_k) / (w_k \in X)$

*Proof:*

- i) For any  $w_i \in X$ , using successive applications of  *$\parallel$ -Or +  $\mathcal{U}$ -Left Equivalence*, we get  $(\parallel \delta(w_k) / (w_k \in y_i)) \parallel \delta(w_i) \sim \delta(w_i)$ .  
Thus by  *$\mathcal{U}$ -Right Weakening*, for any  $w_i \in X$ , we get:  
 $(\parallel \delta(w_k) / (w_k \in y_i)) \parallel \delta(w_i) \sim \parallel \delta(w_k) / (w_k \in X)$ .
- ii) By a series of applications of  *$\parallel$ -Or +  $\mathcal{U}$ -Left Equivalence* on i), we get  $(\parallel \delta(w_k) / (w_k \in Y)) \parallel (\parallel \delta(w_k) / (w_k \in X)) \sim \parallel \delta(w_k) / (w_k \in X)$ , that is,  
 $\parallel \delta(w_k) / (w_k \in \mathcal{W}_u(\alpha)) \sim \parallel \delta(w_k) / (w_k \in X)$ .
- iii) By hypothesis  $\mathcal{W}_u(\alpha) \neq \emptyset$ , thus  $\parallel \delta(w_k) / (w_k \in \mathcal{W}_u(\alpha)) \cong_{\mathcal{U}, \mathcal{L}} \alpha$ .  
Thus by  *$\mathcal{U}$ -Left Equivalence* on ii),  $\alpha \sim \parallel \delta(w_k) / (w_k \in X)$ , iff  
 $C_{\perp}(\alpha) \Vdash_{\mathcal{U}, \mathcal{L}} \parallel \delta(w_k) / (w_k \in X)$ .  $\square$

**Lemma 5.**  $\parallel \delta(w_k) / (w_k \in X) \Vdash_{\mathcal{U}, \mathcal{L}} C_{\perp}(\alpha)$

*Proof:*

Suppose the contrary, *i.e.* that there is a  $w_k \in X$  such that  $\delta(w_k) \not\Vdash_{\mathcal{U}, \mathcal{L}} C_{\perp}(\alpha)$ . For simplicity's sake, say  $w_k$  is  $w_1$ .

- i) By the derived rule *And*,  $\alpha \sim C_{\perp}(\alpha)$ .
- ii) For any  $w_k \in X$ ,  $w_k \Vdash \alpha$ , iff  $\delta(w_k) \Vdash_{\mathcal{U}, \mathcal{L}} \alpha$ .  
Thus  $\parallel \delta(w_k) / (w_k \in X) \Vdash_{\mathcal{U}, \mathcal{L}} \alpha$ , so by *Supra- $\mathcal{U}$ -Consequence*,  
 $\parallel \delta(w_k) / (w_k \in X) \sim \alpha$ .
- iii) In Lemma 4's proof (item iii) we showed that  $\alpha \sim \parallel \delta(w_k) / (w_k \in X)$ . Thus by the derived rule *Equivalence* on i) and ii),  $\parallel \delta(w_k) / (w_k \in X) \sim C_{\perp}(\alpha)$ .
- iv) By *Reflexivity*,  $\parallel \delta(w_k) / (w_k \in X) \sim \parallel \delta(w_k) / (w_k \in X)$ , thus by *And* on iii),  
 $\parallel \delta(w_k) / (w_k \in X) \sim (\parallel \delta(w_k) / (w_k \in X)) \wedge C_{\perp}(\alpha)$ .
- v)  $(\parallel \delta(w_k) / (w_k \in X)) \wedge C_{\perp}(\alpha) \cong_{\mathcal{U}, \mathcal{L}}$   
 $(\delta(w_1) \wedge C_{\perp}(\alpha)) \parallel (\delta(w_2) \wedge C_{\perp}(\alpha)) \parallel \dots \parallel (\delta(w_n) \wedge C_{\perp}(\alpha))$ .
- vi) By hypothesis,  $\delta(w_1) \not\Vdash_{\mathcal{U}, \mathcal{L}} C_{\perp}(\alpha)$ , iff  $w_1 \not\Vdash C_{\perp}(\alpha)$ .  
By hypothesis  $\mathcal{U}$  is precisification-free, which implies that there is no  $w$  in  $\mathcal{U}$  such that  $w \Vdash \delta(w_1) \wedge C_{\perp}(\alpha)$ .  
Thus  $\delta(w_1) \wedge C_{\perp}(\alpha) \cong_{\mathcal{U}, \mathcal{L}} \perp$ .
- vii) Necessarily, there is a  $w_2 \in X$  such that  $w_2 \neq w_1$ : for suppose not, then  $\parallel \delta(w_k) / (w_k \in X) = \delta(w_1)$ , thus by  *$\mathcal{U}$ -Right Weakening* on iv) and vi),  $\delta(w_1) \sim \perp$ , so by  *$\mathcal{U}$ -Consistence*,  $\delta(w_1) \Vdash_{\mathcal{U}, \mathcal{L}} \perp$ , iff  $w_1 \notin \mathcal{U}$ , which is absurd.
- viii) By v), vi) and vii),  $(\parallel \delta(w_k) / (w_k \in X)) \wedge C_{\perp}(\alpha) \cong_{\mathcal{U}, \mathcal{L}} \perp$ .

- $(\delta(w_2) \wedge C_{\perp}(\alpha)) \parallel \dots \parallel (\delta(w_n) \wedge C_{\perp}(\alpha)) \cong_{\mathcal{U}, \mathcal{L}^1}$   
 $(\|\delta(w_k)/(w_k \in \{w_2, \dots, w_n\})\| \wedge C_{\perp}(\alpha)).$
- ix)** By  *$\mathcal{U}$ -Right Weakening* on iv) and viii),  
 $\|\delta(w_k)/(w_k \in X)\| \parallel (\|\delta(w_k)/(w_k \in \{w_2, \dots, w_n\})\| \wedge C_{\perp}(\alpha)),$   
 thus by  *$\mathcal{U}$ -Right Weakening*,  
 $\|\delta(w_k)/(w_k \in X)\| \parallel \|\delta(w_k)/(w_k \in \{w_2, \dots, w_n\})\|.$
- x)**  $w_1 \in X$ , so by definition of  $X$ ,  $\delta(w_1) \parallel \delta(w_2) \not\parallel \delta(w_2)$ .  
 So by ix), there is a  $w_3 \in X$  such that  $w_1 \neq w_3 \neq w_2$ .
- xi)** By hypothesis  $\mathcal{U}$  is precisification-free, so for any  $w_i \in X$ ,  $\delta(w_i)$  is maximal-consistent.  
 Thus, since  $\delta(w_1) \parallel \delta(w_2) \not\parallel \delta(w_2)$ , by *Injectivity* on ix) and x),  
 $\delta(w_1) \parallel (\|\delta(w_k)/(w_k \in \{w_3, \dots, w_n\})\| \parallel \|\delta(w_k)/(w_k \in \{w_3, \dots, w_n\})\|).$   
 Since for any  $i$  ( $3 \leq i \leq n$ ),  $\delta(w_1) \parallel \delta(w_i) \not\parallel \delta(w_i)$ , successive applications of the rule *Injectivity* lead to a contradiction.  
 Thus  $\delta(w_1) \parallel_{\overline{\mathcal{U}, \mathcal{L}}} C_{\perp}(\alpha)$ .

A similar reasoning can be made for any  $w_k \in X$  (one just has to change the indexation).  
 So, for any  $w_k$  in  $X$ ,  $\delta(w_k) \parallel_{\overline{\mathcal{U}, \mathcal{L}}} C_{\perp}(\alpha)$ , iff  
 $\|\delta(w_k)/(w_k \in X)\| \parallel_{\overline{\mathcal{U}, \mathcal{L}}} C_{\perp}(\alpha)$ . □

**Lemma 6.** *If  $\mathcal{W}_{\mathcal{U}}(\alpha) \neq \emptyset$ , then, for any  $\mathcal{L}^{\parallel}$ -formula  $\beta$ , we have  $\alpha \parallel \beta$  iff  $\alpha \parallel_{\mathcal{M}} \beta$ .*

*Proof:*

Assume  $\mathcal{W}_{\mathcal{U}}(\alpha) \neq \emptyset$ . By Lemmas 4 and 5,  $C_{\perp}(\alpha) \cong_{\mathcal{U}, \mathcal{L}^1} \|\delta(w_k)/(w_k \in X)\|$ . For any  $\mathcal{L}^{\parallel}$ -formula  $\beta$ , we have:

$$\begin{aligned}
 & \alpha \parallel \beta \text{ iff} \\
 & C_{\perp}(\alpha) \parallel_{\overline{\mathcal{U}, \mathcal{L}^{\parallel}}} \beta, \text{ iff} \\
 & \|\delta(w_k)/(w_k \in X)\| \parallel_{\overline{\mathcal{U}, \mathcal{L}^{\parallel}}} \beta, \text{ iff} \\
 & \text{for any } w_k \text{ such that } w_k \in X, w_k \parallel_{\overline{\mathcal{U}, \mathcal{L}^{\parallel}}} \beta, \text{ iff} \\
 & \text{for any } w_k \text{ such that } w_k \text{ is } < \text{-minimal for } \alpha, w_k \models \beta, \text{ iff} \\
 & \alpha \parallel_{\mathcal{M}} \beta
 \end{aligned}$$
□

**Lemma 7.**  *$\mathcal{M} = (\mathcal{U}, <)$  is a finite  $\mathcal{L}^{\parallel}$ -smooth precisification-free partial worlds model such that  $\parallel_{\mathcal{M}} = \parallel$ .*

*Proof:*

Immediate, from Lemmas 2-6. □

**Theorem 8.** *If  $\parallel$  is a relation on  $\mathcal{L}^{\parallel}$  satisfying PW and such that  $\parallel_{\overline{\mathcal{U}, \mathcal{L}^{\parallel}}}$  is induced by a precisification-free subset  $\mathcal{U}$  of  $\mathcal{W}_{\mathcal{L}^{\parallel}}^{\mathcal{P}}$ , then  $\parallel$  admits a finite  $\mathcal{L}^{\parallel}$ -smooth precisification-free partial worlds model.*

*Proof:*

Immediate, from Lemma 6. □

**Corollary 9.** *PW is sound and complete for inference relations induced on  $\mathcal{L}^{\parallel}$  by finite  $\mathcal{L}^{\parallel}$ -smooth precisification-free partial worlds models.*

**Remark.**  *$\parallel$ -Transitivity is valid in any  $\mathcal{L}^{\parallel}$ -smooth model, whether  $<$  is transitive or not. However,  $\parallel$ -Transitivity forces the above construction to be transitive. Thus, for*

any finite  $\mathbf{L}^\parallel$ -smooth precisification-free model  $\mathcal{M}$ , there is a transitive (and irreflexive) finite precisification-free model  $\mathcal{M}'$ , such that  $\|\sim_{\mathcal{M}} = \|\sim_{\mathcal{M}'}$ . On the other hand, any strict ordered model is  $\mathbf{L}^\parallel$ -smooth, provided that  $\mathcal{U}$  is finite. Thus, although finite strict ordered precisification-free partial worlds models form a strict subclass of finite  $\mathbf{L}^\parallel$ -smooth ones, they give rise to the same set of induced inference relations (contrary to what is the case in the KLM framework, where smoothness and strict ordering generate two different sets of induced inference relations, even in the finite case).

## 6.2 Representation theorem for finite precisification-free ranked models

A binary relation  $<$  on a set  $E$  is a *modular order* iff  $<$  is a strict partial order that satisfies the property<sup>1</sup>

**Modularity** (*1st version*): for any  $x, y$  and  $z$  in  $E$ , if  $x < y$ , then  $z < y$  or  $x < z$ .<sup>2</sup>

If we write  $y \leq z$  to mean that  $z \not< y$ , then an equivalent formulation of *Modularity* is:

**Modularity** (*2nd version*): for any  $x, y$  and  $z$  in  $E$ , if  $x < y$  and  $y \leq z$ , then  $x < z$ .

In practice, *Modularity* orders  $E$ 's elements into ranks.

If  $\mathcal{M} = (\mathcal{U}, <)$  is a partial worlds model such that  $<$  is a modular order, we say that  $\mathcal{M}$  is a *ranked* partial worlds model.

Let  $\mathbf{L}$  be a finite propositional language,  $\mathcal{U} \subseteq \mathcal{W}_E^p$ , and  $\mathcal{M} = (\mathcal{U}, <)$  a ranked partial worlds model. Since  $\mathcal{U}$  is finite and  $<$  is a strict partial order,  $\mathcal{M}$  is a finite  $\mathbf{L}^\parallel$ -smooth partial worlds model. Thus  $\|\sim_{\mathcal{M}}$  satisfies the rules from PW. In addition,  $\|\sim_{\mathcal{M}}$  satisfies the following rule *Rankedness*:

**Rankedness** For any  $\mathbf{L}^\parallel$ -formulas maximal-consistent in  $\mathcal{U}$   $\alpha_1, \alpha_2$  and  $\alpha_3$  (all  $\neq$ ), if  $\alpha_1 \parallel \alpha_2 \|\sim_{\mathcal{M}} \alpha_2$  and  $\alpha_1 \parallel \alpha_3 \not\|\sim_{\mathcal{M}} \alpha_3$ , then  $\alpha_2 \parallel \alpha_3 \|\sim_{\mathcal{M}} \alpha_2$

(see Appendix E for proof). *A fortiori*, if  $\mathcal{M} = (\mathcal{U}, <)$  is a precisification-free finite ranked partial worlds model, then  $\|\sim_{\mathcal{M}}$  satisfies  $\text{PW} \cup \{\text{Rankedness}\}$ .

Conversely, if  $\mathbf{L}$  is a finite propositional language,  $\mathcal{U}$  a precisification-free subset of  $\mathcal{W}_E^p$ ,  $\|\sim_{\mathcal{U}, \mathbf{L}^\parallel}$  the  $\mathcal{U}$ -consequence relation induced on  $\mathbf{L}^\parallel$  by  $\mathcal{U}$  and  $\|\sim$  a relation on  $\mathbf{L}^\parallel$  satisfying  $\text{PW} \cup \{\text{Rankedness}\}$ , then using the fact that when  $\mathcal{U}$  is precisification-free (and *modulo* classical equivalence)  $\{\alpha \in \mathbf{L}^\parallel / \alpha \text{ is maximal-consistent in } \mathcal{U}\} = \{\delta(w) / w \in \mathcal{U}\}$ , one easily checks that the construction suggested in section 6.1 above gives rise to a finite ranked precisification-free model. So  $\text{PW} \cup \{\text{Rankedness}\}$  is sound and complete for inference relations induced on  $\mathbf{L}^\parallel$  by finite ranked precisification-free partial worlds models.

## 7. Modelling automatic inferences and learning

Finally, we shall give precisions on how partial worlds models can be used to model automatic inferences and learning. In a first part, we shall provide a formal version of the modelling of automatic inferences sketched in section 2.2 above. Then we shall briefly discuss the learning processes attached to automatic inferences, and the modelling of these processes.

<sup>1</sup>Definition taken from (Lehmann & Magidor, 1992, p. 19).

<sup>2</sup>(Lehmann & Magidor, 1992, p. 19).

## 7.1 Modelling automatic inferences

Let  $\mathcal{A}$  be the considered agent. We show how to choose a propositional language  $\mathbf{L}$  and build a partial worlds model  $\mathcal{M} = (\mathcal{U}, <)$  such that for any  $\mathbf{L}$ -formulas  $\alpha$  and  $\beta$ ,  $\alpha \Vdash_{\mathcal{M}} \beta$  if and only if  $\mathcal{A}$  is disposed to infer  $\beta$  from  $\alpha$ , in the sense defined on p. 8 above.

Let  $\mathcal{F}^+$  be the set of all the features that  $\mathcal{A}$  is physiologically able to perceive. We choose a propositional language  $\mathbf{L}$  such that there is a bijection  $\sigma: \mathcal{F}^+ \rightarrow \text{Var}(\mathbf{L})$ . Since  $\mathcal{F}^+$  is finite (see p. 6 above),  $\mathbf{L}$  is finite.

As remarked in section 2.2, agents are able to conceive the negation of the features they can perceive, and to reason over these negated features just as if they were (negative) features. Let  $\eta$  be the function which to each feature  $f$  in  $\mathcal{F}^+$  associates its negation *not- $f$* , and  $\mathcal{F}^-$  the image of  $\mathcal{F}^+$  by  $\eta$ .  $\mathcal{F} = \mathcal{F}^+ \cup \mathcal{F}^-$  is the set of all the features that  $\mathcal{A}$  is able to mentally handle.

Now we define a function  $\rho: \mathcal{F} \rightarrow \{\lambda \in \mathbf{L} / \lambda \text{ is a literal}\}$ :

If  $f \in \mathcal{F}^+$ , then  $\rho(f) = \sigma(f)$ , otherwise,  $\rho(f) = \neg\sigma(\eta^{-1}(f))$ .

It is immediate that  $\rho$  is a bijection. It defines a relation of representation between the literals from  $\mathbf{L}$  and the features from  $\mathcal{F}$ , in the sense that the literal  $\lambda$  shall stand for the feature  $f$  in the model if and only if  $\lambda = \rho(f)$ . One readily checks that for any  $p \in \text{Var}(\mathbf{L})$  and any  $f \in \mathcal{F}^+$ ,  $p = \rho(f)$  if and only if  $\neg p = \rho(\text{not-}f)$ .

Let  $\mathcal{R}$  be the set of  $\mathcal{A}$ 's most precise mental representations. As specified in section 2.2, we regard  $\mathcal{A}$ 's mental representations as sets of positive and negative features, which means that for any  $r$  in  $\mathcal{R}$ ,  $r \subseteq \mathcal{F}$ . In addition, since  $\mathcal{R}$  contains only  $\mathcal{A}$ 's most precise mental representations, it holds that for any  $r$  and  $r'$  in  $\mathcal{R}$ ,  $r \not\subseteq r'$  (where  $\subset$  denotes the strict inclusion). We also suppose that  $\mathcal{A}$ 's mental representations are consistent, *i.e.* that for any  $r$  in  $\mathcal{R}$  and any  $f$  in  $\mathcal{F}^+$ ,  $\{f, \text{not-}f\} \not\subseteq r$ . Obviously,  $\mathcal{R}$  is finite, since  $\mathcal{F}^+$ , hence also  $\mathcal{F}$ , are.

Let  $\rho': \mathcal{R} \rightarrow \mathcal{W}_{\mathbf{L}}^p$  be the function which to each  $r$  in  $\mathcal{R}$  associates the partial world  $w$  such that  $\delta(w) = \bigwedge \rho(f) / f \in r$ .  $\rho'$  defines a relation of representation between partial worlds for  $\mathbf{L}$  and  $\mathcal{A}$ 's most precise mental representations. Namely, a partial world  $w \in \mathcal{W}_{\mathbf{L}}^p$  shall stand for  $r$  in the model if and only if  $w = \rho'(r)$ .

We define  $\mathcal{U} = \{w \in \mathcal{W}_{\mathbf{L}}^p / \exists r \in \mathcal{R} \text{ s.t. } w = \rho'(r)\}$ . It is immediate that  $\mathcal{U}$  is finite and precisification-free.  $\mathcal{U}$  will represent  $\mathcal{R}$  in the model. Given a partial world  $w$ , the fact that  $w \in \mathcal{U}$  means that  $\mathcal{A}$  knows some class of objects/situations such that  $r = \rho'^{-1}(w)$  is the mental representation of this/these object(s)/situation(s) in its mind. Given a literal  $\lambda$  from  $\mathbf{L}$ , the fact that  $w \models \lambda$  means that according to this mental representation (*i.e.*, in  $\mathcal{A}$ 's view), the corresponding object(s)/situation(s) satisfy/ies the feature  $f = \rho^{-1}(\lambda)$ .

We define a binary relation  $<$  on  $\mathcal{U}$  by:

For any  $w$  and  $w'$  in  $\mathcal{U}$ ,  $w < w'$  iff  $\rho'^{-1}(w)$  is *more vivid* in  $\mathcal{A}$ 's mind than  $\rho'^{-1}(w')$ .

The exact properties of  $<$  depend on what 'being more vivid than' means when it comes to memories, which, according to our account, ultimately depends on the precise neural mechanisms that govern interactions between concept-neurons. But these are still poorly understood and all we can do for now is to make suppositions based on the intuition we have of our own retrieving of memories. In this regard, it seems that the fact that not all our memories appear equally vivid to us and that we feel this difference in vividness as somewhat quantitative (as is evidenced by our usual speaking of memories

as being ‘more’ or ‘less’ vivid relative one to another), should bring us to ask that  $<$  be at least irreflexive and transitive, and probably also modular. Therefore, it seems that  $\mathcal{M} = (\mathcal{U}, <)$  should be a precisification-free finite ranked partial worlds model.  $\mathcal{M}$  can be seen as representing the agent’s worldview, or, depending on the perspective we take, as representing its inferential system.

Using  $\mathbf{L}$  as a basis, we define  $\mathbf{L}^\parallel$  as specified in section 5.1, the meaning of  $\parallel$  being as discussed in this same section. The inference relation  $\parallel_{\mathcal{M}}$  induced by  $\mathcal{M}$  on  $\mathbf{L}^\parallel$  is to be interpreted as the set of  $\mathcal{A}$ ’s dispositions to infer, that is, as its general knowledge about things. More specifically, if  $\alpha$  and  $\beta$  are  $\mathbf{L}^\parallel$ -formulas, the fact that  $\alpha \parallel_{\mathcal{M}} \beta$  represents the fact that  $\mathcal{A}$  is disposed to infer  $\beta$  from  $\alpha$ . As shown in section 6,  $\parallel_{\mathcal{M}}$  is closed under the rules from  $\text{PW} \cup \{\text{Rankedness}\}$ . These rules are properties of the inference relation  $\parallel_{\mathcal{M}}$ , which means that they appear as a by-product of the posited inferential process. They should thus be interpreted as rules that arise from the functioning of a brain that runs this process. In this sense, the suggested modelling of automatic inferences gives an insight into how it is possible that intelligent reasoning and logic emerge in a brain at work.

## 7.2 Modelling of learning: an informal glimpse

Crucially, the suggested modelling allows to imagine a modelling of learning that would articulate with it in a natural manner. In general terms, learning can be defined as a revision of one’s knowledge, so as to take into account newly acquired information. In the context of automatic inferences, it boils down to a revision of the agent’s dispositions to draw automatic inferences, that is, to a revision of the inference relation  $\parallel_{\mathcal{M}}$ . In practice, this means that the information entering the agent’s inferential system triggers both the drawing of automatic inferences according to its current dispositions to infer, and a subsequent modification of these same dispositions. This revision can take different forms, depending on how incoming information incorporates with the agent’s current knowledge. On the face of it, one may distinguish three broad cases.

A first one is when the agent learns the existence of a given class of objects/situations. For example, suppose that ‘black’, ‘white’ and ‘swan’<sup>1</sup> are features the agent is able to perceive, and suppose that it knows about white swans, but ignores that there are also black ones. Following the analysis provided in section 2.2, this means that there is a ‘white swan’ item in the set of its mental representations, but no ‘black swan’ one. This is represented in the logical framework by the fact that there is a partial world  $w$  in  $\mathcal{U}$  that satisfies  $\rho(\text{‘white’}) \wedge \rho(\text{‘swan’})$ , but none that satisfies  $\rho(\text{‘black’}) \wedge \rho(\text{‘swan’})$ . Suppose that now the agent sees a black swan: that is, its perceptual system detects the co-occurring features ‘black’ and ‘swan’, and sends the corresponding information to its inferential system. Since  $\mathcal{A}$  has no ‘black swan’ mental representation, this information does not match any of the representations of the objects and situations that it knows, yielding its astonishment. On the logical side, this is rendered by the fact that  $\rho(\text{‘black’}) \wedge \rho(\text{‘swan’}) \parallel_{\mathcal{M}} \perp$ , where  $\perp$  is to be interpreted as an error signal (roughly, the agent concludes that ‘something is wrong’). But then it gets over its surprise, and admits that black swans do, in fact, exist. In other words, it revises its worldview by adding a ‘black swan’ element to the set of its mental representations, so as to make it consistent with the collected information. Astonishment is the start signal for this revision process.

---

<sup>1</sup>Again, for simplicity, we regard these as features (see footnote page 7).

A second case is when the agent supplements its knowledge of a given class of objects/situations with additional information. For example, suppose that  $\mathcal{A}$  knows white swans, but never had the opportunity to observe their legs, and thus has no idea whatsoever about what colour are their legs. Accordingly, its mental representation of white swans says nothing about the colour of their legs. This is represented in the framework by the fact that the partial world  $w \in \mathcal{U}$  that stands for  $\mathcal{A}$ 's mental representation of white swans neither satisfies nor falsifies any of the propositional variables that stand for legs' colour features. Put formally, the truth-assignment  $\mathbf{w}$  that corresponds to  $w$  is undefined for any propositional variable  $p$  such that  $p = \rho(f)$ , with  $f$  a legs' colour feature. But suppose that now the agent sees a white swan walking on the riverside, and observes that its legs are in fact black. Again, its inferential system returns an error signal, since none of  $\mathcal{A}$ 's mental representations satisfies at the same time the features 'white', 'swan' and 'black legs'; and again, this launches a learning process. But this time the agent supplements its knowledge about white swans by adding the feature 'black legs' to its previous mental representation of white swans.

Finally, a third case is when the vividness of a given representation is reinforced through repetition of experience. For example, suppose that the agent knows black and white swans, but that in its environment white swans outnumber black ones, so that it encounters white swans much more often than black ones. Presumably, when thinking about swans, it is white swans that should come to its mind in the first place, since a more often reactivated memory is generally more vivid than a less often reactivated one. As a consequence,  $\mathcal{A}$  should be disposed to infer 'white' from 'swan'. In the logical framework, this is represented by the facts that  $w_{ws} < w_{bs}$  (where  $w_{ws}$  and  $w_{bs}$  are the partial worlds in  $\mathcal{U}$  that respectively stand for  $\mathcal{A}$ 's mental representations of white and black swans) and  $\rho(\text{'swan'}) \Vdash_{\mathcal{M}} \rho(\text{'white'})$ . But suppose that now the number of black swans gradually increases over time, and that finally black swans become more numerous than white ones. One expects  $\mathcal{A}$ 's representation of black swans to become more and more vivid, and finally to become more vivid than its representation of white swans. As a consequence, its disposition to infer 'white' from 'swan' should gradually fade away, and be eventually replaced by a disposition to infer 'black' from 'swan'.

It appears that in all three cases the considered learning process can be represented in the framework by an appropriate modification of the model  $\mathcal{M} = (\mathcal{U}, <)$ . More precisely, the addition of a new element to the set of  $\mathcal{A}$ 's mental representations can be represented by the addition of the corresponding partial world in  $\mathcal{U}$ , while the supplementation of a mental representation with a new feature can be represented by the replacement of the partial world that stands for the considered mental representation by the suitable precisification of this same partial world (in the given example, the partial world  $w$  such that  $\delta(w) = \rho(\text{'white'}) \wedge \rho(\text{'swan'})$  should be removed from  $\mathcal{U}$ , and its precisification  $w'$  such that  $\delta(w') = \rho(\text{'white'}) \wedge \rho(\text{'swan'}) \wedge \rho(\text{'black legs'})$  should be added). Finally, a change in the vividness of  $\mathcal{A}$ 's mental representations can be accounted for by a suitable modification of  $<$ .

If one regards  $\mathcal{M}$  as representing the agent's worldview, then such a modification would reflect the fact that, at the mental level of description, learning can be seen as the revision of one's worldview. If, alternatively, one regards  $\mathcal{M}$  as representing the agent's inferential system, it would reflect the fact that at the neural level learning consists in a rearrangement of the agent's inferential system.

In any case, the suggested revision of  $\mathcal{M}$  would result in a revision of the inference relation it induces. The latter should naturally be interpreted as a revision of the agent's dispositions to infer — that is, as a revision of its general knowledge about things<sup>1</sup>.

---

<sup>1</sup> Note that the set  $\mathcal{F}^+$  of the features  $\mathcal{A}$  is physiologically able to perceive (and thus the language  $\mathcal{L}$ ), is left unchanged by the considered learning processes. This is necessary, since to be able to have the idea that a given object/situation satisfies such and such features, the agent must already be able to perceive them.

We would thus get a natural articulation between the modelling of automatic inferences and that of their associated learning processes. It should be noted that the latter would be indefinitely iterable, since a precisification-free model would always be revised into another, uniquely defined, precisification-free model. This remains of course to be properly formalized, and in particular it should be specified how, in each of the above three cases,  $<$  should be revised. This will be the subject of a further paper.

## References

- Bartos, M. (2008). Hunting prey with different escape potentials — alternative predatory tactics in a dune dwelling salticid. *The Journal of Arachnology*, 35, (pp. 499–508).
- Chittka, L. & Niven J. (2009). Are bigger brains better? *Current Biology*, 19, (pp. 995–1008).
- Dubois, D. (2008). On ignorance and contradiction considered as truth-values. *Logic Journal of the IGPL*, 16:2, (pp.195-216).
- Gabbay, D. M. & Schlechta, K. (2008). Cumulativity without closure of the domain under finite unions. *The Review of Symbolic Logic*, 1:3, October 2008 (pp. 372–392).
- Gabbay, D. M. & Schlechta, K. (2009). Roadmap for preferential logics. *Journal of Applied Non-classical Logics*, 19:1, (pp. 43–95).
- Jackson, R. R. & Cross, F. R. (2011). Spider Cognition. *Advances in Insect Physiology*, 41, (pp. 115–174).
- Kraus, S., Lehmann, D. & Magidor, M. (1990). Non-monotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44 (Page numbers are those of the arXiv version).
- Kripke, S. (1965). Semantical analysis of intuitionistic logic I. In J. Crossley and M. Dummett, editors, *Formal systems and recursive functions* (pp. 92–130). North-Holland, Amsterdam.
- Lehmann, D. & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, 55 (Page numbers are those of the arXiv version).
- Leitgeb, H., (2004). Inference on the low-level, an investigation into deduction, non-monotonic reasoning and the philosophy of cognition. *Kluwer Academic Publishers*.
- Lin, L. & al. (2007). Neural encoding of the concept of nest in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, 104:14, (pp. 6066–6071).
- Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, 13, (pp. 587–597).
- Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, 13:1, (pp. 90–99).
- Tsunoda, K. & al. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4:8, (pp. 832–838).

## Appendix A.

For any  $w \in \mathcal{U}$  and any  $L^{\parallel}$ -formula  $\alpha$ ,  $\delta(w) \Vdash_{\langle \mathcal{U}, L^{\parallel} \rangle} \alpha$  iff  $w \Vdash \alpha$ .

*Proof:*

$\implies$ ) Let  $w \in \mathcal{U}$  and let  $\alpha$  be a  $L^{\parallel}$ -formula such that  $\delta(w) \Vdash_{\langle \mathcal{U}, L^{\parallel} \rangle} \alpha$ .  
 $w \Vdash \delta(w)$ , so by the definition of  $\Vdash_{\langle \mathcal{U}, L^{\parallel} \rangle}$ ,  $w \Vdash \alpha$ .

- $\Leftarrow$ ) Let  $w \in \mathcal{U}$  and let  $\alpha$  be a  $\mathbf{L}^\perp$ -formula such that  $w \models \alpha$ . We proceed by induction on the construction of  $\mathbf{L}^\perp$ -formulas:
- i) (Base case) If  $\alpha$  is a  $\mathbf{L}$ -formula, then  $\delta(w) \vdash \alpha$  (see section 3.1). Thus by supra-classicality of  $\Vdash_{\mathcal{U}, \mathbf{L}^\perp}$ ,  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \alpha$ .
  - ii) (Induction step) If  $\alpha$  is not a  $\mathbf{L}$ -formula, then, either  $\alpha = \beta \parallel \gamma$  or  $\alpha = \beta \wedge \gamma$ . Suppose that we have already proven that for any  $w' \in \mathcal{U}$ ,  $w' \models \beta$  entails  $\delta(w') \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta$  and  $w' \models \gamma$  entails  $\delta(w') \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \gamma$ .
    - 1. If  $\alpha = \beta \parallel \gamma$ , then by hypothesis
      - $w \models \beta \parallel \gamma$ , iff (by def. of  $\parallel$ )
      - $w \models \beta$  or  $w \models \gamma$ , thus (by induction hypothesis)
      - $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta$  or  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \gamma$ .
        - If  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta$ , then by the def. of  $\Vdash_{\mathcal{U}, \mathbf{L}^\perp}$ ,
        - $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta \parallel \gamma$ , i.e.  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \alpha$ .
        - Otherwise,  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \gamma$ , so by the def. of  $\Vdash_{\mathcal{U}, \mathbf{L}^\perp}$ ,
        - $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta \parallel \gamma$ , i.e.  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \alpha$ .
    - 2. If  $\alpha = \beta \wedge \gamma$ , then by hypothesis
      - $w \models \beta \wedge \gamma$ , iff (by def. of  $\wedge$ )
      - $w \models \beta$  and  $w \models \gamma$ , thus (by induction hypothesis)
      - $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta$  and  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \gamma$ , iff
      - $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \beta \wedge \gamma$ , i.e.  $\delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \alpha$ . □

## Appendix B.

### *Injectivity:*

**For any formula  $\alpha$  maximal-consistent in  $\mathcal{U}$ ,**

**if  $\alpha \parallel \beta \parallel \gamma \Vdash_{\mathcal{M}} \beta \parallel \gamma$  and  $\alpha \parallel \beta \not\Vdash_{\mathcal{M}} \beta$ , then  $\alpha \parallel \gamma \Vdash_{\mathcal{M}} \gamma$ .**

*Proof:*

Let  $\alpha$ ,  $\beta$  and  $\gamma$  be  $\mathbf{L}^\perp$ -formulas such that  $\alpha$  is maximal-consistent in  $\mathcal{U}$ ,  $\alpha \parallel \beta \parallel \gamma \Vdash_{\mathcal{M}} \beta \parallel \gamma$  and  $\alpha \parallel \beta \not\Vdash_{\mathcal{M}} \beta$ , and let  $w \in \mathcal{U}$  be  $<$ -minimal for  $\alpha \parallel \gamma$  (if there is no such  $w$ , then, trivially,  $\alpha \parallel \gamma \Vdash_{\mathcal{M}} \gamma$ ). Suppose that  $w \not\models \gamma$ :

- i) Then  $w \models \alpha$ .
- ii) By hypothesis  $\alpha$  is maximal-consistent, so  $\alpha$  is a  $\mathbf{L}$ -formula. Since  $w \models \alpha$ ,  $\delta(w) \vdash \alpha$ .
- iii)  $w \models \alpha \wedge \delta(w)$ , so  $\alpha \wedge \delta(w) \Vdash_{\mathcal{U}, \mathbf{L}^\perp} \perp$ . By def. of maximal-consistent formulas,  $\alpha \vdash \delta(w)$ . So by ii),  $\alpha \equiv \delta(w)$ .
- iv) By hypothesis  $\alpha \parallel \beta \not\Vdash_{\mathcal{M}} \beta$ , iff
  - $\exists w' \in \mathcal{U}$  s.t.  $w'$  is  $<$ -minimal for  $\alpha \parallel \beta$  and  $w' \not\models \beta$ . Thus  $w' \models \alpha$ , iff  $\delta(w') \vdash \alpha$ .
  - Since  $\alpha$  is maximal-consistent,  $\alpha \vdash \delta(w')$ , so  $\alpha \equiv \delta(w')$ .
- v) By iii) and iv),  $w = w'$ , thus by iv),  $w$  is  $<$ -minimal for  $\alpha \parallel \beta$  and  $w \not\models \beta$ .
- vi) Since  $w \models \alpha$ ,  $w \models \alpha \parallel \beta \parallel \gamma$ . We show that  $w$  is  $<$ -minimal for  $\alpha \parallel \beta \parallel \gamma$ :
  - Suppose the contrary, then there is a  $w'' < w$  such that  $w'' \models \alpha \parallel \beta \parallel \gamma$ . By  $\mathbf{L}^\perp$ -smoothness,  $w \not\prec w''$ , so  $w'' \neq w$ , thus since  $\alpha$  is maximal-consistent,  $w'' \not\models \alpha$ .
  - So  $w'' \models \beta \parallel \gamma$ . If  $w'' \models \beta$ , then  $w'' \models \alpha \parallel \beta$ , so  $w$  is not  $<$ -minimal for  $\alpha \parallel \beta$ , which contradicts v). If  $w'' \not\models \beta$ , then  $w'' \models \gamma$ , so  $w'' \models \alpha \parallel \gamma$ , so  $w$  is not  $<$ -minimal for  $\alpha \parallel \gamma$ , which contradicts the hypothesis. Thus  $w$  is  $<$ -minimal for  $\alpha \parallel \beta \parallel \gamma$ .
- vii) By hypothesis  $\alpha \parallel \beta \parallel \gamma \Vdash_{\mathcal{M}} \beta \parallel \gamma$ , so by vi),  $w \models \beta \parallel \gamma$ .
  - By v)  $w \not\models \beta$ , so  $w \models \gamma$ , which contradicts the hypothesis. □

## Appendix C.

### $\parallel$ -Disjunct Equivalence:

If  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ ,  $\alpha_2 \parallel_{\mathcal{M}} \alpha_1$  and  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \gamma$ , then  $\alpha_2 \parallel \beta \parallel_{\mathcal{M}} \gamma$ .

*Proof:*

Let  $\alpha_1, \alpha_2, \beta$  and  $\gamma$  be  $\mathbf{L}^\parallel$ -formulas such that  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ ,  $\alpha_2 \parallel_{\mathcal{M}} \alpha_1$  and  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \gamma$ .

- i) By hypothesis,  $\alpha_2 \parallel_{\mathcal{M}} \alpha_1$  so by *U-Right Weakening*,  $\alpha_2 \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ .
- ii) By *Reflexivity*,  $\alpha_1 \parallel_{\mathcal{M}} \alpha_1$ , so by *U-Right Weakening*,  $\alpha_1 \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ .  
By  $\parallel$ -Or on i),  $\alpha_1 \parallel \alpha_2 \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ .
- iii) By *Reflexivity*,  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ , so by  $\parallel$ -Or on ii),  $(\alpha_1 \parallel \alpha_2) \parallel (\alpha_1 \parallel \beta) \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ .  
Since  $(\alpha_1 \parallel \alpha_2) \parallel (\alpha_1 \parallel \beta) \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha_1 \parallel \alpha_2 \parallel \beta$ , by *U-Left Equivalence*  
 $\alpha_1 \parallel \alpha_2 \parallel \beta \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ .
- iv) By *Reflexivity*,  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \alpha_1 \parallel \beta$ , so by *U-Right Weakening*,  
 $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \alpha_1 \parallel \alpha_2 \parallel \beta$ .
- v) By hypothesis  $\alpha_1 \parallel \beta \parallel_{\mathcal{M}} \gamma$ , so by the derived rule *Equivalence*, on iii) and iv),  
 $\alpha_1 \parallel \alpha_2 \parallel \beta \parallel_{\mathcal{M}} \gamma$ .
- vi) By hypothesis,  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$  and by *Reflexivity*  $\alpha_2 \parallel_{\mathcal{M}} \alpha_2$ , so by  $\parallel$ -Or  $\alpha_1 \parallel \alpha_2 \parallel_{\mathcal{M}} \alpha_2$ .  
So by *U-Right Weakening*,  $\alpha_1 \parallel \alpha_2 \parallel_{\mathcal{M}} \alpha_2 \parallel \beta$ .
- vii) By *Reflexivity*  $\beta \parallel_{\mathcal{M}} \beta$ , so by *U-Right Weakening*,  $\beta \parallel_{\mathcal{M}} \alpha_2 \parallel \beta$ .  
So by  $\parallel$ -Or on vi),  $\alpha_1 \parallel \alpha_2 \parallel \beta \parallel_{\mathcal{M}} \alpha_2 \parallel \beta$ .
- viii) By *Cautious Monotony* on v) and vii),  $(\alpha_1 \parallel \alpha_2 \parallel \beta) \wedge (\alpha_2 \parallel \beta) \parallel_{\mathcal{M}} \gamma$ .  
Since  $(\alpha_1 \parallel \alpha_2 \parallel \beta) \wedge (\alpha_2 \parallel \beta) \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha_2 \parallel \beta$ , by *U-Left Equivalence*,  
 $\alpha_2 \parallel \beta \parallel_{\mathcal{M}} \gamma$ . □

## Appendix D.

### $\parallel$ -Transitivity:

If  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ ,  $\dots$ ,  $\alpha_{n-1} \parallel_{\mathcal{M}} \alpha_n$ , then  $\alpha_1 \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_n$ .

*Proof:*

Let  $\alpha_1, \alpha_2, \dots$ ,  $\alpha_n$  be such that  $\alpha_1 \parallel_{\mathcal{M}} \alpha_2$ ,  $\dots$ ,  $\alpha_{n-1} \parallel_{\mathcal{M}} \alpha_n$ .

- i) By hypothesis, for any  $i(1 \leq_N i <_N n)$ ,  $\alpha_i \parallel_{\mathcal{M}} \alpha_{i+1}$ , and, by *Reflexivity*  
 $\alpha_{i+1} \parallel_{\mathcal{M}} \alpha_{i+1}$ .  
Thus by  $\parallel$ -Or, for any  $i(1 \leq_N i <_N n)$ ,  $\alpha_i \parallel \alpha_{i+1} \parallel_{\mathcal{M}} \alpha_{i+1}$ .
- ii) By *Reflexivity*, for any  $i(1 \leq_N i <_N n)$ ,  $\alpha_{i+1} \parallel_{\mathcal{M}} \alpha_{i+1}$ , so by *U-Right Weakening*,  
for any  $i(1 \leq_N i <_N n)$ ,  $\alpha_{i+1} \parallel_{\mathcal{M}} \alpha_i \parallel \alpha_{i+1}$ .
- iii) By i),  $\alpha_{n-1} \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_n$ , so by successive applications of the derived rule  
*U-Disjunct Equivalence* on i) and ii), one gets  $\alpha_1 \parallel \alpha_2 \parallel \dots \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_n$ .
- iv) By *U-Right Weakening* on iii),  $\alpha_1 \parallel \alpha_2 \parallel \dots \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_1 \parallel \alpha_n$ .
- v) By *Cautious Monotony* on iii) and iv),  $(\alpha_1 \parallel \alpha_2 \parallel \dots \parallel \alpha_n) \wedge (\alpha_1 \parallel \alpha_n) \parallel_{\mathcal{M}} \alpha_n$ .  
 $(\alpha_1 \parallel \alpha_2 \parallel \dots \parallel \alpha_n) \wedge (\alpha_1 \parallel \alpha_n) \cong_{\mathcal{U}, \mathbf{L}^\parallel} \alpha_1 \parallel \alpha_n$ , so by *U-Left Equivalence*,  
 $\alpha_1 \parallel \alpha_n \parallel_{\mathcal{M}} \alpha_n$ . □

## Appendix E.

### Rankedness:

For any  $\mathbf{L}^\parallel$ -formulas  $\alpha_1, \alpha_2$  and  $\alpha_3$  maximal-consistent in  $\mathcal{U}$  (all  $\neq$ ),  
if  $\alpha_1 \parallel \alpha_2 \parallel_{\mathcal{M}} \alpha_2$  and  $\alpha_1 \parallel \alpha_3 \not\parallel_{\mathcal{M}} \alpha_3$ , then  $\alpha_2 \parallel \alpha_3 \parallel_{\mathcal{M}} \alpha_2$ .

*Proof:*

Let  $\mathcal{M} = (\mathcal{U}, <)$  be a finite ranked partial worlds model, and let  $\alpha_1, \alpha_2$  and  $\alpha_3$  be (all  $\neq$ )  $\mathbf{L}^{\#}$ -formulas maximal-consistent in  $\mathcal{U}$  and such that  $\alpha_1 \parallel \alpha_2 \not\|_{\mathcal{M}} \alpha_2$  and  $\alpha_1 \parallel \alpha_3 \not\|_{\mathcal{M}} \alpha_3$ . In addition, let  $w \in \mathcal{U}$  s.t.  $w$  is  $<$ -minimal for  $\alpha_2 \parallel \alpha_3$  (if there is no such  $w$  in  $\mathcal{U}$ , then, trivially,  $\alpha_2 \parallel \alpha_3 \not\|_{\mathcal{M}} \alpha_2$ ). Suppose that  $w \not\| \alpha_2$ :

- i) Then  $w \Vdash \alpha_3$ .  
Thus  $w \Vdash \alpha_1 \parallel \alpha_3$ .
- ii) By hypothesis  $\alpha_1 \parallel \alpha_3 \not\|_{\mathcal{M}} \alpha_3$ , iff  
there is a  $w'$  in  $\mathcal{U}$  such that  $w'$  is  $<$ -minimal for  $\alpha_1 \parallel \alpha_3$  and  $w' \not\| \alpha_3$ .  
Since  $<$  is modular, by i),  $w' \leq w$ .
- iii) Since  $w' \Vdash \alpha_1 \parallel \alpha_3$  and  $w' \not\| \alpha_3$ ,  $w' \Vdash \alpha_1$ . Thus  $w' \Vdash \alpha_1 \parallel \alpha_2$ .
- iv) Since  $w' \Vdash \alpha_1$  and  $\alpha_1$  is maximal-consistent,  $\alpha_1 \equiv \delta(w')$ .
- v) We show that  $w' \not\| \alpha_2$ : suppose the contrary, then since  $\alpha_2$  is maximal-consistent, we get  $\alpha_2 \equiv \delta(w')$ , thus by iv)  $\alpha_1 \equiv \alpha_2$ , which contradicts the hypothesis.
- vi) By hypothesis  $\alpha_1 \parallel \alpha_2 \not\|_{\mathcal{M}} \alpha_2$ , so by iii) and v),  $w'$  is not  $<$ -minimal for  $\alpha_1 \parallel \alpha_2$ . Since  $\mathcal{M}$  is finite,  $<$  is  $\mathbf{L}^{\#}$ -smooth, thus there is  $w'' < w'$  such that  $w''$  is  $<$ -minimal for  $\alpha_1 \parallel \alpha_2$ .  $w'' \Vdash \alpha_2$ , so  $w'' \Vdash \alpha_2 \parallel \alpha_3$ .
- vii) By vi)  $w'' < w'$ , and by ii),  $w' \leq w$ , so by modularity of  $<$ ,  $w'' < w$ .
- viii) By vi),  $w'' \Vdash \alpha_2 \parallel \alpha_3$ , so by vii)  $w$  is not  $<$ -minimal for  $\alpha_2 \parallel \alpha_3$ , which contradicts the hypothesis. □