A Formal Model through Homogeneity Theory of Adaptive Reasoning

by

Roberto Garigliano and Derek Long

Technical Monograph PRG-71
ISBN 0-902928-53-8
February 1989

Oxford University Computing Laboratory
Programming Research Group
8-11 Keble Road
Oxford OX1 3QD
England

Oxford University Computing Laboratory
Programming Research Group
8 11 Keble Road
Oxford OX1 3QD
England

A Formal Model through Homogeneity Theory of Adaptive Reasoning

Roberto Garigliano and Derek Long

Abstract

We address the problem of how to deal with inaccurate, incomplete and changing information. We proceed by formally comparing existing reasoning systems in order to precisely define which features are needed and which should be avoided. In order to do so, we construct a formal theory, which we call *homogeneity theory*. The equivalence of the transformation rule of identity, for each reasoning system, to some expressions of homogeneity theory is proved. An order is then induced among the systems analysed using the expressions of homogeneity theory.

Some natural criteria are formally defined in order to evaluate the respective power of the systems, which lead to a second order of the systems, embedded in the first. An intermediate model and finally the model, called *adaptive reasoning system*, are specified. They are proved to be stronger than the systems previously examined, according to the criteria defined.

The central idea in the adaptive reasoning system is the attempt to recognize patterns of behaviour in sources of information, or areas of interest, and regulate the reliability of the sources and the stability of the areas accordingly. The adaptive reasoning system is equipped with yardsticks against which to judge and regulate its own performance.

Acknowledgements

Derek Long's contribution could hardly be overestimated: I am most grateful for his critical spirit and his very scientific detachment from his own opinions, which make the most strenuous discussion such a rewarding experience. Because of our continuous collaboration during and after the completion of this work, it is only right for us to share credit for it.

I wish to thank my supervisor, Joe Stoy, for his openness in helping me through a field of study which was not his own, and for his readiness and ability in smoothing down problems that could have otherwise caused me delays and frustrations, always in that pleasant style so characteristic of him.

The Programming Research Group and St Cross College have, between them, created a most relaxed, enjoyable and beneficial environment.

To Dr. Moshe' Machover and Prof. Marco Mondadori I owe most of the logic I know and the very fruitful suggestion to try and use it in computer science.

Dr. Richard Bird has provided essential encouragement in a moment of great depression, when I could not even find a problem. Prof. Bennett has given me a substantial dose of self-confidence by believing in my work at an early stage.

Hary, Kavi, Ella, Maureen, Nigel and Dave have supported in the thousand occasions when a good friend is needed. Marco Capria has offered some very sharp criticism. Liz Knight has very kindly proof-read the text.

To my parents, Francesca and Glauco, I owe the motivation and the energy that come from a continuous flow of love. This thesis has its roots there.

Contents

Chapter 4 A Formal Analysis of Five Classes of Reasoning Systems

Chapter 1

Introduction

1.1 The Crisis of Growth in A.I.

Artificial Intelligence is a new field, at most forty years old, depending on the preferred birth date. Most of the advances have taken place in the last twenty years, since the development of the first expert systems.

It is no wonder, then, that the subject appears to be in a rather chaotic state, with a generally acknowledged lack of theoretical foundations, an unfortunate tendency to the hyperbolic in titles of papers, and an inflated use of terms for which there are no standard definitions (often no definitions at all) or for which the new use does not match the accepted one. At the same time the area is loaded with meticolous analysis, calculations and implementations so disproportionate with respect to the weakness of their central ideas that one could be forgiven for thinking that the former are intended to disguise the latter. On top of all this, A.I. is more and more attracting an enthusiastic interest from industry and governments (probably born out of fear) which, while extremely beneficial to the researchers and their institutions, tends to put an enormous pressure to deliver here and now: this is usually not the best recipe for healthy, steady growth.

This situation however is not due merely to the exuberance of youth; two other important factors play a key role. The first is the nature of the subject itself: on one hand, the final goal is such that it can only produce great enthusiasm or complete scepticism - hardly anything in the middle - once the real objective is visualised; on the other hand, there are intermediate goals of many kinds to be achieved along the way, even for those who cannot see the final aim or do not believe in it.

The second factor is the intrinsic interdisciplinarity of the subject: logic, mathematics, philosophy of several kinds, linguistics, psychology, statistics, sociology, anatomy and physiology (human and animal) all have an interest in A.I., a contribution to make and a potential gain to extract. It is only to be expected that such a wide ranging collaboration should cause clashes of methodology, terminology and backgrounds, attempts to steer in opposite directions, and gaps in communications, with consequent duplication of efforts.

To summarise, A.I. is a young subject, with a very ambitious goal and lots of sub-goals and by-products; it is the cross-point for several previously unrelated disciplines; and it is very hot, in the sense that new developments can spring up (and disappear) at a very high rate. All this is of course very nice, expecially if one shares the interest for the ultimate objective and has some longing for a less specialised, more interconnected ideal of science. On the other hand, this situation brings with itself the problems of confusion, lack of rigour and wild claims already mentioned.

We would therefore suggest that the present work be read on two different levels: as a contribution to the methodology of A.I., and as a contribution to a particular area of the subject, which we consider central, namely the theory of reasoning engines. The former is embedded in the structure of this work, but not explicitly discussed anywhere else, and for this reason we explain it here in more detail.

1.2 A Methodological Contribution

There are of course many very well organised papers in this area, but we think that there is still the scope and the need for a further attempt at standardization. A.I. is an applied science of the engineering kind, in the sense that the overall goal of the enterprise is to *construct* something, or at least to give a complete *specification for construction*. Any contribution to the field must then start with a problem affecting that overall goal, and discuss some kind of solution to it. The starting point of

the analysis can and often must be very far away from the solution, but the *liaison* must be present already; otherwise the work is better considered as a piece of pure mathematics or linguistics et cetera, which could subsequently be applied to A.I.. It is all too natural for researchers coming from other disciplines to pursue their old interests under a new banner. Tempting as it may be, this practice causes, at best, an excess of material useless in A.I. terms, which ends up obscuring the relevant contribution.

At this point it is worth stating what is increasingly recognised as the overall goal of A.I.: A.I. is concerned with *simulation of successful human behaviour*. Note that the qualification *intelligent* is unnecessary, since trivially if human behaviour is reproduced in general, its intelligent subset is reproduced as well, while unintelligent behaviour is by definition cruder than the intelligent kind, and thus it does not represent any additional task. The advantage in avoiding the term "intelligent" is that our understanding of it is not only very fuzzy but, what is worse, strongly dependent on social criteria. For example, diagnosing with considerable degree of success a lung illness is naturally considered a higher kind of activity than driving a car in a jammed town, simply because very many of us can easily learn the second skill, while only few possess the first. It has nevertheless turned out that the former task is much easier to simulate than the latter, and for reasons of adaptive ability which, at least in our view, capture the common meaning of "intelligence" (as well as the evolutionary one) better than a prepackaged set of deductions. The term "successful", on the other hand, is easily understood as "achieving the desired goal", whatever the goal is. The concept of "successful" is much safer to use than that of "intelligent" because the former is naturally relativized, while the latter claims a very controversial absoluteness.

A second point worth noting in this definition is the stress on reproduction, or simulation, of human *behaviour*, instead of human *mental processes*. This is essential in order to distinguish between A.I. and, say, experimental psychology. While the two disciplines can have a profitable interaction, their aims and methods can be very different: essentially, A.I. can use modelling techniques of any kind, as long as they allow a needed result to be reached, while psychology must concentrate on those which are plausibly used by the mind itself. Similarly, a vision system by radar would be just

as good for the A.I. researcher, but useless for the ophthalmic expert.

Once the problem has been identified, the main ideas used in its analysis should be explicitly stated and, if necessary, discussed. This is the part of the methodology inherited from philosophy: the concept is that there are always hidden assumptions behind apparently obvious statements, and when the assumptions are controversial in the field in which they are used then they should be declared.

The third point is to develop as many of the technical tools needed as possible *before* attacking the central problem. The utility of this precept in terms of modularity and, consequently, ease for proving and modifying units is self evident; it forms part of the standard kit for the working mathematician. A great deal has been said about it recently in software engineering but, in our opinion, from a different perspective. The most important feature in the classic mathematical approach is the construction of theories which are much more abstract and general than the particular problem requires, and their use for discovery and proof of properties which would be far too complicate to analyse at a more specific level, but which will nevertheless still hold once the details are filled in. Most of the "formal methods for specifications", on the other hand, do rather the opposite, starting from specifications which lie on the same plane as the final solution, only much more simple-minded. As a result anything proved there has to be proved again and again, as more details are added, in a time-consuming game sometimes called "data refinement".

Another essential point is to show the match between the philosophy and the formalism. It does not have to be carried out at every point - for example part of the supporting mathematical theory could be justified simply by its use in the final model - but it is surely helpful to match the two sides as often as possible, and it also alleviates the final burden. It could also be claimed that, in order to have developed that particular theory in the first place, the researcher must have had some kind of intuition about its connection with the final problem, and attempting to make this relation explicit would not only help the reader, but also the author.

A further point of methodology is how to relate the main goal of A.I. with the partial, tentative achievements that a particular piece of research in A.I. can hope for. The criterion that we have adopted is that of "extendability". If the proposal, specification, or piece of code which inevitably lies at the end of the paper can be envisaged as a useful step towards a stronger construction, able to deal better with a wider range of problems, then it is probably worth having done it. As a prerequisite, this criterion requires the existence of a larger project, a vision that goes beyond the particular solution. We think that without such a general direction no important results in A.I. can be achieved, given the holistic nature of the central problem. Previous results can be adapted, improved, often just nicely packaged, without the direction of research that we claim essential, but all this has more to do with marketing than with science, or at least with A.I. as we intend it. It is clear, though, that even under such a perspective, the judgment of what will prove useful at a later stage can only be subjective and prone to error. For this reason, we consider this guideline to be on a different plane from the others; nevertheless, we have taken the decision not to complete our specifications when to do so, with the means available at this stage, would have suggested a way forward which we believed to be a dead end. This happens expecially in Chapter 6.

1.3 An Overview of the Work

We now outline how the present works fits into the lines of development that we have advocated above.

Chapter 2 is devoted to a description of the problems we are addressing, to a discussion of some philosophical points about knowledge and reasoning relevant to our understanding of these problems, and to an outline of the general direction of research within which this work is to be considered.

Our problem is not new in A.I.: how to cope with inaccurate, incomplete and changing information. That is to say, according to our definition of A.I.'s main goal, how to simulate, at least partially, a successful human

behaviour under such circumstances. In Chapter 2 we argue informally that the solutions proposed until now, namely systems based (or so it is claimed) on classical, intuitionistic, non-monotonic, fuzzy logics et cetera. are unsatisfactory for several reasons, and we give an outline of what we would consider an improvement on them. We then examine some points which have been raised during this discussion, or which underline some part of the analysis. Among these are the nature and role of inconsistencies, the subjective versus the objective view of acquisition of knowledge, the relation between assumption, reasoning and evidence, and the basic assumption of homogeneity.

Finally, we present a direction of research in which the reasoning mechanism is expected to play a central role in many different ways, from consistency recovery (discussed in this work) to the natural language interface or the sensorial data analyser. While this more general project is little more than an intuition with a lot of hope, several parts of it (like the "reasoning by analogy" module) are already specified or in the process of being constructed.

In Chapter 3 we proceed to the definition and exploration of a mathematical theory which we call *homogeneity theory*. We explain how it relates to some of the points made in the previous chapter, and we prove some interesting results in it that are used later in Chapter 6. We then apply it to the analysis of the standard problem of individuals, gaining some understanding that we expect to be very useful in subsequent developments (mainly beyond the present work).

In Chapter 4 we define the concept of a reasoning system in relation to that of a logic, we then transform five important classes of logics into their respective reasoning systems and proceed to analyse them. We argue that the transformation is necessary because in A.I. we need a reasoning *entity* rather than a reasoning *tool*. This part of the work is connected with the need of formally comparing the solutions that have been presented to our problem and showing their weaknesses. In order to do this, a formal framework is constructed in which it is possible to carry out analysis and proofs about the behaviour of the reasoning systems in their interaction with the environment.

Chapter 5 is devoted to the analysis of the five classes of reasoning systems through homogeneity theory. In particular, we prove that the rules of identity convert into rather complex assumptions of homogeneity, and that these in turn can be placed in a partial order. A discussion of the implications of these results follows. We then propose some very natural criteria to judge the performance of the reasoning systems in relation to their environments, namely their ability to "survive" and "react", and we prove that these tests impose an order on the systems which is embedded in the previous one. The consequences of these proofs are then discussed.

In Chapter 6 we first present an intermediate model, called *unforgiving*, which is useful to complete the comparison of the other five. We then proceed to the construction of the final model, belonging to the class of *adaptive* reasoning systems.

The part of the adaptive reasoning systems presented here is concerned with what we call a *consistency recovery mechanism*. As mentioned before, places are left where we foresee modules developed in the future, using additional techniques, fitting. This occurs when a possible solution at this stage does not point towards the future development that we envisage. Thus, notwithstanding its appearance as a functional language program, this model is not to be seen as a complete, implementable specification, but rather as a formally defined step in the direction we intend to pursue. On the other hand we have taken care not to specify any mechanism which would be, once implemented, computationally not feasible.

In the definition and construction of this model, as well as in the complexity analysis for some parts of it, we have drawn heavily from our definitions and results of Chapter 3 about homogeneity theory. Since the adaptive reasoning system is a natural extension of the unforgiving one, we do not prove formally its dominant position in relation to the five previously examined. This is an application of the technique to prove results over simple models that could be easily reworked for the extended models, but where the proofs would then be too long and tedious.

Conclusion are drawn in Chapter 7.

1.4 Reading this Work

In organising the presentation of this work we have adopted virtually throughout the following schema. First comes a general overview of the ideas behind the material to be exposed in the chapter. This is intended to convey a flavour, and to set the right frame of mind: for these reasons it is often left very open and not tightly connected to the formal part that usually follows.

We then proceed to list a series of definition and proofs with very little English explanation; a note at the beginning of the section advises where the explanations are to be found. We are aware of the difficulties that this approach can create, but we have nevertheless decided to use it for the following reason: while any expression, in a formal or natural language, can be interpreted in several different ways according to the background knowledge used, a natural language tends to suggest that one of these "enriched" interpretations is in fact intended by the author. Once this impression is created, one tends to force everything that follows into that frame, instead of going back and modifying the interpretation.

A formal language, on the other hand, not only specifies exactly the minimal common interpretation required, but also, because of its being so alien, positively discourages any intuitive addition of meaning from anybody but the most immersed of mathematicians. We believe that the additional context dependent meaning is absolutely necessary, but that having to pass first through the formal definitions helps the mind to keep some critical detachment from its own interpretations.

The cross-references provided allow, in any case, for a different style of reading. Not only could the explanations be read before the formal definitions and proofs, but also, for example, the basic definitions of the theory of homogeneity could be left until Chapter 5, while the more complex results of Chapter 3 are not used until Chapter 6. We have valued rigour and modularity more than smoothness and immediate applicability, but we recognise that these are mainly questions of taste.

We have explicitly referred in the text to a particular author only when our interpretation could have given rise to controversy, or when it could otherwise have appeared as ours original contribution. Otherwise, even when some particular source could have been identified, we have preferred to omit the reference, while presenting the standard interpretation. We have done so mainly because most of the ideas have been around some time, or exist in many different shades, and we have neither the qualification nor the interest for assigning intellectual patents.

Chapter 2

Problems, Philosophy and Direction of Research

2.1 The Problems

The problem we are addressing is how to cope with *inaccurate, incomplete and changing* information. There can be little doubt that this is the kind of information we receive most of the time, and, by our definition of A.I., coping succesfully with it is consequently a challenge for any A.I. program.

While the precise definition of a *successful* behaviour is clearly open to controversy, we think there are some basic features upon which most of us would agree as forming an essential part of any succcessful response.

The first of these features is clearly the ability to maintain an interaction with the world, despite the deceiving quality of the information the world is providing. This may seem a trivial requisite, but that is because the need to interact is so entrenched in us that we take it for granted. In a machine, however, it has to be reproduced in some way, and this creates some interesting consequences, as we show in Chapter 5.

The second, natural requirement is that the system should learn from previous mistakes or misrepresentations: again, the idea of *learning* is rather fuzzy, but surely must include the concept that if one goes through the same mistake, or accepts the same misrepresentation, twice under similar circumstances, then that person has not learnt from the previous experience. There are some exceptions to this rule, as always, but in most cases it is very sound and intuitive.

Also, the change in behaviour must be consistent with some principle: for example, we would not think much of a random reaction. The principle we follow is that the reaction is expected to minimise the chances of the same problem occuring again in the future, while at the same time least reducing the interaction with the environment. There are clearly other possible ways of balancing these two guidelines, or even other possible guidelines altogether. We suggest that in the humans, at the top level, these principles could be very resistant to change, if not fixed altogether, but that a large body of reasonable changes and adaptations is possible within the framework of fixed guidelines. We outline a discussion of these basic principle, which we call *motivations*, in Section 3.10. The problems of how to balance them, and of what kind of variations are possible in that fixed frame are examined in Chapter 6.

There is then a second, more sophisticated aspect of learning, that is learning about the aptness of one's own reactions. That is necessary because even a very reasonable reaction can turn out to be wrong, or simply badly tuned. We would then expect the system to be able to re-examine any previous decision. In order to do so, a *tension* between different motivations must be created. Also, the system should monitor its own behaviour as well as that of its environment, and the same basic criterion for learning should be applied to its reactions, so that if a pattern of decisions does not achieve the result hoped for, the same kind of reaction will not be used identically a second time, in similar conditions.

Another important feature is connected to the need for taking action. We would expect a system to recognise that the wrong information, believed sufficiently strongly to act upon it, is much more serious than information which was not trusted up to the *action point* in the first place. As a consequence, the reaction should be that much more drastic in the former case.

Finally, we all know that many problems, apparently due to bad information, are in fact a consequence of some kind of misunderstanding. The question of natural language understanding is beyond this work, but we claim that some of these problems have a direct root within the logical structure of the reasoning system, typically the assumption of persistency

over time, which can cause changing information to be taken for inaccurate. Chapter 5 and 6 are partly concerned with this issue.

We point out that, despite the complexity of the problems addressed and the reactions required, the analysis and the partial solutions proposed take place at a very basic level in the organization of a reasoning system: as a matter of fact, most of the work pivots upon the *transformation rule of identity*, while even an elementary rule, like modus ponens, is avoided. This is because we believe that *it is in the simple acts of accepting and preserving data that a large part of these problems arise*, and thus a good solution should reflect this.

There are, on the other hand, several similar problems for which an adequate solution can only involve additional capacities, be it reasoning power, background knowledge or interface sophistication. A typical case is the contradiction between different sources (cf. Section 6.1). Some possible ways forward are outlined in Chapter 7.

2.2 A Criticism of Existing Solutions

We now briefly discuss some solutions which have been proposed to parts of the problem. The arguments presented here are of an intuitive kind, and reflect the intuitions behind the present work. A formal criticism is embedded in the analyses of Chapter 5.

The first point we want to raise is that the logics that have been suggested as the core of a reasoning machine, have been originally devised as tools, not descriptions of a working entity. The translation is not difficult, of course, but it is important to make it, in order to avoid the complex parts of a reasoning process being carried out only in the minds of author and reader, when the merit is claimed for the mechanism (theoretical or implemented). An interesting consequence of this obvious requirement is that any meta-level process used, must be specified inside the system. For the rest of this section we will adapt to the common

terminology, referring to logic as agents.

The Prolog community, on one hand, claims that classical logic is all we need to reason efficiently; our opinion is that Prolog itself is not based on classical logic, but on a form of non-monotonic logic. This is due to the presence of the overriding and of the negation-by-failure features. A detailed discussion of these points can be found in Chapter 4.

The non-monotonic logic community, on the other hand, has focused explicitly on the need to deal with contradictory information, but ignored all the problems about consistency and reliability that a fully committed approach to inconsistent information can create.

The problem of how to manage contradictions which arise is clearly essential in order to deal with uncertain information: in fact, in almost all kinds of problems, uncertainty must be accepted at the meta-level as well as at the object level, otherwise we would find ourselves pretending absolute certainty about the extent to which we are uncertain (in which case we are really talking about precision of approximation, not uncertainty of information).

Of course, while approximation is a perfectly safe concept when properly used, meta-level uncertainty can cause incorrect information and inferences, hence contradictions and the need to recover from them. It follows that the problem of recovering from and reacting to contradictions is central to uncertainty logics as well, like fuzzy logics or Incidence calculus.

Here follows the intuitive base of our criticism of all these approaches, in respect to the problem of consistency recovery.

2.3 The Difficulties with Existing Options

Our criticism of the options based on classical and intuitionistic logics is that, in their pure forms, these logics do not "survive" contradictions at all, in the sense that they cannot admit them and preserve some part of the

theory under examination. It is our contention that, because of this reason, no working system is actually based on the pure forms of these logics, but that they all include features which are either completely extraneous to the spirit of the logic or, at best, embed meta-level operations that, from the outisde observer's point of view, flatten these systems onto (weak) non-monotonic ones.

We identify the following shortcomings with systems based on non-monotonic logics and uncertainty logics.

(i) *No learning from experience* - This is the central point of our criticism and, consequently, the attempt towards its solution is the main part of our work. In a standard non-monotonic logic the only message conveyed by a contradiction is that a piece of information previously believed true is actually false (for the time being). We think that the most important thing that can be learned from the discovery of an inconsistency is something other than this, namely that a source of information is not as trustworthy as was assumed, or that a subject is not in as well-ordered a state as it appeares, or that an (unstable) inference technique is less reliable than it seemed, or a combination of all three. This would also be the natural human reaction: we claim that such a feature should be part of the reasoning core itself, and that this extension is central in order to provide a solution for the problems described in (ii), (iii) and (iv).

(ii) *The need for infinite degrees of certainty* - if an uncertainty logic can alter the truth value of a statement as many times as it is required, then, if the logic is to have a model at all, such a model must be equivalent to one with infinitely many degrees of certainty. If not, the model should sooner or later allow the assignment of two opposite truth values with an identical degree of certainty to the same statement at the same time, and this is in contradiction with the concept of logic itself, non-monotonic or otherwise.

Nevertheless, such a model with infinite degrees of uncertainty is clearly unnatural and impossible to translate into a meaningful human scale of degrees of belief (or anything like that). We note that some versions of uncertainty calculus are in fact equivalent to non-monotonic logics, because

they admit infinitely many "adjustements" (cf. Section 4.10).

(iii) *No fixed "action point"* - We call an "action point" that degree of certainty (or belief) high enough for decisions and positive actions to be taken upon it. It is clear that, for most practical purposes, such a degree of belief is as good as the top one.

It is obviously desirable to have the action point fixed at some point, according to some external parameters like urgency, importance of possible effects etc.. This is because it would be ludicrous to consider "x" degree of certainty enough to take some action now, and "$2x$" not good enough for the same thing a little later, under the same conditions. This is what is bound to happen in the model with infinite ordered sentencial values mentioned in (ii) above.

(iv) *No tending towards stabilization* - Since a non-monotonic logic has to accept a change of truth values when the user so wishes, no stable picture of the world could ever emerge from the kind of poor information that we try to deal with: this is at odds with the everyday experience of most of us, where we try very hard to achieve a relatively stabilazed picture of the world.

In fact, it is so important for humans to obtain a steady, lasting image that many of us give up any attempt at improvement and hold fast to their views, no matter how uselessly.

Also, the need for stability is evident in science, where complex, well founded and useful theories are not discarded simply because a counter example has surfaced; if, on the other hand, sufficient evidence has been accumulated against the old theory, and a new, more powerful candidate has been elaborated, then the collapse of the old theory is not perceived as a simple readjustment, but as a "catastrophe" requiring a deep revision (cf. point (i)).

2.4 The Proposed Solutions

We now examine our proposed solutions to the above problems in general terms. In Chapter 6 we will relate such solutions to a mathematical model.

In order to take any action about the source of information and its subject two conditions must be satisfied: firstly, such a source and subject must be identifiable; secondly, the system must be able to adjust, or even refuse, the input of data from the user.

The first condition is self-explanatory. The second is necessary because the final possible action must be downgrading the level of reliability or even refusing the information from that particular source or about that particular subjects. Of course it is possible to react by extending the period during which a new formula is checked for contradictions with the existing database, before being accepted. It is nevertheless clear that, on top of the limits imposed by decidability and feasibility, such a course of action simply limits the damage of contradictions entering the database and being used as a base for inference, but does not weaken the need to bar new inconsistencies of the same kind, by acting on the origin.

The idea that a machine should question, adjust and possibly refuse information input by the user may evoke science-fiction images for some: the point is that we cannot require a program to be as useful as an expert yet behave like a slave. Most of the existing expert systems avoid the problem by having all the essential information coded in by the knowledge engineer, and leaving the user a few slots in the production rules which he can fill.

Once we aim for a really interactive system, and for real-size problems (unsolvable through huge lists of specific production rules) then the choice is no longer avoidable: any expert, human or machine-simulation, must defend its knowledge from unnecessary inconsistencies and draw conclusions about the reliability of particular sources and the general state of a subject.

Such actions can be visualised as an argument between two groups of

operators, acting in opposition: one group trying to downgrade the reliability of the sources and to lower the expectations about uniformity in an area of knowledge, in accordance with the frequency and the seriousness of the contradictions discovered. The other trying to achieve the opposite effect, arguing that no serious contradictions have been discovered and the flow of information reaching the action point is not satisfactory.

This image will be particularly useful in the interpretation of the adaptive reasoning system in Chapter 6.

2.3 Some Related Questions of Philosophy

The ways in which we have defined our problems and specified some requirements for an acceptable solution involve some philophical assumptions, while these assumptions or others of a similar kind have then provided the intuitions upon which we have elaborated the formal material. We think that they should be made clear and open to discussion.

It is useful to notice that, while in the parts of this work where we propose new points of view, as in Chapters 3 and 6, our philosophical perspective is evident and directly translated into the formalisms. Where we discuss and compare other, existent opinions we often stand on a less committed philosophical ground, in order to provide a common denominator. We believe, nevertheless, that the results of these analyses and comparisons support the views which we present in this section.

We hold a *pragmatic* view of science, and knowledge in general, and the yardstick which we use to judge theories or beliefs is by their *effectiveness* in helping to *control* the environment. When we refer to control we mean either capability to predict correctly (up to some standard), or to prescribe constructions (physical or theoretical) which actually carry out the job they are supposed to do.

While this view is clearly related to *falsificationist* and particularly

neo-empiricist philosophies of science, it diverges from those due to the fact that the problem of Truth, as such, is irrelevant under the pragmatic point of view. This extends also to the meta-level of scientific methodology, since, in our view, the requirements of exactness and adequacy cannot be shifted from a scientific theory to its methodological meta-theory. A methodology is then judged by the overall level of control over the environment which it allows, and expecially to its ability to improve this control.

Several other important concepts are related to this. We briefly mention the most relevant to the present work.

From this perspective there is no way of establishing whether the direction taken is the right one in the long term or whether it is just a local maximum. Therefore the concept of *correctness* of an observation, a theory or a methodology must be replaced with that of *adequacy*, which means that the control we have obtained satisfies our present needs and is in line with the general performance of the best comparable techniques.

One of the problems with such a relativistic philosophy of knowledge is that there seems to be no theoretical starting point, in the sense that there is apparently no well-founded way to proceed for a hypothetical rational entity trying to reason by these rules. A standard answer is that there is no objective rationality at all, and that the philosophy of knowledge simply describes an historic process that just happens to be as it is. While there is definitely some ground for this, strictly descriptive (in the biological and social sense) point of view, we think it is possible to relate it to the more classical requirement for justification, as opposed to description.

The argument is that we do need to exercise some control over our environment in order to survive. This requires that at least part of it be constant enough in its behaviour to be predictable. Herein lies what we call the *basic epistemological bet*. If there is not such an order, we do not lose anything in looking for it, since we could not use our mental power in a better way. If, on the other hand, some order exists, then we stand a chance of finding it only if we assume its existence in the first place. Looking for regularity is then rational in an a-priori sense, given

the need to survive. Note that this argument, unlike the similar one proposed by Reichenbach and others, does not try to justify any particular assumption of regularity (which could of course be wrong), but the need for having a mental organization which embeds the search. The more flexible this organization is, the more the search can be varied according to circumstances, and the more chance there is of finding an acceptable candidate.

The link with the evolutionary approach above is that the most plausible (and also the most likely) candidate for the first hypothesis of regularity is the minimal one, that is, the assumption which requires one division in the immediate environment such that some useful property is more likely to be found on one side than on the other. For exmple, those things which are edible, among those things which can be reached. We discuss formalisation of this point of view in Chapter 3, where we examine the theory of homogeneity and the determination of individuals.

It is interesting to note that, while the initial assumption of homogeneity (regularity) should be as weak as possible, the reasoning system supporting it should be already equipped with a much greater power, in order to take advantage of both successes and failures in the application of that assumption. It could be argued that in the human race this power has developed over generations, not individuals, even if there is a large amount of evidence pointing in the opposite direction. However the evolution has taken place, humans of present time seem to be born with the reasoning power set towards looking for and understanding complex regularities, while at the same time the experiments small children conduct as they play point to rather simple and local assumptions to start with. While, as usual, we as A.I. researchers have no obligation to copy the human way of reaching a complex behaviour, this analysis provides a stimulating starting point that we have used heavily in the construction of homogeneity theory and of the adaptive reasoning sytems.

We can now connect the analysis about assumptions of homogeneity with the criteria described above of effectiveness and adequacy, remembering that even apparently neutral observations are identified and interpreted against background expectations. The three terms of the relation, assumptions,

reasoning and observations, appear then to be linked in a circular fashion. This is clearly a departure from the intuitive point of view, according to which the observations are pure data, reasoning is either correct, thus preserving truth, or is not, and the laws are simply extracted from the data through correct reasoning. A great deal of recent research in philosophy of science has gone into proving how naive and useless this representation is, to the extreme point of negating any objectivity whatsoever in the process of gathering knowledge.

Our idea of a circular relation is that each term is judged according to how it fits with the other two: for example, an assumption can be used to interpret some data, reasoning can then be applied to them, the new data so obtained can be matched with the assumption to make a prediction, but which might not fit with a second set of data. It could be that the assumption is inaccurate, the interpretation imprecise, or the reasoning incorrect. The important thing is that, simplifying a little, from the failure of one piece to go into place, we could learn about a more likely shape of at least one of the other two.

This leads us to the concept of contradictions as potentially helpful occurrences, which is, again, linked to the falsificationist valuation of falsification above verification, but with a more positive stress on the large gain in control power that can be achieved when an assumption is recognised as the origin of the problem. In the classical falsificationist version, while the evolutionary gain in the passage from a theory to another is acknowledged, the emphasis is on the negative achievement of being able to definitely classify the old theory as wrong. The general difference is that falsificationism is really positivism upside down, and it is just as much concerned with the discovery of truth, while the pragmatic approach that we advocate is all concentrated towards succesful adaptations.

A further consequence of this more optimistic way of looking at contradictions is that a contradiction is considered existent from the moment in which it is discovered. This follows from the notion that while an unobserved contradiction may very well have created problems in the relation of the entity with its environment, it will have not done so in the entity's internal representation, upon which the entity acts, of that

relation. This is not to say that "objective" contradictions do not exist (even if we would have difficulty in imagining one), but simply that they are, as such, epistemologically irrelevant. On the other hand, once a contradiction is discovered, then it may easily be dated from a moment in the past.

The apparent paradox is due to the difference between these three concepts: external reality (whatever that means), which is completely beyond this analysis; entity–environment relation perceived through the entity's sensitivity and interpretation, which is again different from that which we would perceive the relation to be from outside; and the entity representation of what the environment is. This seemingly bizarre distinction, and others of an analogous kind, have proved to be very useful to us in the construction of the formal frame for analysing and comparing different reasoning systems, in Chapters 4 and 5.

Finally, we have to act and interact with our environment, and the information upon which we do so is often of a poor quality, and anyway we are an integral part of the way in which this information is gathered and interpreted. It follows that we cannot wait for the *valid* reasoning technique to make inferences. Infact, even when the data are precise enough and no undecidable problems are involved, the blindness of valid inferences, their step-by-step nature and the complexity of the exhaustive search they require, make their use almost impossible except in the most restricted circumstances.

We claim that there are other reasoning techniques, like reasoning by analogy, which are faster, more powerful and easier to handle. The price is that they are *unstable*, that is they can produce false conclusions from true premises. On the other hand, we have argued that a reasoning system must be equipped to deal with inconsistencies independently of the inferences it carries out, simply because of the way interaction with the environment is bound to progress. Once we accept the need to deal with the problem (in a reasonable way), it is clear that we can take advantage of this and employ unstable inference techniques too.

The basic criteria that we envisage in dealing with unstable inferences are

that, firstly, they must be recognised as such; secondly, their performance must be controlled and their value modified according with it; thirdly, as in the case of inconsistencies arising from observed data, the recovery must imply some variation in the behaviour of the system, so that something has been learned from the contradiction.

We then propose the following description of a *reasonable* inference: it is an inference which directly produces useful results (thus, it must also be fast) while, when it fails, it indirectly provides additional useful information about related matters, such as the interpretation of the data or some assumptions behind the inference.

We do not use unstable reasoning techniques in this work because, as we have indicated already, a rather complex analysis and a partial solution to our problems can be devised without using these inference techniques. We have briefly mentioned them here, anyway, because they are central to the advanced solutions that we envisage for certain parts of the adaptive reasoning system. They also play a pivotal role in the direction of research, outlined in the next section, which forms the context for the present work.

2.4 A General Direction of Research

The central idea that characterises our direction of research is that reasoning is central in many activities which we try to simulate. Particularly important are the unstable reasoning techniques, for the reasons of speed, applicability and power already mentioned.

We have identified several such techniques, which of course have points in common with other methods elaborated by researchers in the field. The most important amongst these are: reasoning by analogy, reasoning by non-contradiction, reasoning by circumscription, reasoning by scientific induction, reasoning by effect to cause (sometimes called abduction), statistical reasoning.

The reasoning by analogy technique has been formally analysed [GRL86].

We have reason to believe that in fact all these techniques (and probably still others) are variations on the same theme, and that this theme could be analysed using homogeneity theory. An interpretation of these apparently very different techniques through one single theory would allow us to formulate a calculus of reasoning techniques on a very solid base.

While we consider the "consistency recovery" mechanism essential in order to deal reasonably and succesfully with unstable reasoning, on the other hand we believe that the reasoning techniques will be needed to enhance the power of the mechanism itself. This interaction is further discussed in Chapter 6.

The areas where we foresee interactions of great potential are: the natural language interface, the large knowledge-base management and the interpretation of sensorial data. In all these cases, there are very serious problems due to an excess of information, a requirement for very high speed of processing, a likelihood of information being ambiguous, incomplete or even inconsistent, and often a need to take action based on whatever information has been understood, retrieved or processed at that stage.

We also think that part of the structure used for the consistency recovery mechanism could provide a base, in a more sophisticated model, for the organisation of the reasoning process in general. The ideas of motivations expressed as performance trends to be approximated, of tension between opposite motivations, of cases prepared for and against some options, of trial runs could all have a role to play, expecially when they would be freed from some of the mechanical nature which of necessity afflicts them at the present stage.

Again, we expect homogeneity theory to play a very useful role, helping to individuate homogeneous trends and fractures which are, until now, beyond the power of the system. A typical case would be the decision not to consider past evidence (what we call the *lookback* parameter in Appendix A) because something has happened that suggests a fracture in that sequence of events. The role of the reasoning techniques would then

be to organize and connect the evidence. Homogeneity theory would provide the frame in which to analyse it and the mechanism inherited from the present model would be used as a test bench.

It has to be stressed again that all this represents only a direction of research, provided as a background against which to cast the present work.

Chapter 3

A Formal Approach to Homogeneity

3.1 Principles of Homogeneity

The concepts of homogeneity and of fractures in homogeneity are inextricably intertwined. This is a very natural consequence of our interpretation of the way in which knowledge is gathered (cf. Chapter 2). Nevertheless, for obvious reasons, we must begin our analysis examining one of these two fundamental parts: the limits that we meet from one side will provide exactly the starting point for exploring the other side, and vice versa.

We begin our discussion with the concept of homogeneity. First of all, in order to discuss homogeneity we must isolate a part of the world the behaviour of which we want to analyse, and that group of properties for which we expect this behaviour to be somehow homogeneous.

Once the sample we are interested in has been isolated, we must be able to look inside it to examine its behaviour under the chosen properties. Our interpretation of "looking inside" is to divide the sample into smaller groups which we hope will behave in the same way in relation to a given predicate. Thus a first approximation to homogeneity is the concept of similarity of behaviour of the sample under some screening.

In order to give a more precise meaning to homogeneity, we need to define more clearly what we mean by the behaviour of a property. We shall use the concept of density, which measures to what extent a property applies to a sample.

Within this interpretation, it follows that a sample which is not divided

under some screening is trivially homogeneous for all properties under that screening, which is then useless for considerations of homogeneity.

This, in turn, strongly suggests that the smaller the divisions of our sample generated by the screen, the more informative is the screening about the density of the sample under the property. We shall refer to the size of the divisions as the depth - the smaller the divisions, the greater the depth.

Another problem we must confront is that there are very many properties under which the sample is trivially completely homogeneous: for example, any property which does not apply to any member of the world. It is clear that the only useful way in which homogeneity can be understood and exploited is by finding a boundary, in the world, where the homogeneity breaks. We call this boundary a fracture.

It is clear, then, that a sample is homogeneous under some property if its density under the property is more or less the same for all divisions of the sample. Similarly, a fracture corresponds to a boundary across which there is a sudden change in density.

The useful properties are thus the ones for which there is a sample, highly homogeneous under the property and also for which there is a sufficiently sudden break in density across the boundary between the sample and its complement. We call these properties discriminants, because they discriminate between the sample and the rest of the world.

We now present the basis of a formal model for homogeneity.

3.2 Definitions for Homogeneity

In this Section we introduce the formal framework within which we can capture and discuss homogeneity. An informal treatment of these definitions can be found in Section 3.3.

Definition 3.2.1

If U is a set, and Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$, then:

(i) any $A \subseteq U$ such that $Q(A) = True$ is a $Q\text{-}set$;

(ii) the $Q\text{-}world\ in\ U$ is the set of all $Q\text{-}sets$ in U;

(iii) $Un(U,Q)$ is defined as $\cup Q\text{-}world\ in\ U$.

When no confusion can arise, we will abbreviate "$Q\text{-}world\ in\ U$" to "$Q\text{-}world$" and "$Un(U,Q)$" to "$Un(Q)$". We will also refer to the set U as the "$container$".

Definition 3.2.2

If U is a set, and Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$, then Q is a $screen$ for U if the Q-world in U is neither empty nor equal to $\{U\}$, and the empty set is not in the Q-world in U.

Definition 3.2.3

If A is a set and P is a predicate with type $A \rightarrow Bool$, then the $P\text{-}density$ of A, $D_P(A)$, is defined by:

$$D_P(A) = |\{x \in A : P(x)\}| / |A|.$$

Definition 3.2.4

If A is a set of real numbers then we define the $radius$ of A, $rad(A)$, by:

$$rad(A) = max(A) - min(A).$$

Definition 3.2.5

If U is a set, Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$ and P is a predicate with type $U \rightarrow Bool$, then the *entropy of Q under P in U*, $ent(U,P,Q)$, and the *homogeneity of Q under P in U*, $hom(U,P,Q)$, are defined as follows:

(i) $ent(U,P,Q) = rad(\{ D_P(A) : A \in Q\text{-world in } U \})$;

(ii) $hom(U,P,Q) = 1 - ent(U,P,Q)$.

3.3 An Explanation of the Basic Definitions for Homogeneity

We now compare each of the definitions made in Section 3.2 with the account of the principles of homogeneity given in Section 3.1.

Definitions 3.2.1 and 3.2.2 provide us with a formal basis for screening, using a predicate on the power-set of the world. The generalised union of the Q-world is the "sample" discussed in Section 3.1.

It is interesting to contrast the intuitive and the formal approaches to the definition of a sample: in the former (Section 3.1) we start from the sample considered as a unit and subsequently look inside it to analyse its homogeneity. In the latter approach we begin with the division (screening) and build the sample from it.

The reason for this difference is that in our intuitive discussion we assumed some kind of "feeling" for which pieces of the world could be sufficiently homogeneous to make good samples, and the screening was seen as a way of verifying it. On the other hand, in the formal approach we do not assume any previous knowledge about suitable samples, and the screening technique is the way by which we determine areas of high homogeneity.

In Definition 3.2.3 we formalise the concept of density as the proportion of

a set for which a property holds.

Finally, in Definitions 3.2.4 and 3.2.5, we introduce entropy and homogeneity. These are complementary measures in the range $[0,1]$. Entropy is the difference between maximum and minimum densities of the screening under the property. In this way, the greater the diversity of behaviour of the screening under the property, the greater the entropy, and consequently the lower the homogeneity.

3.4 Fractures in Homogeneity

Having formalised our understanding of homogeneity, we are now in a position to introduce the definitions of fractures and discriminants. We also prove two basic results.

An informal description of the following material can be found at the end of this Section.

Definition 3.4.1

If U is a set, Q is a predicate with type $\mathbb{P}(U) \rightarrow Bool$ and P is a predicate with type $U \rightarrow Bool$, then the *fracture value of Q under P in U*, $fr(U,P,Q)$, is defined by:

$$fr(U,P,Q) = D_P(Un(Q)) - D_P(U - Un(Q)).$$

Definition 3.4.2

If U is a set, Q is a predicate with type $\mathbb{P}(U) \rightarrow Bool$ and P is a predicate with type $U \rightarrow Bool$, then a *picture*, $Uf_x h_u(P,Q)$, is defined by:

$Uf_{x}h_{y}(P,Q)$ iff $|x| \leq |fr(U,P,Q)|$, $sign(x) = sign(fr(U,P,Q))$, and

$$y \leq hom(U,P,Q).$$

Definition 3.4.3

If U is a set, Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$ and P is a predicate with type $U \rightarrow Bool$, then P is an $f_{x}h_{y}$ *discriminant for* Q *in* U if $Uf_{x}h_{y}(P,Q)$.

P is a *positive discriminant* if $x > 0$ and is a *negative discriminant* if $x < 0$.

Theorem 3.4.1

If U is a set, Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$ and P is a predicate with type $U \rightarrow Bool$, then:

(i) $D_{P}(A) = 1 - D_{\sim P}(A)$;

(ii) $fr(U,P,Q) = -fr(U,\sim P,Q)$;

(iii) $hom(U,P,Q) = hom(U,\sim P,Q)$.

Proof:

(i) $D_{P}(A) = |\{x \in A : P(x)\}|/|A|$, by Definition 3.2.3,

$= |A - \{x \in A : \sim P(x)\}|/|A|$

$= 1 - |\{x \in A : \sim P(x)\}|/|A|$

$= 1 - D_{\sim P}(A)$, by Definition 3.2.3.

(ii) $fr(U,P,Q) = D_p(Un(Q)) - D_p(U - Un(Q))$, by Definition 3.4.1,

$= (1 - D_{\sim p}(Un(Q))) - (1 - D_{\sim p}(U - Un(Q)))$, by (i),

$= -(D_{\sim p}(Un(Q)) - D_{\sim p}(U - Un(Q)))$

$= -fr(U,\sim P,Q)$, by Definition 3.4.1.

(iii) $hom(U,P,Q) = 1 - rad(\{D_p(A): A \in Q\text{-world in } U\})$, by Definition 3.2.5,

$= 1 - rad(\{1 - D_{\sim p}(A): A \in Q\text{-world in } U\})$, by (i).

$rad(A) = max(A) - min(A)$, by Definition 3.2.4,

$= (1 - min(\{1 - x: x \in A\})) - (1 - max(\{1 - x: x \in A\}))$, by properties of min and max,

$= rad(\{1 - x: x \in A\})$.

Therefore:

$hom(U,P,Q) = 1 - rad(\{D_{\sim p}(A): A \in Q\text{-world in } U\})$

$= hom(U,\sim P,Q)$, by Definition 3.2.5. □

Proposition 3.4.1

If U is a set, Q is a predicate with type $\mathbb{P}(U) \rightarrow Bool$ and P is a negative $f_s h_q$ discriminant for Q in U, then $\sim P$ is a positive $f_{\sim s} h_q$ discriminant for Q in U.

Proof:

Since P is a negative discriminant, $x < 0$.

By Definition 3.4.2, $sign(x) = sign(fr(U,P,Q))$, so $fr(U,P,Q) < 0$.

By Proposition 3.4.1(ii), $fr(U,P,Q) = -fr(U,\sim P,Q)$, so $fr(U,\sim P,Q) > 0$.

Therefore $sign(-x) = sign(fr(U,\sim P,Q))$, and $|x| \leq |fr(U,\sim P,Q)|$, by above and Definition 3.4.2.

By Definition 3.4.2, $y \leq hom(U,P,Q)$.

Therefore $y \leq hom(U,\sim P,Q)$, by Theorem 3.4.1(iii). The result follows by Definition 3.4.2. □

In Definition 3.4.1 we introduce the fracture value of a screening under a property. This is the difference in density between the sample and its complement in the world. We define a standard, compact notation for the concepts of homogeneity, fractures and discriminants in Definitions 3.4.2 and 3.4.3.

The first result we prove, in Theorem 3.4.1, describes the behaviour of the functions D_p, fr and hom under negation of the property used as a discriminant. The interesting parts of this Proposition are (ii) and (iii). Here we see that fr is asymmetric under negation of the discriminant, while hom is symmetric.

Using these results, we are able to prove, in Proposition 3.4.1, that the negation of a negative discriminant is a positive discriminant for the same screening.

3.5 Classes of Predicates: Properties and Screens

We now generalise the conditions of the previous definitions to allow us to use classes of predicates. An explanation of the purpose of the following definitions and results can be found in Section 3.6.

Definition 3.5.1

If U is a set, Q is a predicate with type $\mathbf{P}(U) \rightarrow Bool$ and $\mathcal{P}$ is a set of predicates each with type $U \rightarrow Bool$, then $\mathcal{P}$ *defines* Q *in* U *within* $f_x h_y$ if:

$$\forall A \subseteq U.\ \exists \mathcal{P}' \subseteq \mathcal{P}.\ A f_x h_y(\wedge \mathcal{P}', Q).$$

Proposition 3.5.1

If $U f_x h_y(P,Q)$ and $D_R(Un(Q)) = 1$ then $U f_x h_y(P \wedge R, Q)$.

Proof:

$\forall A \subseteq Un(Q).\quad D_P(A) = D_{P \wedge R}(A)\ \ldots\ (*)$ by hypothesis.

$x \leq fr(U,P,Q) = D_P(Un(Q)) - D_P(U - Un(Q))$, by Definition 3.4.1,

$= D_{P \wedge R}(Un(Q)) - |\{x: x \in (U - Un(Q)),\ P(x)\}| / |U - Un(Q)|$,
by $(*)$ and Definition 3.2.3,

$\leq D_{P \wedge R}(Un(Q)) - |\{x: x \in (U - Un(Q)),\ P(x) \wedge R(x)\}| / |U - Un(Q)|$,

$\leq D_{P \wedge R}(Un(Q)) - D_{P \wedge R}(U - Un(Q))$, by Definition 3.2.3.

$y \leq hom(U,P,Q) = 1 - rad(\{ D_P(A) : A \in Q\text{-world in } U \})$, by Definition 3.2.5,

$= 1 - rad(\{ D_{P \wedge R}(A) : A \in Q\text{-world in } U \})$, by (*),

$= hom(U, P \wedge R, Q)$, by Definition 3.2.5. □

Corollary 3.5.1

If $\mathcal{P}$ defines Q in U within $f_x h_y$, and $\forall P \in \mathcal{P}\ D_P(Un(Q)) = 1$ then

$$\forall A \subseteq U.\ A f_x h_y(\wedge \mathcal{P}, Q).$$

Proof:

If $\mathcal{P}$ defines Q in U within $f_x h_y$ then $\forall A \subseteq U.\ \exists \mathcal{P}' \subseteq \mathcal{P}.\ A f_x h_y(\wedge \mathcal{P}', Q)$, by Definition 3.5.1. By Proposition 3.5.1, $A f_x h_y((\wedge \mathcal{P}') \wedge P, Q)$ for all P such that $D_P(Un(Q)) = 1$.

Therefore, by hypothesis, $\forall A \subseteq U.\ A f_x h_y(\wedge \mathcal{P}, Q)$. □

Definition 3.5.2

If U is a set, $\mathcal{Q}$ is a set of predicates each with type $\mathbf{P}(U) \rightarrow Bool$ and $\mathcal{P}$ is a set of predicates each with type $U \rightarrow Bool$, then

$$U f_x h_y(\mathcal{P}, \mathcal{Q}) \text{ iff } \forall Q \in \mathcal{Q}.\ \exists \mathcal{P}' \subseteq \mathcal{P}.\ \mathcal{P}' \text{ defines } Q \text{ in } U \text{ within } f_x h_y.$$

Proposition 3.5.2

If $U f_x h_y(\mathcal{P}, \mathcal{Q})$ and $V f_z h_t(\mathcal{P}, \mathcal{Q})$, then $(U \cap V) f_c h_d(\mathcal{P}, \mathcal{Q})$, where $(c,d) = (x,y)$ or $(c,d) = (z,t)$.

Proof:

By Definition 3.5.2, $Uf_xh_y(\mathcal{P},\mathbb{Q})$ implies:

$\forall Q \in \mathbb{Q}.\ \exists \mathcal{P}' \subseteq \mathcal{P}.\ \mathcal{P}'$ defines Q in U within f_xh_y.

By Definition 3.5.1, $\mathcal{P}'$ defines Q in U within f_xh_y implies:

$\forall A \subseteq U.\ \exists \mathcal{P}'' \subseteq \mathcal{P}'.\ Af_xh_y(\wedge\mathcal{P}'',Q)$.

Since $(U \cap V) \subseteq U$, it follows, by transitivity of set inclusion, that

$\forall A \subseteq (U \cap V).\ \exists \mathcal{P}'' \subseteq \mathcal{P}'.\ (U \cap V)f_xh_y(\wedge\mathcal{P}'',Q)$.

Therefore $\mathcal{P}'$ defines Q in $(U \cap V)$ within f_xh_y, by Definition 3.5.1, and, by Definition 3.5.2, it follows that $(U \cap V)f_xh_y(\mathcal{P},\mathbb{Q})$.

The same argument holds for $(c,d) = (s,t)$. □

In the case when $x \geq s$ and $y \geq t$, we choose $(c,d) = (x,y)$. Similarly when the total order is inverted, we choose $(c,d) = (s,t)$.

Proposition 3.5.3

If $Uf_xh_y(\mathcal{P},Q)$ and $A \in Q$-world in U, then $\forall P \in \mathcal{P}.\ hom(A,P,Q) \geq y$.

Proof:

$hom(A,P,Q) = 1 - rad(\{D_p(B): B \in Q\text{-world in } A\})$, by Definition 3.2.5,

$= 1 - rad(\{D_p(B): B \in \{C \subseteq A: Q(C) = True\}\})$, by Definition 3.2.1,

$\geq 1 - rad(\{D_p(B): B \in \{C \subseteq U: Q(C) = True\}\})$, by Definition 3.2.4,

$\geq hom(U,P,Q)$, by Definition 3.2.5,

$\geq y$, by Definition 3.4.2. □

Proposition 3.5.4

If $Uf_x h_x(P,Q)$ and $Uf_s h_s(P,Q')$ then $hom(U,P,Q \wedge Q') \geq max(\{y,t\})$.

Proof:

Without loss of generality, assume $y \geq t$.

$hom(U,P,Q \wedge Q') = 1 - rad(\{D_p(B): B \in (Q \wedge Q')\text{-world in } U\})$,
by Definition 3.2.5,

$= 1 - rad(\{D_p(B): B \in \{C \subseteq U: Q(C) \wedge Q'(C) = True\}\})$,
by Definition 3.2.1,

$\geq 1 - rad(\{D_p(B): B \in \{C \subseteq U: Q(C) = True\}\})$,
by Definition 3.2.4,

$\geq hom(U,P,Q)$, by Definition 3.2.5,

$\geq y$, by Definition 3.4.2. □

3.6 The Use of Classes of Predicates

In Definition 3.5.1 a set of predicates is used to "define" a screen. This set of predicates includes properties whose conjunction provides a good discriminant between some part of the Q-world and some part of the rest of the world.

We then prove Proposition 3.5.1 and Corollary 3.5.1 which together show that conjoining properties that apply to the whole of $Un(Q)$ with any other property does not alter the homogeneity and can only improve the fracture value of the screen under the latter property.

This result tells us that if we have a list of properties which apply universally to a set, we can take their conjunction as the *definition* of the set, without losing any of the power of the individual conjuncts. This is particularly useful when constructing and using "dictionary" definitions of various concepts, since dictionary definitions will be treated as conjunctions of characteristic properties, chosen, nevertheless, for their individual power of discrimination. It is clearly esssential that this power is not compromised by forming this conjunction.

Next we make Definition 3.5.2 which generalises Definition 3.5.1 to a set of screens. This will form a building block in our construction of "world divisions" in Section 3.8.

In Propositions 3.5.2, 3.5.3 and 3.5.4 we consider the effects of various operations on homogeneity and fracture values. Proposition 3.5.2 concentrates on the intersection of containers (cf. remark following Definition 3.2.1).

Proposition 3.5.3 establishes a lower bound on the homogeneity of individual Q-sets, which will be useful when we refine the treatment of an entire Q-world in order to concentrate on particular subsets. It is clear that it is very important to be sure that all the properties of the Q-world are preserved in this transition. In Proposition 3.5.4 we show a condition on the homogeneity of the conjunction of two screens under a given property.

3.7 Homogeneity and the Class of Size-Flat Screens

We now proceed to define a class of screens which play a special role in our later use of homogeneity and fracture values. A fundamental result is proved about this class of screens.

An informal rendering of the following material can be found in Sections 3.8 and 3.9.

Definition 3.7.1

If U is a set, and Q is a screen for U then the *depth of Q in U* is defined as follows:

$$depth(U,Q) = 1 - (min(\{|A| - 1 : Q(A),\ A \subseteq U\}) / |Un(Q)| - 1).$$

Definition 3.7.2

If U is a set, $1 \leq k \leq |Un(Q)|$, and Q is a screen for U then the *k-slice of Q in U* is defined as follows:

$$k\text{-}slice(U,Q) = \{ A : A \subseteq U, Q(A), |A| = k \}.$$

Definition 3.7.3

If U is a set and Q is a screen for U then Q is a *size-flat* screen in U if $Q(A)$ implies $Q(B)$ for all B in the $|A|$-slice of Q in U.

Definition 3.7.4

If U is a set, Q is a screen and P is a predicate with type $U \rightarrow Bool$, then the set of *exceptions to P under Q* is defined by:

$$ex(U,P,Q) = \{ x : \sim P(x), x \in Un(Q) \}.$$

Theorem 3.7.1

If U is a set, Q is a size-flat screen of depth d and P is a predicate with type $U \rightarrow Bool$, then:

$$hom(U,P,Q) = max(\{0,\ 1 - n/k,\ 1 - (|Un(Q)| - n)/k,\ 2 - |Un(Q)|/k\}),$$

where $n = |ex(U,P,Q)|$ and $k = d + (1-d)|Un(Q)|$.

(n is thus the number of exceptions to P under Q, and k is the size of the smallest set for which Q holds).

Proof:

$$hom(U,P,Q) = 1 - rad(\{D_p(A) : A \in Q\text{-world in } U\}), \text{ by Definition 3.2.5,}$$

$$= 1 - (max(\{D_p(A) : A \in Q\text{-world in } U\}) - min(\{D_p(A) : A \in Q\text{-world in } U\})), \text{ by Definition 3.2.4,}$$

$$= 1 - max(\{D_p(A) : A \in Q\text{-world in } U\}) + min(\{D_p(A) : A \in Q\text{-world in } U\}),$$

$$= (1 - max(\{|\{x : P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\})) + (1 - max(\{|\{x : \neg P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\})),$$

by Theorem 3.4.1(i) and Definition 3.2.3.

Since Q is size-flat, then when A is a Q-set, all sets of size $|A|$ are Q-sets.

Therefore, the value of $|\{x : \neg P(x), x \in A\}|/|A|$ is maximised for a particular value of $|A|$ when $ex(U,P,Q) \subseteq A$, or, if $|ex(U,P,Q)| > |A|$, when $A \subseteq ex(U,P,Q)$.

So: $max(\{|\{x: \sim P(x), x \in A\}|/v : A \in Q\text{-world in } U \text{ and } |A| = v\})$

$= |ex(U,P,Q)|/v$, if $|ex(U,P,Q)| \leq v$,

$= 1$, if $|ex(U,P,Q)| > v$.

Thus:

$max(\{|\{x: \sim P(x), x \in A\}|/v : A \in Q\text{-world in } U \text{ and } |A| = v\})$

$= |ex(U,P,Q)|/v$, if $|ex(U,P,Q)|/v \leq 1$,

$= 1$, if $|ex(U,P,Q)|/v > 1$.

This implies that:

$max(\{|\{x: \sim P(x), x \in A\}|/v : A \in Q\text{-world in } U \text{ and } |A| = v\})$

$= min(\{|ex(U,P,Q)|/v, 1\})$,

$= min(\{n/v, 1\})$,

simplifying the above expression.

Therefore:

$hom(U,P,Q) = (1 - max(\{|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$
$+ (1 - max(\{min(\{n/v, 1\}) : A \in Q\text{-world in } U \text{ and } |A| = v\}))$,
using the above expression.

Since:

$max(\{min(\{e, 1\}) : R(e)\}) = min(\{max(\{e : R(e)\}), 1\})$,

it follows that:

$$hom(U,P,Q) = (1 - max(\{\,|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$$
$$+ (1 - min(\{max(\{n/v : A \in Q\text{-world in } U \text{ and } |A| = v\}), 1\})),$$

$$= (1 - max(\{\,|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$$
$$+ (1 - min(\{n / min(\{|A| : A \in Q\text{-world in } U\}), 1\})).$$

Now:

$d = depth(U,Q)$, by hypothesis,

$= 1 - (min(\{|A| - 1 : Q(A), A \subseteq U\})/|Un(Q)| - 1)$, by Definition 3.7.1,

$= 1 - (min(\{|A| : A \in Q\text{-world in } U\}) - 1)/(|Un(Q)| - 1)$.

So $min(\{|A| : A \in Q\text{-world in } U\}) = d + (1 - d)|Un(Q)|$,

$= k$, by hypothesis.

Therefore:

$$hom(U,P,Q) = (1 - max(\{\,|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$$
$$+ (1 - min(\{n/k, 1\})).$$

So:

$$hom(U,P,Q) = (1 - max(\{\,|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$$
$$+ (1 - min(\{\,|\{x: \sim P(x), x \in Un(Q)\}|/k, 1\})),$$

$$= (1 - max(\{\,|\{x: P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}))$$
$$+ max(\{0, 1 - n/k\}),$$ by hypothesis.

A similar argument can be used to simplify the first part of the expression, as follows.

If:

$$|\{x : P(x), x \in Un(Q)\}| \geq k$$

then:

$$max(\{|\{x : P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}) = 1,$$

and if:

$$|\{x : P(x), x \in Un(Q)\}| < k$$

then:

$$max(\{|\{x : P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}) = (|Un(Q)| - n)/k.$$

Now, $|\{x : P(x), x \in Un(Q)\}| = |Un(Q)| - n,$

so the following two equations hold:

$$|\{x : P(x), x \in Un(Q)\}| \geq k \text{ implies } (|Un(Q)| - n)/k \geq 1,$$

$$|\{x : P(x), x \in Un(Q)\}| < k \text{ implies } (|Un(Q)| - n)/k < 1.$$

Thus:

$$max(\{|\{x : P(x), x \in A\}|/|A| : A \in Q\text{-world in } U\}) = min(\{1, (|Un(Q)| - n)/k\}),$$

which implies that:

$$hom(U,P,Q) = 1 - min(\{1, (|Un(Q)| - n)/k\}) + max(\{0, 1 - n/k\}),$$

$$= max(\{0, 1 - (|Un(Q)| - n)/k\}) + max(\{0, 1 - n/k\}),$$

$$= max(\{0, 1 - n/k, 1 - (|Un(Q)| - n)/k, 2 - |Un(Q)|/k\}). \square$$

Corollary 3.7.1

If U is a set, Q is a size-flat screen of depth d, P is a predicate with type $U \rightarrow Bool$, and $2k - |Un(Q)| > k - n > 0$, where $n = |ex(U,P,Q)|$ and $k = d + (1 - d)|Un(Q)|$, then:

$$hom(U,P,Q) = 2 - |Un(Q)| / k.$$

Proof:

If $2k - |Un(Q)| > k - n > 0$ then it is clear that:

$$2 - |Un(Q)| / k > 1 - n/k > 0.$$

Also, it follows that:

$$k > |Un(Q)| - n > |Un(Q)| - k,$$

so that:

$$2 - |Un(Q)| / k > 1 - (|Un(Q)| - n)/k > 0.$$

By Theorem 3.7.1,

$$hom(U,P,Q) = max(\{0,\ 1 - n/k,\ 1 - (|Un(Q)| - n)/k,\ 2 - |Un(Q)| / k\}),$$

and therefore $hom(U,P,Q) = 2 - |Un(Q)| / k$. □

3.8 A Discussion about Classes of Screens

In Definition 3.7.1 we present an important measure which will allow us to consider the behaviour of screens: the measure of depth of a screen, which is a quantification of how far into the Q-world it is known a particular degree of homogeneity extends. This is precisely the concept of depth

referred to in Section 3.1.

Next we define a k-slice, which is a collection of all the Q-sets of the same size (Defn. 3.7.2). We then use this concept to introduce size-flat screens (Defn 3.7.3). The classification of screens is an important task in order to look for common properties of sets of screens. Size-flat screens are one such class, falling into the important category of "unintelligent" screens.

Unintelligent screens are characterised by using conditions which can be checked using only properties of the candidate Q-set itself, that is, without reference to relations with other elements of the container outside the candidate Q-set. For example, the membership of the candidate Q-set in the Q-world, for a size-flat screen, can be checked by simply counting its elements. In contrast, an intelligent screen is one which uses information outside the candidate Q-set.

A very important type of intelligent screen is the "metric screen". That is, a screen using a metric imposed on the elements of some part of the container including the candidate Q-set. A typical application of such a metric would be to use neighbourhoods of elements as a screen. A different type of intelligent screen is the "statistical screen", which uses some form of random sampling to deliver Q-sets.

It might at first appear strange to be interested in unintelligent screens when the use of intelligent screens is so widespread and successful. The reason is that unintelligent screens, when applicable, offer a far greater power and security in computing the degree of homogeneity of the Q-world. On the other hand, the reliability of intelligent screens, when computing the degree of homogeneity of a Q-world, varies greatly with the adequacy of the particular conditions on which the screen is based.

Finally, Definition 3.7.4 identifies a set of exceptions to a given property under some screen. This set is used immediately in Theorem 3.7.1, which we discuss in Section 3.9.

3.9 Analysis of an Important Result about Size-Flat Screens

Theorem 3.7.1 provides a significant expression for the homogeneity of a size-flat screen under a given property, in terms of the depth of the screen, d, and the size of the exception set, n. Theorem 3.7.1 is important for several reasons, which we examine separately below.

The definition of $hom(U,P,Q)$ (Defn. 3.2.5) implies that the naive approach to computing its value is to examine the value of the P-density of each Q-set, which in turn, given a set $Un(Q)$, requires us to check every subset of $Un(Q)$ to depth d. The number of subsets that must be checked is then exponential in the size of $Un(Q)$, which is clearly extremely expensive. Theorem 3.7.1 tells us, however, that the only information that is required to compute $hom(U,P,Q)$ is the size of the exception set in $Un(Q)$, which can be computed in $O(|Un(Q)|)$ steps.

Corollary 3.7.1 highlights an interesting result contained within the expression obtained in Theorem 3.7.1, which is that under certain conditions, the value of $hom(U,P,Q)$ is given by a term which is independent of the number of exceptions to the property P. This is unexpected and to some extent counter-intuitive. When the depth of the screen is 0, it is trivially the case that $hom(U,P,Q) = 1$, but it is interesting to observe that this is the limit value of that term in the expression of Theorem 3.7.1 which is independent of the number of exceptions.

The result of Corollary 3.7.1 can be interpreted as saying that, when either of the terms $1 - n/k$ or $1 - (|Un(Q)| - n)/k$ lies between $2k - |Un(Q)|$ and 0, then the value of $hom(U,P,Q)$ is precisely that of the term $2k - |Un(Q)|$.

We emphasise that the expression given in Theorem 3.7.1 is an equality, and that the value of $hom(U,P,Q)$ is bound to be exactly equal to one of only four precisely defined numerical values. Not only that, but homogeneity is obviously a commodity that must be considered to be highly desirable, yet despite taking the worst case analysis throughout the proof of the Theorem the result is still in the form of a maximum of several values.

Finally, since size-flat screens are absolutely the least intelligent screens, the expression in Theorem 3.7.1 is in fact a lower bound for all screens, intelligent or not. This vindicates our study of size-flat screens , and re-emphasises the power of the Theorem.

3.10 Assumptions of Homogeneity

So far, we have considered homogeneity and fractures as exactly computable values, and we have also proved that these computations are feasible (Theorem 3.7.1). It must be noted now that these computations are possible only when the test of applicability of the property to each element in the $Un(Q)$ can actually be made. This in turn requires knowing all these elements, having access to them and possessing an accurate test. On top of all this, there is the problem of things changing over time. It is clear that for most worlds, in our everyday or scientific experience, homogeneity and fractures values will never be really known. We cannot even talk about approximation as a general rule, even if, in many cases, that is exactly what takes place, since often enough completely wrong estimates can be and are made.

This situation is not an obstacle to the smooth application of our model, just as it is not an obstacle, in real life, to everyday or scientific reasoning. This is because our knowledge of the homogeneity of the world is an even higher goal than the knowledge of relationships between specific objects in the world. Thus, each assumption of homogeneity can be considered as a hypothesis about the homogeneity of the world, as well as a provisional tool with which we can analyse the world.

In Section 5 we will examine what assumptions of homogeneity are made by different classes of reasoning systems, and how explicit they are.

3.11 The Problem of Individuals

A fundamental primitive in any language is the concept of individuals. That is, the atomic constants about which the language is designed and whose relationships with each other are assigned truth values. However, there is a problem in applying these representations, since the entities which are most usefully and most naturally considered as individuals are frequently, in reality, a collection of much smaller entities. Identifying formally when the division of individuals into smaller entities should stop has long been considered a difficult problem.

We claim that, through the ideas introduced in the preceeding sections, these "primitives" can actually be defined from more basic concepts, and when this definition is matched with our interpretation of individuals in the world, the problem of individuals which are really collections of smaller objects is dealt with in a very natural way.

We assume that there is, given, a "world" of sensorial data (real or simulated), which is the set we consider as a container (cf. remark following Defn. 3.2.1), and also *motivation*. We will, for the purposes of this work, consider motivation as a source of requirements that lead naturally to the investigation of certain properties in a hierarchy of importance.

Motivation is a very important subject for study, and we believe that it has a structure which can be formalised and analysed. However, we consider this to be outside the scope of this work.

Since motivation will not be treated here and yet we claim that the first properties we explore are in some way suggested by the motivation, we will give an example of the way we believe such a property might be provided by motivation in human infants.

The motivation in human infants includes such motives as "explore the world", "find things to eat" and "reproduce pleasurable experiences". Together these motives produce the commonly observed behaviour of children

whereby they put every newly discovered object into their mouths. The property that is used to discriminate between objects is "tastes good" (in a wide sense). This property can be used to divide the world into two groups – those that are nice to put in the mouth and those that are not.

It is important to note that, while we consider individuals to be logical building blocks, we observe that they can be constructed only through the application of some properties. That is, statements about the relationship between properties and individuals are intimately connected to the exploration of the world itself.

It is from our initial impressions of the environment that we build the first individuals and equip ourselves for exploring the world. In the following sections we show how the construction of these individuals and the way we attribute properties to them are very different processes from those that are usually conceived.

3.12 Individuals and Appearances

The following formal definitions are discussed in Section 3.13.

Definition 3.12.1

If U is a set and $\mathcal{P}$ is a set of predicates each with type $U \rightarrow Bool$, then a *world division of* U, with depth d, fracture bounded below by x and homogeneity bounded below by y is a set $\{(A_i, Q_i)\}_{i=1}^{k}$ satisfying:

(i) $\bigcup_{i=1}^{k} A_i \subseteq U$;

(ii) $\forall i \neq j \,.\, A_i \cap A_j = \emptyset$;

(iii) each Q_i satisfies:

(a) $Un(Q_i) = A_i$;

(b) Q_i is a size-flat screen with depth d;

(c) $Uf_x h_y(\mathcal{P}, \{Q_i\}_{i=1}^n)$.

Definition 3.12.2

If U is a set and $\{(A_i, Q_i)\}_{i=1}^n$ is a world division with depth, d, and with fracture value and homogeneity bounded below by f and h respectively, then each Q_i is an *individual*.

Note that the values d, f and h are local constants, which will have been provided by functions driven by motivation.

Definition 3.12.3

An *appearance*, $U_{(P,S)} f_x h_y(R,Q)$, for Q an individual, is defined by:

$U_{(P,S)} f_x h_y(R,Q)$ iff $Uf_x h_y(P,Q)$

and S is a Q-set for which $D_R(S) = D_P(S)$.

3.13 A New Understanding of Individuals

In Definition 3.12.1 we introduce world divisions. A world division is a partition (Defn. 3.12.1(ii)) of some subset of the container (Defn. 3.12.1(i)), made in such a way that there is a screen associated with each set in the partition. Each of these screens satisfies certain conditions: firstly, (Defn. 3.12.1(iii)(a)) they isolate precisely that set in the partition with which

they are associated. Secondly, (Defn. 3.12.1(iii)(b)) each screen is size-flat and has a predetermined depth. Finally, (Defn. 3.12.1(iii)(c)) each screen is defined (in the sense of Definition 3.5.1) by some subset of a pre-specified set of properties.

The significance of world divisions is in their application to the determination of individuals. By this we mean a categorisation of the container into a small number of roughly equal (in importance) parts, according to some identified set of interesting and useful properties.

It will be noted that many values in Definition 3.12.1 are left variable: in particular, the number of divisions, k, the specified depth, d, and the specified bound on fracture values and homogeneity, x and y. Nor do we specify the relationship between the union of the defining sets of properties, $\mathcal{P}_1$, and the initial set of properties, $\mathcal{P}$. However, all these values are to some extent interdependent, even if frequently there will be good reasons to optimise one or more of them. For example, if the relative importance of each property in $\mathcal{P}$ is more or less equal, then a criterion could be to maximise the number of properties that occur in some $\mathcal{P}_1$. This, of course, is subject to further constraints: firstly, the fracture value is of paramount importance - if this is not sufficiently high, then there is confusion at the boundaries between divisions. Another constraint is represented by the value of k, which must usually be small.

We now consider the process by which individuals are constructed.

Properties and parameters satisfying the constraints outlined above are provided by the motivation discussed in Section 3.11. Using the values provided by motivation and the first small world of experience, world divisions are constructed. The resulting size-flat screens are checked by examining each Q-set. In these small worlds fracture values and homogeneity are exact lower bounds.

The size-flat screens thus created are the individuals (Defn. 3.12.2).

It might at first appear that the intuitively appropriate candidates for individuals are the elements of the container, or world. A closer

examination of this possibility reveals its flaw: the world we start with is just a collection of sensorial data, with no inherent structure or order, so there are no defined objects about which we can speak or reason. (We underline again that by "sensorial data" we also intend simulated sensorial data, so that, for instance, textual input to a reasoning machine could equally be considered "sensorial data").

In fact, to be considered as an individual, it is essential that an object be identifiable. That is, we must be able to distinguish it uniquely from the rest of the world.

To clarify this, take the example of a basket of apples: it is tempting to believe that we can think of each apple as an individual, but the reality is that the individual is the collection of apples within the basket. We can distinguish the collection of apples in the basket from all other collections of apples, because we have the basket, and we have that collection of apples inside it. We cannot distinguish any single apple from any other, within the basket, when the only screen we can use is whether they are in the basket. If we choose to distinguish the apples in the basket by their position then we have created another screen, induced by the property of position, and thus a new individual - the newly distinguishable apples.

We have already indicated the problem of deciding where to draw the limit in recognising an individual consists of a collection of smaller individuals. Actually, the decision is simply a question of motivation - once the individuals have been recognised for a given property then there is no need (or motive) for further subdivision to find smaller individuals. It is the motivation that indicates the properties that are of interest and these in turn are used to define world divisions, which include the size-flat screens defined by the original properties and these are precisely the individuals we recognise. Notice that the motivation does not decide the individuals, but the properties that must be used to define them and these are decided without knowledge of which individuals they will create.

It is important that an individual can be distinguished from the rest of the world by a discriminant property (Defn. 3.12.3) - if there is no good discriminant then the object is not really an individual, because there is

some other part of the world from which the object cannot be distinguished. A good discriminant is one which induces a screen with a high fracture value from the rest of the world.

The elements of $Un(Q)$ are not individuals, because they are not distinguishable, nor are the Q-sets distinguishable, unless we make an explicit label to identify each piece of sensorial data, or each Q-set, which creates new properties (labels) and thus induces new screens and therefore new individuals.

It follows from this discussion, that the only entity which remains that can be considered as an individual is the screen, which is the justification for the definition we have made.

When we have divided the world into individuals, we want to be able to examine new properties of the individuals. Unfortunately, it is frequently impossible, impractical or too time consuming to test the behaviour of every Q-set under a given property. Therefore, we usually make assumptions about the behaviour of the whole of a Q-world based on an examination of the behaviour of a single Q-set, together with a knowledge of the homogeneity of the individual under its defining properties in the world division inducing it.

We put together all the information on which the assumption is based in an appearance (Defn. 3.12.3). When we use an appearance we usually choose to hide the assumptions and treat the formula as if it were a picture (Defn. 3.4.2).

It is only when the assumptions prove false, and the appearance is in contradiction with an observation of the world, that, if a continued growth of understanding is to be achieved, the assumptions must be examined once again and revised. This in turn requires that the contradiction is recognised within a logical system strong enough to formalise the assumptions which have been made, so that they can be re-examined. (As we have already observed, this revision can in itself be a gain in knowledge). In Chapters 5 and 6 we formally define some necessary properties that a reasoning system must satisfy in order to allow for recognition and examination of the

assumptions involved.

This concludes our treatment of individuals.

We have now formalized part of the general philosophy about assumptions and needs for regularity that we have outlined in Section 2.5. Part of the material of this chapter is used in Chapter 5, mainly the definitions of screens. Other results, particularly Theorem 3.7.1, are profitably applied later in Chapter 6. We believe, anyway, that the value of the theory, as already pointed out in Section 2.6, goes far beyond its use in the present context.

Chapter 4

A Formal Analysis of Five Classes of Reasoning Systems

4.1 Introduction

In Chapter 3 we developed a formal theory of homogeneity and used it to consider one of the most basic problems in understanding and reasoning: recognising and discriminating the different individuals of the world.

It is evident that in order to progress from the first steps in discerning the parts of the world, it is necessary to create a suitable tool with which to manipulate and extend knowledge. For centuries logics have been considered the most rigorous instruments for both describing rational reasoning and actually exercising it, and therefore offer themselves as natural candidates. We are prepared to subscribe to this point of view, provided that a certain extension to the definition of logic is allowed, in order to admit recent entries in the field, like non monotonic logics, or certain kinds of fuzzy logic. On the other hand, it is our contention that even this extended definition is not enough for the needs of A.I., and we propose as an alternative, more restrictive concept, our definition of a *reasoning system*.

The standard model-theoretic definition of a logic is a triple containing a formal language, a set of transformation rules and a set of possible mappings from the well-formed expressions of the language to a set of values. This usually consists of a set of individuals, whose names are the constants of the language, a set of relations between the individuals, which is expressed in the connectives and predicates of the language, and a set of

classes (at least two), which are usually understood to represent "truth" and "falsity", but which in fact can be interpreted in several different ways, from simple flags (e.g. in boolean algebras) to complex epistemological, temporal or deontic values.

While not every interpretation can be applied to these classes, there is always more than one possible. We will refer to them as "semantic classes". The important thing is that there must be one special subclass on which the transformation rules apply. We call this subclass "*T*".

The transformation rules themselves are defined so that the result of any transformation is guaranteed to remain in that same special subclass, *T*. This condition is often called "validity". If the logic is presented in axiomatic form, the axioms can be completely expressed in a richer proof theory (the transformation rules), while the opposite is not true. Therefore, from now on we will drop any reference to axioms.

In order to be applied, a logic also requires a universe of discourse, or theory. Essentially, this is a set of well-formed sentences of the language which are mapped to *T*, and on which the transformation rules can then operate. In the more restrictive definition of a theory, only sentences which can be derived from the original ones through transformations, and their negations, belong to the theory. This view gives rise to some paradoxical conclusions: for example, it follows that the continuum hypothesis does not belong to the axiomatic set theory! A more relaxed approach requires only that the constants and the predicates named in the sentence under scrutiny have all appeared already in some of the original sentences.

If any sentence of the theory maps to more than one semantic value at any particular moment, the theory is said to be *inconsistent*, which means that the transformation rules could then map every sentence to the special subclass, making any useful interpretation impossible. This point underlines our contention that reasoning is about discerning, and that a formal tool which does not help to discern is worse than useless. It is for this reason that we regard logic as applied, not pure mathematics, despite its abstract character and foundational use, since an abstract algebra could be interesting on, say, aesthetic grounds even if it does not model any piece of reality,

while a logic in that condition would clearly be meaningless. It is also clear that the assignment of sentences to the special subclass T is the central part of the mechanism, since the transformation rules operate only on sentences in T, and deliver more members of T.

The important point is that, in the standard definition of a logic, the sentences mapped to T cannot ever be moved to something else, preserving the same theory. As a consequence, if a sentence which was mapped to T is then mapped to something else, it ends up being mapped to two classes, and the theory is considered inconsistent and hence collapsed. When only two semantic classes are used, as in classical logic, this clearly means that no map can be changed; other logics use more than two semantic classes (they are called many-valued logics), in which cases some changes may be possible, but not away from T.

This usually happens because the standard interpretation of T is Truth, in an *ontological* sense, while the additional values tend to be *epistemological*, such as "not known", "not understood", "believed (but not necessarily true)" and so on, and when the posssibility of reasoning at an ontological level is admitted, clearly the epistemological values are subordinate to the ontological one. Our position in this respect is that it is possible and very practical to use ontological concepts on *local* theories, that is from inside the system, but that the only possible interpretation on the *global* theory, that is as seen from outside the system during an interaction, is the epistemological one.

The extension we believe is needed for the concept of logic formalises a tendency which has been manifest for some years, expecially, but not only, amongst logicians working in the field of A.I. It is proposed that the above privilege of the T class be eliminated, so that sentences can be moved away from it without the theory necessarily collapsing. Our reason for proposing this change in the definition of logic, instead of pushing this feature, together with the other ones we are going to propose, into our own concept of reasoning system, is to harmonise the (expanded) use of the term logic and to keep the number of different groupings in the field as low as possible. On the other hand, the (likely) reason for which the

change had been adopted in the first place, by the creators of these new logics, was the desire to emulate the human ability to make a change of mind about a particular aspect of a theory without having to relinquish the whole construction.

If the rationale for introducing changes in such a time-honoured a concept as that of logic is to better equip a system for simulating human reasoning activities, then the system should be analysed and judged with regard to its *behaviour as it is observable from outside*, exactly as we would do in judging the reasoning performance of a human being. What usually happens instead is that our logical tools are examined from the point of view of their performance *in our hands*, and then the conclusion is drawn that the instrument has provided the core of the action, and thus it is expected that embedding it in a machine should more or less be enough to reproduce the overall behaviour. The very serious flaw in this approach is that the really important part of the task has been carried out by the human completely outside the logical tool, using other mechanisms of which we are not usually aware.

An example should clarify this point. A mathematician examines a theorem that he has just proved. Ignoring the fact that any real size mathematical theorem is far too complex to be described in terms of formal logic, let us assume that the instrument of classical logic is used to do the check. Suppose a contradiction is found, that is a sentence which maps both to T and to a second value, at the same time. If not too depressed, the mathematician will trace back the source of the contradiction to either a wrong passage or to an original assumption, which could then, say, be weakened, and subsequently the proof can be attempted again. The mathematician has acted in a very rational way, of course. If the question is raised of what logic has been employed during this course of action, the mathematician will surely reply "classical logic". That is unfortunately wrong! The only information that could have been received from the logical tool is that the theory has collapsed. The processes of tracing back the source, and, even more, of salvaging most of the theory and weakening the (apparently) responsible assumption in the right way are completely beyond the logic as formally expressed. Nevertheless these processes are clearly the real core of the activity.

It may help, in order to accept this apparent inability of highly rational people to recognise and describe the processes really going on in their minds, to mention that the same situation has happened in the natural sciences, where for centuries scientists have believed themselves to be acting on Bacon's "collection" rules, and it has taken a great deal of effort from the philosophers of science to prove how wrong that reconstruction was.

For these reasons, we propose a new concept, namely that of *reasoning system*. A reasoning system is defined as a logic of the extended kind described above, but with some additional properties.

The first property is that the system is expected to work in isolation, which is to say that the transformation rules express all the operations that the system is able to perform. In this way, a reasoning system can very easily be seen as a skeletonised program. It is interesting to remember, though, that this is not necessarily so, and that the description of a reasoning system could apply, under certain condition, to a human as well.

A condition that follows from this definition is that any input registered by the system *must* be mapped to some value: that is because, once the signal is received, the system is doing something with it, even just ignoring it, and this amounts to mapping it to some value, which will have its own order relation with other values in that system domain.

A second consequence is that any eventual meta-level must be specified *inside* the system, if it is to play any role in the actual performance of the system. While of course such a meta-level could be potentially arbitrarily large, being expressed as a generation rule, in any moment it can be only finite, and only that part of the hierarchy can then be used. We would like to underline the strong links between this point of view and the constructivist arguments about the potential infinite, generation rules and programs as constructive proofs.

It is clear that there is a straightforward transformation from a logic into a reasoning system. Since several logics have been proposed to tackle the problems that we are addressing, we will compare them in order to

formally specify which requirements are the needed and which are the undesirable properties. We firstly translate these logics, namely classical, intuitionistic, non monotonic, uncertainty and interval, into reasoning systems. We think that such a translation is very close to the spirit in which these logics are said to perform against our problems. Anyway, the translation will not touch the standard formalism in which each logic is expressed (when there is such a thing). Instead, it will show itself in the way in which the systems are put into operation and their performances checked.

In order to stay as close to the originals as possible, and also for reasons of readibility, we express our transformation rules with an axiomatic notation, so that, for example, "$\alpha_n \vdash \alpha_{n+1}$" is to be read as "from the information '$\alpha \mapsto T$ at state n' we can derive '$\alpha \mapsto T$ at state $n+1$'". When the transformation rules become too complex for this notation, as in Chapter 6, we revert to the functional notation. The traditional "bar" notation of Natural Deduction texts is very effective pictorially, but far too cumbersome for systems of this complexity.

In this chapter we proceed to develop a general structure which can be used to reflect the way in which the transformation rules of a reasoning system can be used to maintain, improve or create knowledge. Using this structure and a set of five classes of reasoning systems, some derived from well-known logics, we show how the reasoning systems can be integrated into a common framework and considered from a common viewpoint.

We highlight some properties of each particular reasoning system and analyse the way in which the reasoning systems approach the fundamental tasks before them. The most powerful tool we use in this analysis is once again the theory of homogeneity.

4.2 A Formal Model for the Progression of Understanding

The observed homogeneity of individuals under certain properties can and does change over time. In order to model this, understand it and reason about it a reasoning system is needed which is supported by a model for changing understanding. By "changing understanding" we intend a progression of knowledge through time. We develop a set of tools by which classes of reasoning systems can be compared and evaluated subject to certain criteria.

In this section we do not restrict the definitions to refer to any particular reasoning system, but we use the general description of reasoning systems given in Section 4.1.

We now present the formal definitions for exploration sets and the supporting structure. An informal presentation of these definitions is given in Section 4.3.

Definition 4.2.1

A set, Sem, is a *semantic set* if $|Sem| \geq 2$, and the distinguished element, T, is a member of Sem.

Definition 4.2.2

Let Sem be a semantic set, and let L be a language, $L' \subseteq L$, and wff' be the set of well-formed formulae in L'. Then:

(i) a *perceived state* is any partial relation $wff' \leftrightarrow Sem$;

(ii) an *open state*, *o-state*, is any partial map $wff' \nrightarrow Sem$;

(iii) a *possible world* is any total map $wff' \rightarrow Sem$.

Definition 4.2.3

Let L be a language, $L' \subseteq L$, wff' be the set of well-formed formulae in L', and Sem be a semantic set. Let TR be a set of *transformation rules* of type:

$$L \cup (wff' \leftrightarrow Sem) \cup (L \times (wff' \leftrightarrow Sem)) \rightarrow L \cup (wff' \leftrightarrow Sem) \cup (L \times (wff' \leftrightarrow Sem)).$$

Let PW be the set of possible worlds, for L' and Sem, and I be the set $\cup\{(wff')^n : n \in \mathbb{N}\}$.

Then $RS = ((L, L'), TR, PW, I)$ is a *Reasoning System*.

When a sequence of well-formed formulae in I has been given to a reasoning system, the system must organise the set of formulae so that every formula is mapped to at least one value.

Definition 4.2.4

Let S be a set of perceived states for some reasoning system, then a *perceived progression* (*pp*) is the pair $(S, <)$, where $<$ is a total order on S.

Definition 4.2.5

Let $(S, <)$ be a perceived progression and $s \in S$. Then $(S, <, s)$ is an *originated perceived progression* (*opp*).

We use the usual definition of an isomorphism between originated perceived progressions, so that $f: S \rightarrow T$ is the unique isomorphism between $(S, <, s)$ and $(T, <, t)$ iff f is a one-to-one correspondence, preserving the order, $<$, and identifying s and t.

Definition 4.2.6

If $\mathfrak{E}_R$ is a set of originated perceived progressions for some reasoning system, R, and $\sqsubset$ is a total order on $\mathfrak{E}_R$, then $(\mathfrak{E}_R, \sqsubset, (S, <, s))$ is an *exploration set* (*es*), if it satisfies:

(i) $(S, <, s) \in \mathfrak{E}_R$;
(ii) $\forall (T, <, t) \in \mathfrak{E}_R . (S, <, s) \sqsubseteq (T, <, t)$.

$(S, <, s)$ is called the *origin of exploration* of the exploration set.

Definition 4.2.7

If $(\mathfrak{E}_R, \sqsubset, (S, <, s))$ is an exploration set then, for each $u \in S$, $[u]$ is the equivalence class defined as follows:

(i) $[u] = \{ t \in T : (T, <, t) \in \mathfrak{E}_R$ and
$u \mapsto t$ in the isomorphism between $(S, <, s)$ and $(T, <, t) \}$.

If $(T, <, t), (T', <, t') \in \mathfrak{E}_R$, and $u \in T$, then $T'[u]$ denotes the perceived state in T', u', such that $u \mapsto u'$ in the isomorphism between $(T, <, t)$ and $(T', <, t')$.

The $\mathfrak{E}_R$-*structure*, $[\mathfrak{E}_R]$, is defined by:

(ii) $[\mathfrak{E}_R] = \{ [t] : t \in S \}$.

There is an ordering on the $\mathfrak{E}_R$-*structure* induced by the ordering, $<$, on S, in the natural way.

It is clear that the $\mathfrak{E}_R$-*structure* is isomorphic to some continuous sequence of the integers (not necessarily infinite), with their natural order and identifying the origin of S with 0. Thus, for all $t \in S$, $t+1$ is the unique successor of t (if it has one).

Definition 4.2.8

If $(\mathcal{E}_R, \sqsubset, (S, <, s))$ is an exploration set then $(\mathcal{E}_R, \sqsubset, S, match)$ is a *valid exploration set* (*ves*) if:

(i) $match : (\{T : (T, <, t) \in \mathcal{E}_R\}) \rightarrow [\mathcal{E}_R]$;

(ii) $\forall (U, <, u), (T, <, t) \in \mathcal{E}_R$.
$(U, <, u) \sqsubset (T, <, t)$ implies $match(U) \sqsubset match(T)$;

(iii) $match(S) = [s]$.

S is called the *origin* of $\mathcal{E}_R$.

Definition 4.2.9

Let $RS = ((L, L'), TR, PW, I)$ be a reasoning system. Then a perceived state, s, is *closed* if it is not open and cannot be transformed, using the rules in **TR** exclusively, into an open state.

4.3 An Informal Presentation of Exploration Sets

In Section 4.2 we have presented the definitions that we require for our model of a progression of understanding.

Definition 4.2.1 defines two basic characteristic of a semantic domain, namely that has at least two elements and that the special value, T, belong to it. The reasons for this definitions have been given in Section 4.1.

In Definition 4.2.2(i) we introduce the concept of perceived states. They are defined simply as any partial relation between the well formed formulae in the syntax and the semantics of a given reasoning system. In

Definition 4.2.2(ii) an open state is defined as any partial mapping from the well formed formulae in the syntax and the semantics of a given reasoning system. In other words, in an open state each well formed formula maps at most to one semantic value. Definition 4.2.2(iii) specifies a possible worlds as an open state where each well formed formula maps exactly to one semantic value.

Definition 4.2.3 formalises the description of a reasoning system which we have given in Section 4.1.

Definition 4.2.4 presents the most basic structure we use to consider the development of knowledge and understanding.

It is the perceived progression, and will be used as a representation of the subjective view of the way in which events and knowledge have grown, changed or persisted. Clearly, the progression is linear - all of us have a subjective view of time as a linear progression. For most purposes we can consider this progression as infinite in both directions - most people do not have a particular point in their subjective time which they consider to be its beginning, nor one which they see as its end.

An originated perceived progression (Defn. 4.2.5) is a perceived progression with an identified origin. This is used in order to synchronise different perceived progressions, held one-at-a-time by the same subject, which all represent an understanding of the same world. The origins are all considered to be representative of the same perceived state as it is viewed with increasing and changing knowledge. The sequence of perceived progressions held by the subject is called an exploration set (Defn. 4.2.6).

The exploration set has an internal structure which is the canonical structure of the originated perceived progressions included in it. This is named in Definition 4.2.7. We then use the structure to define valid exploration sets, in which the structure of the exploration set itself is embedded in each of the perceived progressions it contains (Defn. 4.2.8). The reason for this is that the perceived progressions must also contain some kind of representation (seen from the outside) of the fact that the view of the world has itself changed.

Finally, in Definition 4.2.9 we introduce the concept of a closed state, the importance of which will become apparent later, in Section 5.

Some of the structure that is presented in Section 4.2 is not used until we have made a study of the basic properties of the reasoning systems that we compare. However, for completeness it has been presented in one piece.

4.4 The Role of Reasoning Systems in Knowledge Acquisition

Now we turn our attention to the use of reasoning systems as a tool for the preservation and development of knowledge.

In order to achieve the first of these goals, a reasoning system must address several tasks. Here we consider each of these tasks in turn and indicate which rules of the reasoning systems embed the particular solution (or failure to offer a solution) to each goal.

The first and most important of these tasks is to infer, from knowledge of the past and present behaviour of the world, the way in which the future behaviour will progress. All reasoning systems are provided with rules of deduction in some form, but the basis on which these transformations rest are the *rules of identity*, and *dynamic monotonicity*, which tell us whether the premises used for the applicatinn of a rule remain constant during that application.

Therefore, although the rules of identity and dynamic monotonicity are not themselves tools for extracting new knowledge about the future from knowledge of the past, the possibility for a reasoning system to predict the future is based on the strength of these rules.

The next task which confronts reasoning systems is to resolve contradictions that arise out of inaccurate prediction about the future, or discovery of contradictory information. A failure to face this task must be paid for by a collapse of the reasoning system when such contradictions arise, because

once contradictions are freely accepted then there is no way to continue to reason successfully. As we will see, there are many different approaches to the resolution of inconsistencies, from the most trivial to the highly sophisticated. All the methods are expressed through the *rules of consistency*, dynamic monotonicity and *non-monotonicity*.

The third task which the reasoning systems face is expressing their view of the past, based on their knowledge of the present. There are two main directions that are taken - one is to consider the world platonistically, as a fixed and unchanging entity and consider gaining knowledge as a progressive exploration of this world. The other approach is to view knowledge as "the world" and see it as a constructive development from the past and into the future. The particular view that is taken is expressed through the *rules of double negation* and of the *excluded middle*.

4.5 Uncertainty Structures and Observations

In this Section we make and explain a definition that will be used to build some of the reasoning systems we discuss in Section 4.6. The intuitive meaning of these definitions is given at the end of this section.

Definition 4.5.1

An *uncertainty structure* (*us*) is the pair $(C, <)$, where C is a set, the elements of which are called *uncertainty values*, and $<$ is a total order on C, such that:

$$\exists\, c \in C.\ \forall d \in C.\ \ c \leq d\ .$$

c is called *bot*. If C is finite, then $(C, <)$ is a *finite uncertainty structure* and the element, e, satisfying $\forall d \in C.\ d \leq e$, is called *top*.

Definition 4.5.2

Let R be a reasoning system and $(S_R, <, s)$ be an originated perceived progression. Then $\alpha_n^c \mapsto T$, in o-state $n \in S_R$, is a *non-derivable* statement if it could not have been produced, by application of any of the transformation rules of R, from the o-state, $n-1$.

Definition 4.5.3

Let $(S_R, <, s)$ be an originated perceived progression for a reasoning system, R. Then the set of non-derivable statements for S_R, $ND(S_R)$, is defined as follows:

$$ND(S_R) = \{ \alpha_n^c : \alpha_n^c \text{ is a non-derivable statement} \}.$$

Definition 4.5.1 presents a simple structure, called an uncertainty structure. It is, essentially, a totally ordered set of elements called certainty values. There is only one constraint, which is that there be a least value, called *bot*. If the structure is finite, then there is, of necessity, a maximum value. Of course, there can be a maximum value in an infinite structure as well. In either case, the maximum value is called *top*.

The structure will be used to provide a scale of values, which can be interpreted as the degrees of conviction with which a certain statement is believed to be true (or false). The uncertainty structure plays its most important role in the treatment of uncertainty *RS*s (cf. Section 4.7.4), where we will restrict ourselves to finite structures. Later, in the treatment of adaptive *RS*s the structure will also play a significant part.

In Definition 4.5.2 we introduce the concept of an non-derivable statement. This is defined as any statement appearing in a perceived progression which could not have been derived from the knowledge embodied in the previous states, using the rules of the reasoning system. The set of all these statements is defined in Definition 4.5.3.

These two definitions will be important when we consider the behaviour of the classes of reasoning systems which we introduce in Section 4.7. The important point contained in the definition of an non-derivable statement, is that the statement cannot have been derived internally, and must therefore have entered from "outside". In the application of the reasoning systems to genuine problems it is clear that there is a need to consider its ability to interact with its environment. This will be dealt with in much greater detail in Chapter 5. However, it is not difficult to see that the greater the variety of inputs that a reasoning system can cope with, the greater its flexibility in the interaction with its environment.

4.6 Steps towards a Classification of Reasoning Systems

We now examine certain notation and its meaning, as it will be employed in our introduction of the rules for the reasoning systems we study.

It should be apparent that the use of any entailment in order to reach some consequence from a given premise represents a change in the perceived state of knowledge - that is, the use of the entailment rule carries the perceived state of knowledge from one state to the next, progressing through a subjective time. This is easily seen to be the case when one considers the way in which entailment is interpreted: "since these premises are true, *then* we infer that these consequences *follow*" - clearly, the words "then" and "follow" imply a progression through subjective time.

We point out that this is not the case for the symbol "$\not\vdash$", which represents the impossibility of leaving the state on the left and reaching the one on the right without additional information. Rules involving this symbol are really redundant, since the non-existence of their duals is actually enough: we have given rules including it only to contrast with their duals when these are actually employed by other reasoning systems.

We note that the subjective time is not "real time", nor the variable recognised as time by physicists, but is measured only by the perceived act

of changing understanding.

To represent this formally, we will use the structure we defined in Section 4.2, taking a valid exploration set, $(\mathfrak{E}_R, \sqsubset, S, match)$, for the given reasoning system, R, and considering the progression in time to be movement from perceived state to perceived state in a perceived progression in $\mathfrak{E}_R$. Thus, we can index a statement, α, in the reasoning system, with a perceived state: so that "α_n" refers to the apparent value of α in the perceived state, n.

The transformation rule equivalent to entailment will be denoted by the usual symbols for classical *RSs* and intuitionistic *RSs* ("$\vdash$" and "$\mid\!\!-$" respectively). For the other classes of reasoning systems that we study, an index will be used to distinguish the different entailments.

For convenience, we will use the standard notation "$|\alpha|$" as an abbreviation for "either α or $\sim\alpha$". This notation is used particularly for expressing guards that must be satisfied for the operation of a rule.

The school of Intuitionism long ago recognised the value of a different interpretation of negation to that used by classical logicians. That is, negation is taken to mean "the absence of a proof". This can be interpreted in two subtly different ways: either as a simple absence of positive proof, as Dummett does [DU77], or as an absence of proofs both ways, which is equivalent to an "I don't know either way" statement. This approach is taken, among others, by Kripke [BM77]. We will follow the latter interpretation. The reason for this choice will be explained in Section 4.7.2.

It is then clear that a statement, α, made in an intuitionistic *RS* must be interpreted as saying "I have a proof of α". Of course, there is still a need for the negation used in classical *RS*, so that one can make the statement: "I have a proof that α is not true". This use of two different negations requires the use of two different symbols - we use "$\sim$" to denote the classical negation and "$\dashv$" to denote the lack of proofs. We follow in this the distinction made by Fitting [FI69], following Kleene [KL52], in his intuitionistic falsification tableaux: this notation is not

standard and many authors use only one symbol, leaving it to the reader to understand the meaning.

The three main properties of the concept denoted by "$\dot{\sim}$" are, firstly, that there can be at most one attached to any statement, secondly, this must be the leftmost negation in any sequence of negations, and finally, that there can be no use, in substitution, of statements prefixed by this negation. The reason for these constraints is that this form of negation is really a meta-statement pushed into the object-level and must refer to the whole of an otherwise purely object-level statement. Notice, for example, that the statement "$\dot{\sim}\dot{\sim}\alpha$", which with our chosen interpretation, should be read "I don't know wether I don't know wether α is true or false", corresponds to an absurdity; that is, it is always false.

We will later show that the rules of classical *RS*s can be reformulated using starred-negation, without any significant change in the meaning of the rules.

Several reasoning systems have been suggested and explored which feature some kind of degrees of belief, uncertainty or other similar concept. We will model them using an uncertainty structure (Defn. 4.5.1). Thus "α_n^c" is read "α is known with certainty c in perceived state n".

Classes of reasoning systems which use finite intervals of perceived time will also be introduced.

4.7 Rules for Five Classes of Reasoning Systems

In this Section we introduce the reasoning systems that we will compare and contrast in the following sections. Explanation of any particular features of each of the reasoning systems will also be given.

Only the rules in which we are interested are listed, and these in a very

general form. The rules which interest us are, of course, those which connect to the tasks listed in Section 4.4. Each set of rules defines a class of reasoning systems satisfying the very general criteria posed by the rules. The classes are given the name of their best known member. When other well-known reasoning systems are also members of a class, this will be indicated.

The particular rules which are considered are the rules of identity, consistency, non-monotonicity and dynamic monotonicity, double negation and the excluded middle.

Here we concern ourselves only with propositional reasoning systems. It is clear that all these classes can be extended to first-order, and that each such extension would present characters and problems of its own. It is also true, on the other hand, that basic propositional rules are never altered by a transformation to first-order; in any case, whatever power the extension can add, it will be useless if the reasoning system has not been able to face the more basic tasks of surviving, avoiding or solving inconsistencies and preserving knowledge.

We present the following classes of reasoning systems:

classical *RS*s, intuitionistic *RS*s, non-monotonic *RS*s, uncertainty *RS*s and interval *RS*s. Later we will introduce two further classes of reasoning systems, called unforgiving *RS*s and adaptive *RS*s.

As we have already stated (Section 4.6), all the formulae will be subscripted with a perceived state (Defn. 4.2.1) taken from a particular originated perceived progression (Defn. 4.2.3), $(opset, <, origin)$.

4.7.1 Classical Reasoning Systems

In the class we call classical *RSs* the following rules hold:

(a) $\alpha_n \vdash \alpha_{n+1}$; rule of identity

(b) $\alpha_n, \sim\alpha_n \vdash$; rule of consistency

(c) $\sim\sim\alpha_n \vdash \alpha_{n+1}$; rule of double negation

(d) $\vdash \alpha_n \vee \sim\alpha_n$. rule of the excluded middle

As we mentioned above, these rules could have been rewritten using the starred-negation notation, without any effect, and the same is true for all the other classes of reasoning systems: that is, adding redundant notation does not alter the meaning of a formalisation.

We now prove this by giving an alternative set of rules for classical *RSs* using the starrred-negation notation of Intuitionism, and transforming it back to the original one.

Proposition 4.7.1.1

The following set of rules implies the one given above for Classical *RSs*:

(a') $\alpha_n \vdash \alpha_{n+1}$; rule of identity

(b') $\alpha_n, \sim\alpha_n \vee \dot{\sim}\alpha_n \vdash$; rule of consistency

(c') $\sim\sim\alpha_n \vee \dot{\sim}\sim\alpha_n \vdash \alpha_{n+1}$; rule of double negation

(d') $\vdash \alpha_n \vee \sim\alpha_n \vee \dot{\sim}\alpha_n$. rule of the excluded middle

Proof:

Rule (a') is the same as (a). rule (c') implies rule (c).

Rule (b'), $\alpha_n, \sim\alpha_n \vee \doteq\alpha_n \vdash$, is equivalent, by substitution, to $\sim\alpha_n, \sim\sim\alpha_n \vee \doteq\sim\alpha_n \vdash$.

By rule (c'), $\sim\sim\alpha_n \vee \doteq\sim\alpha_n \vdash \alpha_{n+1}$.

Hence $\sim\alpha_n, \sim\sim\alpha_n \vee \doteq\sim\alpha_n \vdash$ implies $\sim\alpha_n, \alpha_{n+1} \vdash$, which is rule (b).

Rule (d'), $\vdash \alpha_n \vee \sim\alpha_n \vee \doteq\alpha_n$, can be transformed by substitution into:

$\vdash \sim\alpha_n \vee \sim\sim\alpha_n \vee \doteq\sim\alpha_n$.

Applying rules (c'), the above becomes:

$\vdash \sim\alpha_n \vee \alpha_n$, which is equivalent to

$\vdash \sim\alpha_n \vee \alpha_n$, that is, rule (d). □

Notice that $\doteq\sim\alpha_n \vdash \alpha_{n+1}$, which is a consequence of rule (c'), is equivalent to (*), $\doteq\alpha_n \vdash \sim\alpha_{n+1}$, using substitution and rule (c') again.

We note that Proposition 4.7.1.1 proves an implication instead of a bi-implication: this is simply because the language used in the second set of rules is richer than the first. Thus it is impossible to obtain expressions in the second language by transforming expressions in the first, because of absence of characters.

The second set of rules is not, however, more powerful than the first, exactly as giving more than one name to an object does not create more objects.

The same kind of proof can be constructed for all the other reasoning systems, but we believe that this proposition has proved the general point

beyond need for further results.

It is also clear that, since the canonical semantics for classical *RS*s has only two values, adding the notation for uncertainty would not make any difference. The variable would simply be interpreted always as "*top*". The intuitive reason here is slightly different from in the previous case: what happens is that a new name in a language can only be assigned an old, available meaning, whatever its original one.

Notice that the rule, (c'), in the second set of rules, could appear to support the idea of "negation by failure". This impression could be conveyed by an interpretation of the starred-negation, according to the intuitionistic tradition, as "It is not proven that", and using the form (*).

This interpretation can be succesfully made only because, as we prove in the proposition, the rules of classical *RS*s are such that they *force* the meaning of starred-negation to coincide with the usual negation, whatever the interpretation. Once again, the rules are what really matters, not the notation.

The concept of "absence of proof" is thus really meaningful only when used in an intuitionistic framework, exactly because negation by failure is explicitly excluded.

Attempts at mixing concepts derived from a philosophy of evolution, like Intuitionism, with a platonistic, unchanging world, as is that enforced by classical *RS*s, can only yield confusion, without offering any possibility of new inferences at all.

We note that some modal or temporal *RS*s fall into the class of classical *RS*s, notwithstanding their additional connectives. In fact, as far as this classification is concerned, the only important behaviour is that related to the rules listed, and this is unaffected by the presence of syntactical marks for extra meta-properties.

4.7.2 Intuitionistic Reasoning Systems

In the class we call intuitionistic *RS*s the following rules apply:

(a) $\alpha_n \vdash \alpha_{n+1}$; rule of identity

(b) $\alpha_n, \sim\alpha_n \vee \div\alpha_n \vdash$; rule of consistency

(c) $\div\sim\alpha_n \nvdash$; rule of double negation

(d) $\vdash \alpha_n \vee \sim\alpha_n \vee \div\alpha_n$. rule of the excluded middle

This set of rules is not the standard one given for intuitionistic *RS*s: in particular, rule (c) is usually expressed as "$\sim\sim\alpha_n \nvdash$" (or, rather, by the absence of "$\sim\sim\alpha_n \vdash \alpha_{n+1}$"). Also, no rule of the excluded middle is normally provided: in fact, its absence is considered a signature of Intuitionistic philosophy.

We now proceed to justify our claim that the above set of rules do correspond to the intuitionistic philosophy.

All intuitionistic logicians recognise the rule (*), "$\sim\sim\alpha_n \nvdash \alpha_{n+1}$", as essential for the expression of the intuitionistic philosophy.

Although marked negations have been used in intuitionistic tableaux [FI69] and in semantics for intuitionistic *RS*s [BM77], they have not appeared in any formalisation. We claim that, under all the three standard interpretations of intuitionistic negation, rule (*) does not hold. The three interpretations of "$\sim\alpha_n$" are:

(i) Until time n I have no proof that α is true;

(ii) At time n I know that I will never find a proof that α is true;

(iii) At time n I don't know wether α is true or false.

The first two interpretations are presented, among others, by Dummett [DU77], while the third can be inferred from Kripke's work on semantics.

It must be noted that we do accept the intuitionistic stand against the rule of the excluded middle as a central feature of intuitionistic philosophy. Our claim here is that rule (d) renders that principle in a much more precise form. That is because, once it has been accepted that a second negation symbol is needed according to the argument above, we are *forced* to recognise (at least) three values for intuitionistic *RS*s.

The name for rule (d) has been chosen for reasons of symmetry with the formalisation for classical *RS*s, (often called the rule of excluded third) even if, in rule (d), the value excluded is actually (at least) the fourth.

We are then in a position to maintain that intuitionistic *RS*s are many-valued reasoning systems, whatever the notation chosen to disguise the fact. In this we link our interpretation to the work of Lukasiewicz [LU70], who first recognised the necessary connection between a constructive view of knowledge and many-valued reasoning systems.

It is interesting to note that rule (d) prevents interpretation (i) of starred-negation, which would require not only an additional disjunct in rule (d), but also an additional rule, such as $\alpha_n \vdash \dot{\sim}\sim\alpha_{n-1}$, in order to distinguish between $\dot{\sim}\sim\alpha_n$ and $\dot{\sim}\alpha_n$.

Also observe that we have no rule of identity for starred-negation, that is we do not allow it to propagate from state to state. This can be intuitively understood remembering that starred-negation would be interpreted as (iii), being thus inconsistent with the other two values. The presence of the rule of excluded middle nevertheless guarantees that, in each state, a value is given to every formula whose terms and predicates are understood.

In fact, of the interpretations listed, only interpretation (iii) is consistent with the rules, since the first two are not satisfied by the rule of the excluded middle. However, the rules as they are given do not force

interpretation (iii). When the truth values are T, F and D ("don't know"), then the value of $\dot{-}\alpha_n$ can be either D or F when α_n is T or F, and still remain consistent. This is because our constraints on starred-negation do not allow us to talk about the value of $\dot{-}\alpha_n$ when the value of α_n is known to be T or F (only the rule of the excluded middle allows the introduction of starred-negation).

Now, consider the weakening of the constraints on the ways in which starred-negation can be used, so that we can have the value $\sim\dot{-}\alpha_n$, or similar, by removing the restriction that starred-negation must always be the outermost of any string of negations.

This alone does not force the interpretation that we have given, but with the additional rule,

$$\alpha_n \vee \sim\alpha_n \vdash \sim\dot{-}\alpha_{n+1},$$

we do force the interpretation, so that $\dot{-}\alpha_n$ is F if the value of α_n is T or F.

It is interesting to note that if we also relax the restriction on the number of starred-negations which occur in a formula, and use the rule,

$$\alpha_n \vee \sim\alpha_n \vdash \dot{-}\dot{-}\alpha_{n+1},$$

instead of the one above, then the value of $\dot{-}\alpha_n$ is forced to be D when α_n is T or F. This is consistent with an interpretation of starred-negation as a refusal to talk about a statement. We have already made the point, in Section 4.7, that with the interpretation (iii), the value of $\dot{-}\dot{-}\alpha_n$ is always F, using a semantic proof. On the other hand, with the alternative interpretation as refusal to talk about a statement, $\dot{-}\dot{-}\alpha_n$ takes the meaning "I am obliged to say whether α is true or false, at time n".

Those modal or temporal *RS*s which reject the double negation rule of classical *RS*s, that is, have a third value in their domain, are based on intuitionistic *RS*s.

4.7.3 Non-Monotonic Reasoning Systems

We use a finite uncertainty structure (Defn. 4.5.1) $(\{bot, top\}, <)$ with $bot < top$.

In the class we call non-monotonic *RS*s the following rules apply:

(a) $\alpha_n^{\,c} \models \alpha_{n+1}^{\,bot}$, $(c = max(\{t : |\alpha|_n^{\,t}\}))$; rule of non-monotonicity

(b) $\alpha_n^{\,c}, \neg\alpha_n^{\,c} \models$; rule of consistency

(c) $\models \alpha_n^{\,bot} \vee \neg\alpha_n^{\,bot}$. rule of the excluded middle

There is no standard formalisation for non-monotonic *RS*s: as a matter of fact, there is not agreement over what a non-monotonic *RS* is precisely, and many of the candidates have no formalisations anyway. It is likely that our formalisation would nevertheless be considered unorthodox by some experts in the field, because of the use of an uncertainty structure. We claim that:

(i) our set of rules reflects the common understanding behind the ideas inspiring non-monotonic *RS*s;

(ii) without such a structure non-monotonic *RS*s flatten over database management systems or worse;

(iii) notwithstanding the use of the uncertainty structure, our system for non-monotonic *RS*s is not equivalent to the set of rules for uncertainty *RS*s (iv), because of a fundamental difference between the rules of non-monotonicity and dynamic monotonicity.

We will present the argument to support each of these points in turn, but first we prove a result about the non-derivable statements for the non-monotonic *RS*s.

We introduce the notation "$\not\mapsto$" to denote "does not map to".

Proposition 4.7.3.1

Let $(S_N, <, s)$ be an originated perceived progression, containing only open states (cf. Defn. 4.2.2), for a non-monotonic RS, N. Then $\alpha_n^c \in ND(S_N)$ iff $\alpha_n^c \mapsto T$ in state $n \in S_N$ with $c = top$.

Proof:

($\Rightarrow$)

Suppose $\alpha_n^c \in ND(S_N)$. Then, by Definition 4.5.3, $\alpha_n^c \mapsto T$ in perceived state n and could not be derived with the rules of N.

rule (c) grants that $\alpha_{n-1}^{bot} \vee \sim\alpha_{n-1}^{bot}$.

Suppose $\sim\alpha_{n-1}^{bot} \mapsto T$ and $\alpha_{n-1}^{top} \mapsto T$. Then rule (a) implies $\alpha_n^{bot} \mapsto T$., so that $c \neq bot$.

Suppose $\sim\alpha_{n-1}^{bot} \mapsto T$ and $\alpha_{n-1}^{top} \not\mapsto T$. Then rule (b) implies $\alpha_{n-1}^{bot} \not\mapsto T$ and rule (a) implies $\sim\alpha_n^{bot} \mapsto T$. Therefore, by rule (b), $c \neq bot$.

Suppose $\alpha_{n-1}^{bot} \mapsto T$ and $\sim\alpha_{n-1}^{top} \mapsto T$. Then rule (a) implies $\sim\alpha_n^{bot} \mapsto T$ and, by rule (b), $c \neq bot$.

Finally, suppose $\alpha_{n-1}^{bot} \mapsto T$ and $\sim\alpha_{n-1}^{top} \not\mapsto T$. Then rule (a) implies $\alpha_n^{bot} \mapsto T$, so $c \neq bot$.

Therefore, $c = top$.

($\Leftarrow$)

Conversely, suppose $\alpha_n^c \mapsto T$ in state $n \in S_N$ with $c = top$.

Then, since there are no rules which introduce statements marked *top*, it cannot be that α_n^c could be derived internally.

Therefore $\alpha_n^c \in ND(S_N)$. □

Point (i)

The principal ideas on which non-monotonic *RSs* are based are as follows:

1. they must be able to survive contradictions, accepting the new information as the good one;
2. they must prefer information input by the user to that deduced from inside (otherwise they would progress from one false deduction to the next, building their own imaginary world);
3. they must recognise a contradiction when one is input or derived;
4. they must, of course, preserve information (which has not been overridden) from one state to the next.

In order to satisfy point (2) there must be some way of distinguishing between information input from outside and information derived from inside: the role of the reasoning system is then to manipulate these distinguished pieces of information according to point (2).

These points are all satisfied by the rules we have given: the first and second points by the use of rule (a), marking input with *top* certainty (by the result of Proposition 4.7.3.1) and internal deductions with *bot* certainty. The third point is precisely that made by rule (b). Notice that the only ways in which contradictions can arise are by deduction from inconsistent premises, leading to contradictory statements marked *bot*, or when the user inserts a statement and its negation at the *same* time. Most implementations expect only one statement to be entered at a time, so that only the former reason is likely to be of concern. Finally, the fourth point is satisfied by rule (a), which allows information to be preserved provided it is marked as internal (*bot*).

We would like to note that, despite the apparently contraditory nature of

the points listed above (the first with the third, for instance), the set of rules we have proposed is internally consistent.

Point (ii)

Having shown that our set of rules is sufficient for expressing the above points, we now show that it is also necessary.

Firstly, if the uncertainty indices are removed, the set of rules collapses into classical *RSs*.

Assuming the uncertainty indices are removed, some of the other rules must be changed as well, in order to maintain the non-monotonic nature of the reasoning system.

A natural candidate would be to use the progression of states itself as a way to order the formulae and thus having an order of survival in case of contradictions. Rule (a) would then be replaced by, say, rules (a'), "$\alpha_n \models \alpha_{n+1}$", and (a"), "$\alpha_n, \sim\alpha_{n+1} \models \sim\alpha_{n+2}$", by which the new formula always replace the old in case of inconsistencies. The problem with this attempt is that, using rule (a'), the old formula can always be propagated into the next state, until it reaches the state when the new one is inserted, thus causing unsolvable contradiction by rule (b).

It is clear that rules (a'), (a") and (b) cannot all be maintained. Therefore one possibility could be to eliminate the rule of identity, (a'). This approach would violate point (4) above, since memory would be lost.

The next plausible attempt on this line could be to weaken the rule of identity, for example by imposing the condition that identity propagation is suspended when new information is input, and then retreived later. This device, apart from being a rather low-level mechanism, does not work. The reason is that, since there are no tags to distinguish input from suspended memory, the retrieved memory must be considered as new input, and would thus override the previous information, violating point (2).

A last attempt could be renouncing the rule of consistency itself: the system

would now be a simple database, accepting everything and preserving all the inconsistencies.

The same argument holds if the uncertainty structure is kept, but other rules are renounced as above.

Point (iii)

The important thing to notice here is that, in rule (a), the certainty value of the formula surviving the contradiction decreases, while in the dual rule, (a), in uncertainty *RSs* the certainty of the surviving formula does not change.

If an rule of double negation is given in the intuitionistic form, then a "don't know" value is introduced.

Alternatively, this rule can appear in a classical form (like (*) above), as "$\vdash \sim\alpha_n^{c} \models \alpha_n^{best}$".

It is interesting to note that in this context negation-by-failure is neither flattened onto classical *RSs* nor leads to immediate contradiction. This is because one of the possible source of inconsistencies, the one caused by clash with external input, is already dealt with by axion (a). This does not make negation-by-failure safe, anyway, since by the use of this rule a purely internal contradiction could always develop, and that wuld be fatal.

A solution to this last problem is by using a more complex uncertainty structure, and marking the formulae inferred by negation-by-failure with an index lower still than that used for other internal deductions.

We will not pursue this route any further here, though, since this concept is completely extraneous to the ideas of non-monotonicity, and belong to the adaptive class of reasoning systems which we will develop later.

We note that, adding the negation-by-failure rule, our set of rules for non-monotonic *RSs* provides a precise description of the behaviour, as far as classes of *RSs* are concerned, of the mechanism behind the PROLOG

programming language.

There is a second possible formalisation for non-monotonic *RSs*, based on the following infinite uncertainty structure (Defn. 4.5.1), $(\{bot, top\} \times \{n \in oppset : n \geq origin\}, <^{\cdot})$ (cf. comment preceding treatment of classical *RSs*).

The order "$<^{\cdot}$" is defined as follows:

- $(bot, n) <^{\cdot} (top, n)$ and - $(top, n) =^{\cdot} (bot, n+1)$.

The set of rules which is then used is as follows:

(a') $\alpha_n^c \models \alpha_{n-1}^c$, $(c = max(\{d : |\alpha|_n^d\}))$; rule of dynamic monotonicity

(b') $\alpha_n^c, \sim\alpha_n^c \models$; rule of consistency

(c') $\models \bigvee \{\alpha_n^c \vee \sim\alpha_n^c : c \in C\}$. rule of the excluded middle

We will refer to the first set of rules as "finite NM *RSs*", and to this second set as "infinite NM *RSs*". It can be seen that using this new uncertainty structure and set of rules, almost the same behaviour is achieved as with the finite NM *RSs*. In order to make the behaviours identical, the second set of rules must be altered to force the infinite NM *RSs* to view the states from the "past" as equivalent, so that (b') is supplemented with:

(b'') $\alpha_n^c, \sim\alpha_n^d \models$, $(c, d <^{\cdot} (bot, n))$.

Assuming that there is a way of distinguishing information input from outside from that derived internally (as we proved was the case in the finite NM *RSs*, Propn. 4.7.3.1) , we claim that point (2) is satisfied by the infinite NM *RSs*, by marking information input at perceived state n with certainty (top, n). This works because this way rule (a') allows the last

input to overrride previous contradictory information, however obtained.

4.7.4 Uncertainty Reasoning Systems

We use a finite uncertainty structure (Defn. 4.5.1) $(C, <)$.

The rules of the class we call uncertainty *RSs* are as follows:

(a) $\alpha_n^{\ c} \vdash \alpha_{n+1}^{\ c}$, $(c = max(\{d : |\alpha|_n^{\ d}\}))$; rule of dynamic monotonicity

(b) $\alpha_n^{\ c}, \neg\alpha_n^{\ c} \vdash$, $(bot < c)$; rule of consistency

(c') $\vdash \vee\{\alpha_n^{\ c} \vee \neg\alpha_n^{\ c} : c \in C - \{bot\}\} \vee \alpha_n^{\ bot}$. rule of the excluded middle

It is interesting to note the similarity between these rules and the infinite NM reasoning systems rules above. The only differences arise because the uncertainty structure that is used is finite in this case, and the *bot* elements are considered as equivalent in this set of rules. In fact, the latter difference can be modified in the infinite NM *RSs* rules, to achieve a class of non-monotonic *RSs* using a "don't know" value.

An important consequence of the use of a finite uncertainty structure, together with rule (a), is that the domain of uncertainty *RSs* has the property of dynamic monotonicity. The meaning of this is that in order to change a belief that a statement is true to a belief that it is false, there must be an increase in the certainty. Since certainty cannot increase indefinitely in a finite structure, there must come a point at which there can be no more changes of mind. This is the fundamental difference with the non-monotonic *RSs*.

If an infinite uncertainty structure with a *top* value is used then a reasoning system based on a classical *RSs* framework can be built, in which certainties (or probabilities) are specified for each statement and assumed to

be good forever. These reasoning systems have all the properties of classical *RSs* as far as this analysis is concerned. Several kinds of classic probability *RSs* fall into this group. Otherwise, the reasoning systems with infinite uncertainty structures are non-monotonic *RSs*, such as most of the reasoning systems based on fuzzy set theory. Even if the uncertainty structure is finite, it is still essential to have the dynamic monotonicity rule, or else the reasoning system flattens onto one of the preceding reasoning systems.

4.7.5 Interval Logics

For the purposes of these rules we will require an extra index which we call a *persistency counter*. It is drawn from the set $\mathbb{N} \cup \{\omega\}$.

The rules of the class we call Interval *RSs* are as follows:

(a) ${}^{p}\alpha_n \vdash {}^{p-1}\alpha_{n+1}$, $(\omega \neq p > 0)$;

(b) ${}^{\omega}\alpha_n \vdash {}^{r(\alpha,n)}\alpha_{n+1}$;

rules of identity

(c) $\alpha_n , \sim\alpha_n \vee \dashv\alpha_n \vdash$; rule of consistency

(d) $\vdash \alpha_n \vee \sim\alpha_n \vee \dashv\alpha_n$. rule of excluded middle

In rule (b), $f: Syn \times oppset \rightarrow \mathbb{N} \times \{\omega\}$ defines an interval of persistency for the statement, α, at perceived state, n, with f satisfying the single restriction that if $f(\alpha, n) = \omega$, for some α and some n, then

$$\forall m \in oppset . (n < m) \Rightarrow f(\alpha, m) = \omega.$$

In the class of interval *RSs* that we have proposed, a pattern is expected in the behaviour of the world, which is embodied in the function, f. The

use of this function is to determine an interval, for a given statement known in a particular perceived state, for which it is expected that the truth of the statement will persist. The use of this form of expected behaviour is very easy to see in human reasoning, for example, when humans invest money, it is usually with the premise that they will live long enough to reap the benefits of their investment. On the other hand, they do not expect to live forever. It is also clear that humans expect different intervals of persistency, depending on the statement in question. In the previous example, the investors usually expect their life span to extend to several years, but they probably expect the institution in which they invest to persist much longer.

Although interval *RS*s have in common with non-monotonic *RS*s the fact that it is possible to review the behaviour of a statement, they are actually very different. non-monotonic *RS*s do not try to understand, or commit themselves to a prediction of, the behaviour of the world they perceive. They simply accept change without question, reshaping their world to fit the new information and not seeing any pattern in the changes.

Unlike non-monotonic *RS*s, interval *RS*s are prepared to pay the price of incorrect prediction about the future, by entering an inconsistent state, and thus collapsing, when the truth of a statement changes before it is expected to.

It is important to realise that the persistency counters are *not* uncertainty indices, but only a way to measure the expected life of a statement. The function f which determines the interval of persistency of the statement is clearly a meta-level function, and thus opaque to the rules at the object-level. The presence of counters in rules (a) and (b) is then a signal aiding the outside observer to understand what is going on; for the same reason, these counters are not introduced in rules (c) and (d), in which the important information conveyed is about each particular perceived state, and the counter is then irrelevant.

This kind of meta-level functions will assume a central role in the adaptive *RS*s.

Note that the interpretation of starred-negation is, once again, constrained by the rules that we adopt about substitution. In fact, if substitution of starred-negation for another formula is allowed, then the rules of identity hold for starred-negated formulae as well, which means that we are allowed to make predictions about the persistency of our ignorance.

If, on the other hand, such a substitution is not allowed, then starred-negated formulae can be introduced only by the rule of the excluded middle, which means that ignorance cannot be reasoned about, but only admitted when knowledge is absent.

This concludes our preliminary discussion of the classes of reasoning systems that we wish to examine.

Chapter 5

A Comparison of Five Classes of Reasoning Systems

5.1 Introduction

In this chapter we analyse the five classes of reasoning systems and compare them, firstly through the lens of homogeneity theory and then using the tests of successful interaction.

We define a set of screens and properties and prove that each rule of identity of a reasoning system is equivalent to an assumption of homogeneity of a particular property on the relevant screen. The screens are then proved to naturally determine a partial order.

We then set the context for studying the interactive power of the systems. We construct, for each system, the possible interaction that reveals its weaknesses, either in the field of reaction or in that of survival.

A second order based on these performances is constructed. This order is embedded in the first one.

All these results have relevance from many perspectives, throwing light on the nature of the rules of identity, on the role of assumptions of homogeneity, and particularly, because of the embedding of the orders, on the connections between basic rules like identity and complex interactive performances. These points are all discussed at the end of the chapter.

5.2 Various Screens for Viewing Perceived Progressions

The definitions presented in this section are used in the first stage of our comparitive analysis of the reasoning systems we have considered. Their informal explanation is presented in Section 5.3.

Definition 5.2.1

Let $(S_R, <, s)$ be an originated perceived progression for a reasoning system, R. Then Q_c: $\mathbb{P}(S_R) \rightarrow Bool$ is defined as follows:

$$Q_c(X) \Leftrightarrow (X \subseteq S_R \wedge X \neq \emptyset).$$

Definition 5.2.2

Let $(S_R, <, s)$ be an originated perceived progression for a reasoning system, R, and t is a perceived state in S_R. Then Q_{int}^t: $\mathbb{P}(S_R) \rightarrow Bool$ is defined as follows:

$$Q_{int}^t(X) \Leftrightarrow (X \subseteq \{u : u \in S_R \text{ and } t < u\} \wedge X \neq \emptyset).$$

Definition 5.2.3

Let $(S_R, <, s)$ be an originated perceived progression for a reasoning system, R, and t be a perceived state in S_R. Then Q_N^t: $\mathbb{P}(S_R) \rightarrow Bool$ is defined as follows:

$$Q_N^t(X) \Leftrightarrow (X \subseteq \{t, t+1\} \wedge X \neq \emptyset).$$

Definition 5.2.4

Let $(S_R, <, s)$ be an originated perceived progression for a reasoning

system, R, t be a perceived state in S_R with $t < u$, $d \in [0,1]$ and $K \subseteq S_R$ satisfying:

$$\forall n \in S_R . \forall u, v \in K . (u < n < v) \Rightarrow n \in K.$$

Then $Q_1^{\langle t,d,K \rangle} : \mathbb{P}(S_R) \rightarrow Bool$ is defined as follows:

$$Q_1^{\langle t,d,K \rangle}(X) \Leftrightarrow (X \subseteq K \wedge |X| \geq d + (1-d)|K| \wedge (\forall u, v \in X . \forall w \in K . (u < w < v \Rightarrow w \in X))).$$

This concludes the set of screens we require. We now define two properties which we will need for our analysis in Section 5.4.

Definition 5.2.5

Let s be a perceived progression for some reasoning system, R. Then the predicate $P_{\Box}^{c}(s)$ is defined to be true iff $\alpha_s^{d} \mapsto T$ for some d such that $c \leq d$.

Definition 5.2.6

Let s be a perceived progression for some reasoning system, R. Then the predicate $R_{\Box}^{c}(s)$ is defined to be true iff $\alpha_s^{d} \mapsto T$ for some d such that $c < d$.

Definition 5.2.7

If $(\xi_R, \sqsubset, Orig.path)$ is a valid exploration set, $(S, <, s) \in \xi_R$ and $P : S \rightarrow Bool$, then $\mathfrak{F}$ is the set of *first states for* P *in* S iff:

$\forall n \in S.$ (i) $P(n)$ holds;

and (ii) $\forall t \in S.\ (n \text{ is a successor of } t) \Rightarrow (P(t) \text{ does not hold})$;

or $\forall t \in S.\ (t < n) \Rightarrow (P_a^c(t) \text{ does hold}).$

implies $n \in \mathfrak{F}$.

Proposition 5.2.1

Let $(\Sigma_a, \sqsubset, Orig, match)$ be a valid exploration set, $(S, <, s) \in \Sigma_a$, $P: S \rightarrow Bool$, and $\mathfrak{F}$ be the set of first states for P in S, then if $\mathfrak{F} = \emptyset$, $P(t)$ is false for every $t \in S$.

Proof:

Suppose $t \in S$ and $P(t)$ holds.

Then let $T = \{u: u \in S,\ u < t\}$.

If $P(u)$ holds for all $u \in T$ then, by Definition 5.2.7, $t \in \mathfrak{F}$, contradicting the hypothesis.

Therefore, $\exists u \in T$ such that $P(u)$ does not hold.

So, the state $v = max(\{u: u \in T,\ P(u) \text{ does not hold}\})$ is well-defined.

Then, if $P(v+1)$ holds, $v+1 \in \mathfrak{F}$ by Definition 5.2.7, contradicting the hypothesis.

So $P(v+1)$ does not hold, and if $v+1 \in T$ then this contradicts the definition of v. But, if $v \in T$ and $v+1 \notin T$, then, by construction of T, $v+1 = t$, so that, by supposition, $P(v+1)$ holds, which is a contradiction.

Therefore there is no $t \in S$ for which $P(T)$ holds. □

5.3 Informal Aspects of Some Screens on Perceived Progressions

In Section 5.2, four different screens operating on perceived progressions are defined: Q_C, Q_{Ins}^n, Q_N^n and $Q_I^{(t,d,K)}$. These will be used in Section 5.3 to analyse the assumptions that are made in the fundamental rules of the classes of reasoning systems considered in Section 4.7.

The first and simplest of the screens, Q_C (Defn. 5.2.1), takes as its Q_C-sets all those non-empty subsets of the perceived states in the progression to which it is applied. This screen is the least discerning of all the four, and identifies the entire set of states as the unique individual (cf. Section 3.13).

The second of the screens, Q_{Ins}^n (Defn. 5.2.2), is actually a class of screens, with n as the parameter of the class. It simply considers all non-empty subsets of the perceived states following (and including) the state n in the progression. It therefore produces a different individual for each state, which is all the perceived "future" of that state.

Definition 5.2.3 introduces the class of screens, Q_N^n, again with parameter n. This screen considers the state n and its successor in the perceived progression (and the subsets containing each alone). Once again there are as many individuals as states, and each individual is a pair of consecutive states. This screen is more discerning than the first, but extremely myopic.

Finally, in Definition 5.2.4, we introduce the most sophisticated class of screens, $Q_I^{(t,d,K)}$, with three parameters: a state, a depth and a set of consecutive perceived states. This screen looks at sequences of consecutive states in the subset K of the progression, examining the set only to a depth d and assuming t is the first state in the sequence in K. The individuals in this case are all the various intervals of consecutive states in the progression.

Having considered the four screens that we use, we present two properties, in Definitions 5.2.5 and 5.2.6, which take a statement and a certainty value as parameters and are designed to consider the behaviour of the particular statement, with the given certainty, in a given state. The first, P_α^c, is used

to check if $\alpha \mapsto \mathcal{T}$ with a certainty value greater than or equal to c, while the second, R_{α}^{c}, is used to check whether $\alpha \mapsto \mathcal{T}$ in a state with a certainty value greater than c.

In Definition 5.2.7, we introduce the important concept of first states for some property in a given progression. This is a set of states of one of two types: those for which the property holds and did not hold in the immediately preceding state, and those for which the property holds and held in all the preceding states. The reason for the consideration of the second type of states, which might appear a rather peculiar set to consider as "first states", is that if all the states of a perceived progression, until some state, have the given property then, if the progreesion does not have a starting state, there will be an infinitely long sequence of states with the property and having no first state! We consider any state in such a sequence as of equal importance as any more natural "first state".

The Section concludes with a single result (Propn. 5.2.1) proving that a progression with no first state for a given property contains no state with that property at all. This result is used in the following section, where we continue our comparative study of the reasoning systems presented in Section 4.7.

5.4 A Formal Treatment of Identity through Homogeneity

In this Section we proceed to match each of the reasoning systems introduced in Section 4.7 with a screen from the selection in Section 5.2, in order to show the way in which the rules of each reasoning system force a particular view of their perceived progressions. The study concentrates on the assumptions of homogeneity that lie behind the particular rule of identity (or its equivalent) adopted by each class of reasoning systems.

The final result of this section proves part of the partial order between the assumptions that are proved equivalent to each of the rules of identity. The rest of the partial order is also given, although the proofs are omitted,

since they are identical in pattern.

A discussion of the meaning of these results can be found in Section 5.5.

Proposition 5.4.1

Let $(\mathcal{E}_c, \sqsubset, Orig, match)$ be a valid exploration set for a classical RS, C, $\alpha \in wff$ in the reasoning system, $(S_c, <, s) \in \mathcal{E}_c$, and $\mathcal{F}$ be the set of first states for P_α^c in S_c. Then, given the validity of the rules (b), (c) and (d), listed in 4.7.1, the assumption of the validity of rule (a) is equivalent to the assumption $hom(S_c, P_\alpha^c, O_c) = 1$.

Proof:

(⇒)

Suppose $n \in \mathcal{F}$ and $t \in S_c$. Then let $u = min(\{n, t\})$ and $v = max(\{n, t\})$.

Rule (d) (excluded middle) implies $\alpha_u \vee \sim\alpha_u$. Suppose $\sim\alpha_u$.

Then, repeated use of the rule of identity, (a), implies:

$$\sim\alpha_u \vdash \sim\alpha_{u+1} \vdash \sim\alpha_{u+2} \vdash \ldots \vdash \sim\alpha_v.$$

Now, $n = u$ or $n = v$. Thus, $\sim\alpha_n \mapsto T$ (with *top* certainty, since the certainty in classical RSs can only be *top*).

By Definition 5.2.7, $P_\alpha^c(n)$ is true, so that $\alpha_n \mapsto T$ (again, c can only be *top*).

By rule (b) (consistency), $\sim\alpha_n, \alpha_n \vdash$, so that state s is closed, contradicting the hypothesis.

Therefore, $\alpha_u \mapsto T$ (and $\sim\alpha_u$ is false).

Again, by the repeated use of the rule of identity:

$$\alpha_u \vdash \alpha_{u+1} \vdash \alpha_{u+2} \vdash \ldots \vdash \alpha_v.$$

Now, $t = u$ or $t = v$, so that α_t. Then, $P_\alpha^{top}(t)$ is true.

Therefore, the P_α^c-density of any subset of perceived states in S_c is 1. Thus, $hom(S_c, P_\alpha^c, Q_c) = 1$.

If $\mathfrak{F} = \emptyset$ then, by Proposition 5.2.1, the P_α^c-density of any subset of perceived states in S_c is 0, so that $hom(S_c, P_\alpha^c, Q_c) = 1$.

($\Leftarrow$)

Let $t \in S_c$, and $\alpha_t \mapsto T$.

By hypothesis, $hom(S_c, P_\alpha^c, Q_c) = 1$, so that:

$$1 - rad(\{ P_\alpha^c\text{-density of } X: Q_c(X)\}) = 1 \qquad \text{(Defn. 3.2.5).}$$

So: $$rad(\{ P_\alpha^c\text{-density of } X: X \subseteq S_c \wedge X \neq \emptyset\}) = 0 \qquad \text{(Defn. 5.2.1).}$$

Therefore, $max(\{ P_\alpha^c\text{-density of } X: X \subseteq S_c \wedge X \neq \emptyset\}) =$

$$min(\{ P_\alpha^c\text{-density of } X: X \subseteq S_c \wedge X \neq \emptyset\})$$

(Defn. 3.2.4).

Thus, there is some value, k, such that, for every X such that $Q_c(X)$ is true, the P_α^c-density of X is k.

Now, if $P_\alpha^c(t)$ is true, and $Q_c(\{t\})$, so that there is a Q_c-set with P_α^c-density 1. So the P_α^c-density of all Q_c-sets is 1.

Hence, the P_α^c-density of $\{t+1\}$ is 1 and $P_\alpha^c(t+1)$ is true.

This implies $\alpha_t \vdash \alpha_{t+1}$, which is the rule of identity. □

Proposition 5.4.2

Let $(\Xi_{Int}, \sqsubseteq, Orig, match)$ be a valid exploration set for a intuitionistic reasoning system , Int, $\alpha \in wff$ in the reasoning system, $(S_{Int}, <, s) \in \Xi_{Int}$ and $\Im$ be the set of first states for P_α^{ε} in S_{Int}. Then, given the validity of the rules (b), (c) and (d), listed in 4.7.2, the assumption of the validity of rule (a) is equivalent to the assumption $\forall\, n \in \Im\,.\, hom(S_{Int}, P_\alpha^{\varepsilon}, Q_{Int}^{n}) = 1$.

Proof:

($\Rightarrow$)

Let $n \in \Im$ and $t \in S_{Int}$, such that $n \leq t$.

Rule (d) (excluded middle) implies $\alpha_t \vee \sim\alpha_t \vee \div\alpha_t$. Suppose $\sim\alpha_t$.

By hypothesis, $P_\alpha^{\varepsilon}(n)$ is true, so that $\alpha_n \mapsto T$ (with *top* certainty, since this is the only value available in Intuitionistic *RS*s).

Then, repeated use of the rule of identity, (a), implies:

$$\alpha_n \vdash \alpha_{n+1} \vdash \alpha_{n+2} \vdash \ldots \vdash \alpha_t.$$

By Rule (b) (consistency), $\sim\alpha_t, \alpha_t \vdash$, so that state t is closed, contradicting the hypothesis.

Suppose $\div\alpha_t \mapsto T$.

Again, by the repeated use of the rule of identity, we have $\alpha_t \mapsto T$, and by rule (b) (consistency), $\div\alpha_t, \alpha_t \vdash$, which is a contradiction.

Therefore, $\alpha_t \mapsto T$, and $hom(S_{Int}, P_\alpha^{\varepsilon}, Q_{Int}^{n}) = 1$, which gives:

$$\forall\, n \in \Im\,.\, hom(S_{Int}, P_\alpha^{\varepsilon}, Q_{Int}^{n}) = 1.$$

($\Leftarrow$)

Let $t \in S_{int}$, and $\alpha_t \mapsto T$.

Suppose that there is no $n \in \mathfrak{Z}$ for which $n \leq t$.

Then let $T = \{ u: u \in S_{int},\ u < t \}$, and suppose $P_\alpha^{\,c}(u)$ for every $u \in T$. Then, by Definition 5.2.7, $T \leq \mathfrak{Z}$, contradicting the assumption that there is no $n \in \mathfrak{Z}$ for which $n \leq t$.

So, the state $v = max(\{ u: u \in T,\ P_\alpha^{\,c}(u)$ does not hold$\})$ is well-defined.

Then, if $P_\alpha^{\,c}(v+1)$ holds, $v+1 \in \mathfrak{Z}$ by Definition 5.2.7, and so $v+1 \notin T$. Then, by Definition 5.2.7, $v+1 \in \mathfrak{Z}$, and if $v \in T$ and $v+1 \notin T$, then, by construction of T, $v+1 = t$.

So $t \in \mathfrak{Z}$, which is a contradiction of the assumption.

Therefore there is some $n \in \mathfrak{Z}$ such that $n \leq t$.

By hypothesis, $hom(S_{int}, P_\alpha^{\,c}, Q_{int}^{\,n}) = 1$, so that:

$$1 - rad(\{ P_\alpha^{\,c}\text{-density of } X: Q_{int}^{\,n}(X) \}) = 1 \qquad \text{(Defn. 3.2.5).}$$

So: $rad(\{ P_\alpha^{\,c}\text{-density of } X: (X \subseteq \{ u: u \in S_{int} \text{ and } t < u \} \wedge X \neq \emptyset) \}) = 0$

(Defn. 5.2.2).

Therefore,

$$max(\{ P_\alpha^{\,c}\text{-density of } X: (X \subseteq \{ u: u \in S_{int} \text{ and } t < u \} \wedge X \neq \emptyset) \}) =$$

$$min(\{ P_\alpha^{\,c}\text{-density of } X: (X \subseteq \{ u: u \in S_{int} \text{ and } t < u \} \wedge X \neq \emptyset) \})$$

(Defn. 3.2.4).

Thus, there is some value, k, such that, for every Y such that $Q_{Int}^{n}(Y)$ is true, the P_{α}^{c}-density of Y is k.

Now, $P_{\alpha}^{c}(t)$ is true, and $Q_{Int}^{n}(\{t\})$, so that there is a Q_{Int}^{n}-set with P_{α}^{c}-density 1. So the P_{α}^{c}-density of all Q_{Int}^{n}-sets is 1.

$Q_{Int}^{n}(t+1)$ holds, so $\{t+1\}$ is a Q_{Int}^{n}-set.

Hence, the P_{α}^{c}-density of $\{t+1\}$ is 1 and $P_{\alpha}^{c}(t+1)$ is true. This implies:

$$\alpha_t \vdash \alpha_{t+1},$$

which is the rule of identity for intuitionistic *RS*s. □

We remind the reader that "$|\alpha|$" is used to abbreviate "either α or $-\alpha$".

Proposition 5.4.3

Let $(\Sigma_N, \sqsubset, Orig, match)$ be a valid exploration set for a non-monotonic *RS*, N, $\alpha \in wff$ in the reasoning system, $(S_N, <, s) \in \Sigma_N$, and:

$$Inp(S_N, \alpha) = \{ n \in S_N : \alpha_n \in ND(S_N) \}.$$

Then, given the validity of the rules (b) and (c) listed in 4.7.3, the assumption of the validity of rule (a) is equivalent to the assumption:

$$\forall n \in S_N .\ ((n \in Inp(S_N, \alpha) \vee n, n+1 \notin Inp(S_N, |\alpha|)) \Rightarrow hom(S_N, P_{\alpha}^{bot}, Q_N^{*}) = 1).$$

Proof:

(⇒)

Let n be a state in S_N.

Suppose $n \in Inp(S_N, \alpha)$.

Then $\alpha_n^{top} \mapsto T$, since only internally derived statements are marked *bot*. But, by Rule (a), if $\alpha_n^{top} \mapsto T$ then $\alpha_{n+1}^{bot} \mapsto T$ must hold, so $P_\alpha^{bot}(n+1)$ holds.

Now, $P_\alpha^{bot}(n)$ holds and $P_\alpha^{bot}(n+1)$ holds, so, by Definition 5.2.3, $hom(S_N, P_\alpha^{bot}, Q_N^*) = 1$.

Alternatively, suppose $n, n+1 \notin Inp(S_N, |\alpha|)$.

Then none of α_n^{top}, α_{n+1}^{top}, $\sim\alpha_n^{top}$ and $\sim\alpha_{n+1}^{top}$ maps to T, since only internally derived statements are marked *bot* (Propn. 4.7.3.1).

So, by rule (a), if $\alpha_n^{bot} \mapsto T$ then $\alpha_{n+1}^{bot} \mapsto T$, and conversely, if α_n^{bot} does not map to T then $\sim\alpha_n^{bot} \mapsto T$ (excluded middle), so that $\sim\alpha_{n+1}^{bot} \mapsto T$ and α_{n+1}^{bot} does not map to T. In either case, $hom(S_N, P_\alpha^{bot}, Q_N^*) = 1$.

Thus the assumption is proved.

($\Leftarrow$)

Let $n \in S_N$ such that $P_\alpha^{bot}(n)$ holds.

Suppose $\alpha_n^{top} \mapsto T$.

Then, since only statements that are not derived internally are marked *top* (Propn. 4.7.3.1), $n \in Inp(S_N, \alpha)$. So, by assumption, $P_\alpha^{bot}(n+1)$, and since this is true regardless of the particular S_N, it must be that $\alpha_n^{top} \vdash \alpha_{n+1}^{bot}$.

That is, $\alpha_n^{bot}, \alpha_n^{top} \vdash \alpha_{n+1}^{bot}$ and $\sim\alpha_n^{bot}, \alpha_n^{top} \vdash \alpha_{n+1}^{bot}$.

Suppose conversely that α_n^{top} does not map to T.

Then, $\alpha_n^{bot} \mapsto T$, since $P_\alpha^{bot}(n)$ holds. If $\sim\alpha_n^{top} \mapsto T$, then the rule of double negation allows the use of the second part of the rule derived

above, to obtain: $\sim\alpha_n^{top}, \alpha_n^{bot} \models \sim\alpha_{n+1}^{bot}$.

So, assume $\sim\alpha_n^{top}$ does not map to T. Then there is some S_M for which $n+1 \notin Inp(S_M, |\alpha|)$, and, by the assumption, $hom(S_M, P_\alpha^{bot}, Q_M^n) = 1$, so that $P_\alpha^{bot}(n+1)$ must hold.

Therefore, the rule that leads to this behaviour must be $\alpha_n^{bot} \models \alpha_{n+1}^{bot}$, when neither α_n^{top} nor $\sim\alpha_n^{top}$ maps to T.

Putting together these parts the final rule is:

$\alpha_n^c \models \alpha_{n+1}^{bot}$ when $c = max\{d : |\alpha|_n^d\}$,

which is the rule (a). □

Proposition 5.4.4

Let $(\mathfrak{E}_U, \sqsubset, Orig, match)$ be a valid exploration set for an uncertainty RS, U, using certainty structure $(C, <)$, $\alpha \in wff$ in the reasoning system, $(S_U, <, s) \in \mathfrak{E}_U$ and $\mathfrak{F}$ be the set of first states for P_α^c in S_U. Then, given the validity of the rules (b) and (c) listed in 4.7.4, the assumption of the validity of rule (a) is equivalent to the assumption:

(i) $\forall n \in \mathfrak{F}.\ hom(S_U, P_\alpha^c \vee R_{\sim\alpha}^c, Q_{Int}^n) = 1$;

(ii) $\exists (T_U, <, t).$ $((T_U, <, t)$ is an opp for U, containing only o-states$) \wedge$

$(\alpha_t^c \mapsto T,\ c = max\{d : |\alpha|_t^d\}) \wedge (\forall d \in C.\ hom(T_U, P_\alpha^d, Q_{Int}^t) = 1)$.

Proof:

(⇒)

Let $n \in \mathfrak{F}$ and $t \geq n$.

By Definition 5.2.7, if $t = n$ then $P_\alpha^c(t)$.

Suppose $t > n$ and $P_\alpha^c(t)$ does not hold. Then, without loss of generality, suppose t is the first state (in the order $<$) after n such that $P_\alpha^c(t)$ does not hold.

Now, $P_\alpha^c(t-1)$ holds, so $\alpha_{t-1}^{\ d} \mapsto T$ for some $d \geq c$ - assume this is the largest value of d for which $\alpha_{t-1}^{\ d} \mapsto T$.

Then rule (a) implies $\alpha_t^{\ d} \mapsto T$, unless $d \neq max(\{e : |\alpha|_n^{\ e}\})$. But, by supposition, $P_\alpha^c(t)$ does not hold, so assume $d \neq max(\{e : |\alpha|_n^{\ e}\})$.

Thus, there must be some value $e \geq d$ for which $\sim\alpha_{t-1}^{\ e} \mapsto T$ - assume this value is the maximum of all such values. Then rule (a) implies $\sim\alpha_t^{\ e} \mapsto T$, and thus $R_{\sim\alpha}^{\ c}(t)$ holds.

So, $hom(S_U, P_\alpha^c \vee R_{\sim\alpha}^{\ c}, O_{Inst}^{\ c}) = 1$, which proves (i).

Let $(T_U, <, t_0)$ be defined as follows (taking all integers as the set of indices for states in T_U and ordering the states by the order of their indices inherited from the integers):

$\forall$ integers, i, let $t_i = (\alpha^c \mapsto T)$.

Then the only statement that maps to T in any state in T_U is α^c. This is an opp containing only o-states and cannot become closed using the rules of U and $\forall d \in C$. $hom(T_U, P_\alpha^{\ d}, O_{Inst}^{\ c}) = 1$.

($\Leftarrow$)

Let $t \in S_U$, and $\alpha_t^{\ c} \mapsto T$, with $c = max(\{d : |\alpha|_t^{\ d}\})$.

Suppose that there is no $n \in \mathfrak{F}$ for which $n \leq t$.

Then let $T = \{u : u \in S_{Inst},\ u < t\}$, and suppose $P_\alpha^{\ c}(u)$ for every $u \in T$. Then, by Definition 5.2.7, $T \in \mathfrak{F}$, contradicting the assumption that there is no $n \in \mathfrak{F}$ for which $n \leq t$.

So, the state $v = max(\{u : u \in T,\ P_a^c(u)$ does not hold$\})$ is well-defined.

Then, if $P_a^c(v+1)$ holds, $v+1 \in \mathfrak{F}$ by Definition 5.2.7, and so $v+1 \notin T$. Then, by Definition 5.2.7, $v+1 \in \mathfrak{F}$, and if $v \in T$ and $v+1 \notin T$, then, by construction of T, $v+1 = t$.

So $t \in \mathfrak{F}$ which is a contradiction of the assumption.

Therefore there is some $n \in \mathfrak{F}$ such that $n \leq t$.

By assumption (i), $hom(S_0, P_a^c \vee R_{\sim a}^c, Q_{Ins}^n) = 1$, so that:

$1 - rad(\{P_a^c \vee R_{\sim a}^c$-density of $X: Q_{Ins}^n(X)\}) = 1$ (Defn. 3.2.5).

So:

$rad(\{P_a^c \vee R_{\sim a}^c$-density of $X: (X \subseteq \{u : u \in S_{Ins}$ and $t < u\} \wedge X \neq \emptyset)\}) = 0$

(Defn. 5.2.2).

Therefore,

$max(\{P_a^c \vee R_{\sim a}^c$-density of $X: (X \subseteq \{u : u \in S_{Ins}$ and $t < u\} \wedge X \neq \emptyset)\}) =$

$min(\{P_a^c \vee R_{\sim a}^c$-density of $X: (X \subseteq \{u : u \in S_{Ins}$ and $t < u\} \wedge X \neq \emptyset)\})$

(Defn. 3.2.4).

Thus, there is some value, k, such that, for every X such that $Q_{Ins}^n(X)$ is true, the $P_a^c \vee R_{\sim a}^c$-density of X is k.

Now, $P_a^c(t)$ is true, and $Q_{Ins}^n(\{t\})$, so that there is a Q_{Ins}^n-set with $P_a^c \vee R_{\sim a}^c$-density 1. So the $P_a^c \vee R_{\sim a}^c$-density of all Q_{Ins}^n-sets is 1.

$Q_{Ins}^n(t+1)$ holds, so $\{t+1\}$ is a Q_{Ins}^n-set.

Hence, the $P_\alpha^c \vee R_{\sim\alpha}^c$-density of $\{t+1\}$ is 1 and $P_\alpha^c(t+1) \vee R_{\sim\alpha}^c(t+1)$ is true.

By assumption (ii):

$\exists (T_U, <, t)$. $((T_U, <, t)$ is an *opp* for U containing only α-states) $\wedge$

$(\alpha_t^c \mapsto T,\ c = max\{d : |\alpha|_t^d\}) \wedge (\forall d \in C.\ hom(T_U, P_\alpha^d, Q_{|\alpha|}^t) = 1)$.

The rules which apply in S_U must also apply in T_U, so it must be that:

$$\alpha_n^c \models \alpha_{n+1}^c,\ (c = max(\{d : |\alpha|_n^d\}))\ .$$

If $t \in S_U$, and $\alpha_t^c \mapsto T$, with $c \neq max(\{d : |\alpha|_t^d\})$, and suppose $\alpha_t^* \mapsto T$ with $c = max(\{d : |\alpha|_t^d\})$. Then the assumptions still apply with α_t^* replacing α_t^c and the same rule is derived as above.

Finally, if $t \in S_U$, and $\alpha_t^c \mapsto T$, with $c \neq max(\{d : |\alpha|_t^d\})$, and $\sim\alpha_t^* \mapsto T$ with $c = max(\{d : |\alpha|_t^d\})$. Then the assumptions still apply with $\sim\alpha_t^*$ replacing α_t^c and the same rule is derived as above, with $\sim\alpha_t^*$ replacing α_t^c.

Therefore, the rule that is derived from the assumptions (i) and (ii) is $\alpha_n^c \models \alpha_{n+1}^c$, $(c = max(\{d : |\alpha|_n^d\}))$, which is precisely rule (a). □

Definition 5.4.1

Let n be a perceived state for an interval RS (cf. Section 4.7.5), and $\alpha \in wff$ in the reasoning system. Then the predicate $\Omega_\alpha(n)$ is defined to be true iff $\alpha \mapsto T$ in state n with a persistency counter ω.

Proposition 5.4.5

Let $(\Sigma_I, \sqsubset, Orig, match)$ be a valid exploration set for an interval RS, I, $\alpha \in wff$ in the reasoning system, $(S_1, <, s) \in \Sigma_I$ and $\mathfrak{F}$ be the set of

first states for $P_\alpha^{\,c} \wedge \Omega_\alpha$ in S_1. Then, given the validity of the rules (c) and (d) listed in 4.7.5, the assumption of the validity of rules (a) and (b) is equivalent to the assumption:

$$\forall n \in \mathfrak{F}.\ hom(S_1, P_\alpha^{\,c}, Q_1^{(n,1,K)}) = 1,$$

where:

if $f(\alpha, n) \in \mathbf{N}$ then $\exists u \in S_1.\ K = \{t \in S_1 : n \leq t \leq u\}$ and $|K| = f(\alpha, n) + 2$

and

if $f(\alpha, n) = \omega$ then $K = \{t \in S_1 : n \leq t\}$.

Proof:

(⇒)

Let $n \in \mathfrak{F}$. Then, by Definition 5.2.7, either $P_\alpha^{\,c}(n-1) \wedge \Omega_\alpha(n-1)$ does not hold, or $\forall t \in S_1.\ t \leq n \Rightarrow P_\alpha^{\,c}(t) \wedge \Omega_\alpha(t)$.

Suppose $f(\alpha, n) \in \mathbf{N}$. Then, by rules (b) and (c), if all the states in S_1 are open, $\alpha_{n+1} \mapsto T$ and has persistency $f(\alpha, n)$. Then, for the subsequent $f(\alpha, n)$ states $\alpha \mapsto T$, so that there is a continuous sequence of $f(\alpha, n) + 2$ states for which $\alpha \mapsto T$, starting from state n.

Suppose $f(\alpha, n) = \omega$. Then, by constraint on f,

$$\forall m \in S_1.\ (n < m) \Rightarrow f(\alpha, m) = \omega.$$

Therefore, by rule (b), $\forall m \in S_1.\ (n < m) \Rightarrow \alpha_n \mapsto T$.

It follows that $hom(S_1, P_\alpha^{\,c}, Q_1^{(n,1,K)}) = 1$.

($\Leftarrow$)

Let $t \in S_1$, and $\alpha_c \mapsto T$.

Suppose that there is no $n \in \mathfrak{Z}$ for which $n \leq t$.

Then let $T = \{u: u \in S_1, \ u < t\}$, and suppose $P_\alpha^c(u) \wedge \Omega_\alpha(u)$ for every $u \in T$. Then, by Definition 5.2.7, $T \subseteq \mathfrak{Z}$, contradicting the assumption that there is no $n \in \mathfrak{Z}$ for which $n \leq t$.

So, the state $v = max(\{u: u \in T, \ P_\alpha^c(u) \wedge \Omega_\alpha(u)$ does not hold$\})$ is well-defined.

Then, if $P_\alpha^c(v+1)$ holds, $v+1 \in \mathfrak{Z}$ by Definition 5.2.7, and so $v+1 \notin T$. Then, by Definition 5.2.7, $v+1 \in \mathfrak{Z}$, and if $v \in T$ and $v+1 \notin T$, then, by construction of T, $v+1 = t$.

So $t \in \mathfrak{Z}$ which is a contradiction of the assumption.

Therefore there is some $n \in \mathfrak{Z}$ such that $n \leq t$.

By hypothesis, $hom(S_1, P_\alpha^c, \Omega_1^{(n,1,K)}) = 1$,

where:

if $f(\alpha, n) \in \mathbb{N}$ then $\exists u \in S_1 . K = \{t \in S_1: n \leq t \leq u\}$ and $|K| = f(\alpha, n)+2$

and

if $f(\alpha, n) = \omega$ then $K = \{t \in S_1: n \leq t\}$.

So:

$1 - rad(\{P_\alpha^c - \text{density of } X: \Omega_1^{(n,1,K)}(X)\}) = 1$ (Defn. 3.2.5).

So: $rad(\{P_\alpha^c - \text{density of } X: \Omega_1^{(n,1,K)}(X)\}) = 0$.

Therefore,

$$max(\{\ P_\alpha^{\ c}\text{-density of } X\colon Q_1^{(n,1,K)}(X)\}) = min(\{\ P_\alpha^{\ c}\text{-density of } X\colon Q_1^{(n,1,K)}(X)\})$$

(Defn. 3.2.4).

Thus, there is some value, k, such that, for every X such that $Q_1^{(n,1,K)}(X)$ is true, the $P_\alpha^{\ c}$-density of X is k.

Now, $P_\alpha^{\ c}(t)$ is true, and $Q_1^{(n,1,K)}(\{t\})$, so that there is a $Q_1^{(n,1,K)}$-set with $P_\alpha^{\ c}$-density 1. So the $P_\alpha^{\ c}$-density of all $Q_1^{(n,1,K)}$-sets is 1.

If $Q_1^{(n,1,K)}(t+1)$ holds, then $\{t+1\}$ is a $Q_1^{(n,1,K)}$-set and hence, the $P_\alpha^{\ c}$-density of $\{t+1\}$ is 1 and $P_\alpha^{\ c}(t+1)$ is true.

This implies that $\forall u \in K - \{n\}.\ \alpha_{u-1} \vdash \alpha_u$ holds. Thus, there is a continuous sequence of $f(\alpha, n)+1$ states for which $\alpha_{u-1} \vdash \alpha_u$ holds.

Clearly, a persistency counter can be used to keep track of where a given state is in such a sequence. Thus, an index can be introduced:

$${}^{p}\alpha_{u-1} \vdash^{p-1} \alpha_u,\ (\omega \neq p > 0);$$

$${}^{\omega}\alpha_{u-1} \vdash^{f(\alpha,n)} \alpha_u.$$

These are the rules of identity, (a) and (b), for interval *RS*s. □

We now show how the assumptions listed in the propositions above can be placed in a partial order in the following way:

If we can show that, for any reasoning system, and any consistent perceived progression in that reasoning system, one of the assumptions of homogeneity in the above propositions implies a second, then the first is before the second in the order.

Conversely, if there is some reasoning system for which there is a consistent perceived progression in which one assumption holds, but a second assumption fails, then the second does not come before the first in the partial order.

There are some fourteen proofs required to prove the entire order, of which the majority (eight, in fact) are all very similar – the proofs that a particular assumption does not imply a second. The pattern for these is to take the particular reasoning system for which the rule of identity is equivalent to the assumption of homogeneity, and then consider some consistent progression which uses a particular feature of the reasoning system not compatible with the second assumption. This then leads to a proof that the second assumption cannot imply the first. Since these proofs are all very straightforward and not particularly instructive, we omit them.

It is also clear that the assumption of homogeneity in the first proposition (which when written in its most general form, for a reasoning system with a certainty structure, $(C, <)$, becomes: $\forall c \in C.\ hom(S_R, P_a^c, Q_c) = 1$) implies the assumptions used in all the other propositions.

For ease of reference, we use the name of the reasoning system for which an assumption has been proved equivalent to the rule of identity, as an abbreviation for each assumption used.

The full partial order is then as follows:

classical *RSs* ➞ intuitionistic *RSs* ➞ uncertainty *RSs* and interval *RSs*

and

classical *RSs* ➞ non-monotonic *RSs*.

All other pairs of reasoning systems in the group are incomparable.

We present here the proofs that intuitionistic *RSs* ⇒ uncertainty *RSs* and interval *RSs*.

For all reasoning systems, R, and for all originated perceived progressions containing only o-states, $(S_R, <, s)$, and for any statement, $\alpha \in wff$ in the reasoning system, which maps to T in some state of S_R with some certainty, c, the following holds:

if R uses the uncertainty structure $(C, <)$, $\mathfrak{Y}^c$ is the set of first states for P_α^c in S_R and $\mathfrak{Y}_\omega^c$ is the set of first states for $P_\alpha^c \wedge \Omega_\alpha$ in S_R then:

$(\forall c \in C. \forall n \in \mathfrak{Y}^c. \ \forall d \in C. hom(S_R, P_\alpha^d, Q_{int}^n) = 1) \Rightarrow$

(I) (i) $\forall c \in C. \forall n \in \mathfrak{Y}^c. hom(S_R, P_\alpha^c \vee R_{\neg\alpha}^c, Q_{int}^n) = 1;$

(ii) $\exists (T_R, <, t).$ $((T_R, <, t)$ a consistent opp for $R) \wedge$

$(\alpha_t^c. c = max(\{d : |\alpha|_t^d\})) \wedge (\forall d \in C. hom(S_R, P_\alpha^d, Q_{int}^n) = 1);$

(II) $\forall c \in C. \forall n \in \mathfrak{Y}_\omega^c. hom(S_R, P_\alpha^d, Q_1^{(n,1,K)}) = 1,$
where K is a continuous sequence of $f(\alpha, n)+2$ states in S_R, starting from n.

Proof:

Since $P_\alpha^c(n) \Rightarrow P_\alpha^c(n) \vee R_{\neg\alpha}^c(n)$ trivially, part (I)(i) follows.

Since $P_\alpha^c(n)$ holds for some $n \in S_R$, by hypothesis, then, by assumption, $\forall d \in C\ hom(S_R, P_\alpha^d, Q_{int}^n) = 1$, so that when $c = max(\{d : |\alpha|_t^d\})$ the opp $(S_R, <, n)$ satisfies (I)(ii).

Finally, since $\mathfrak{Y}_\omega^c \subseteq \mathfrak{Y}^c$, and $Un(Q_1^{(n,1,K)}) \subseteq Un(Q_{int}^n)$, it is clear that (II) follows from the assumption.

Now, the assumption is the general form of the assumption used in Proposition 5.4.2, and (I)(i) and (ii) are the assumptions used in Proposition 5.4.4, while (II) is the assumption used in Proposition 5.4.5.

Therefore, intuitionistic RSs $\Rightarrow$ uncertainty RSs and interval RSs. □

5.5 Analysis of the Results about Identity through Homogeneity

In the previous section we have proved the equivalence of the rules of identity, for the five classes of reasoning systems, to some expressions in homogeneity theory. We now explain why this is a worthwhile analysis and what we have learned from it.

Firstly, we have concentrated on the rules of identity because of the very particular role played by these rules. As we mentioned before in Section 4.4, the rules of identity are the foundations upon which every deduction must be made. This is so for two distinct reasons: the first is that the rules of identity propagate the existing knowledge through time, representing thus the logical memory of the system; the second, which is more subtle, is that any attempt at coping with a changing reality must begin with a modification of these rules. Their connection to any kind of deduction is then clear: we need premises inherited from the past in order to start deductions, and we must also make sure that any formula used in the inference will not change its semantic value while the deduction is carried through.

It is interesting and telling, however, that identity is always taken for granted, seldom discussed, and sometimes even left out as trivial from the formalisation of some systems. We think that there are two main reasons explaining this oddity, one historical-psychological and the other much more technical.

The historical-psychological explanation springs from the observation of the status gained by classical logics (expecially in the Frege-Russell form). While an enormous amount of work has been devoted to details of any sort (often without visible motivation), the foundations of the system have been challenged only through completely antagonistic conceptions, many of which lack the rigour of their opponent. This situation naturally creates polemics instead of analysis, and the clash of philosophies obscures and denies the need, in order to put the discussion on a proper footing, for a common formal base. Partly as a reaction against the damaging quarrels of the first part of this century and partly due to an unfortunate process of

mythification of classical logic (reminiscent of the analogous and disastrous authority achieved by Aristotelian syllogistics in the Middle Ages), the need for an open-minded analysis of essential features like the principle of Identity has been ignored.

On top of all this rests the well known psychological device by which the most complex (and basic) mental operations about the external reality are made to appear, to each of us, as extremely simple and shallow. The basic assumption of homogeneity, which we discussed in Chapter 2, is, perhaps, the most startling example of this mechanism.

The technical reason is that, in order to analyse such a fundamental part of any reasoning system, a tool is required which is more general and basic still. This tool should also be powerful enough to examine several reasoning systems at once, if the exercise is to have any meaning. It is clear that no competitor can be referee, so that a theory altogether different from the reasoning systems under consideration is needed. Given these requirements of generality, strength and fundamentality, very few candidates are left for the position. One possible candidate could be the Topoi theory [GO83], if it were not for its own underlying philosophy, which positions the theory itself in opposition to some of the logics to be examined.

We claim that homogeneity theory presents all the necessary characteristics of rigour and generality, while the plane of its philosophy automatically places it above any direct confrontation with the philosophies embedded in each reasoning system. This can happen because homogeneity theory does not provide a logical calculus of some kind, or a substitute for set theory: it expresses formally a point of view about how a vision of the world could be created, a problem which clearly precedes *any particular* definition of such a vision. This level of problem has been considered, until now, as an essential part of the realm of pure philosophy, and treated through the argumentative techniques characteristic of the field, with the unavoidable degree of confusion associated with that. As far as we know, homogeneity theory is the first formal model to occupy that ground.

There are three practical ways in which we have employed homogeneity theory in the analysis of the problem of identity. Firstly, to define the

screens and predicates we have discussed above; secondly, to prove that, for each class of reasoning systems, a particular combination of predicates and screen is equivalent to the rule(s) of identity, when conjoined to the remaining rules (Propositions 5.4.1, 5.4.2, 5.4.3, 5.4.4 and 5.4.5). Thirdly, we have shown how these combinations of predicates and screens form a partial order (Proposition 5.4.6).

A new understanding, in many directions, can be drawn from these results.

The first lesson is, obviously, about the principle of identity itself. Far from being the simple, self-evident observation it is widely believed to be, it has proved to embody a complex and multiform structure of assumptions. It can now be said with absolute confidence that the principle of identity represents the basic vision about the behaviour of the world (or of our knowledge of the world) through time.

Several points can be made in this regard: firstly, the principle of identity is *not* the same for all reasoning systems, even when the rules look identical (as for classical reasoning system and intuitionistic reasoning systems, or non-monotonic reasoning system and uncertainty reasoning systems). Secondly, the view of the world which it represents can be rather naive or very sophisticated. The classical reasoning systems, on one hand, force a pre-eminence of the "external world" over our knowledge of it, and claim that the only possible recepticle of knowledge is a divine one. This can be clearly seen by the analysis of the screen and predicate used in Proposition 5.4.1, which condemns to eternal collapse any knowledge which is not complete and perfect to start with.

All the other systems, each in their own way, consider the process of acquiring knowledge as part of the reality which they try to capture. The intuitionistic *RS*s do it by requiring absolute persistence from the moment of discovery: this is somehow a total reversal of the classical point of view, notwithstanding the apparent closeness of the formalisms; here in fact it is the knowledge which subordinates the external reality, which was free to assume any value before being known, but is frozen in that position from the moment of discovery on. It could be suggested that the harsh antagonism, often recorded, between classical and intuitionistic logicians, is due

to this clash of beliefs more than to the invoked differences over uncomputable functions.

The non-monotonic reasoning systems present, under the homogeneity analysis, yet another point of view: there, there is no belief whatsoever about the far future, the only requirement being about how a transition is achieved from one state to the next. This stand could be imagined as that of someone who has a (classically inspired) view of how the world *should* be, but is prepared to relinquish it if so told. In addition, nothing can be learnt from such a change, since this disillusioned classical logician *could not* give any meaning to a world that does not behave as it should.

The uncertainty reasoning systems represent a cumulative approach to knowledge, since the strength of our belief weighs more than the truth or falsity of the belief itself. Of course, this is true only up to a point, given the finite uncertainty structure which characterises uncertainty *RS*s. The need for a finite sequence of increasing indices is due to the *realistic* view of belief embodied in this reasoning system: believing here means "being prepared to act upon", and this clearly requires a final "point of action". This also corresponds to the human model, where it is not possible to increase beliefs indefinitely.

It is interesting to contrast this attitude with the "beliefs" represented in the reasoning systems based on infinite non-monotonic logics, like some forms of Fuzzy *RS*, where everything is no more than an opinion, which can be reversed without consequence.

An important point is that the concept of identity in the uncertainty *RS*s is a compromise between the classical and intuitionistic extremes: here, in fact, a world is recognised, in which, sooner or later, it will be necessary to act, but in the meantime an increase in knowledge is considered important enough to outweigh the discovery of falsities in the previously held knowledge. This concept of the balance between events and assumptions, level and meta-level is close to the ideas expressed in Section 3.0, and forms the technical backbone of the class of adaptive *RS*s.

The role of assumptions, analysis and prediction in identity is even clearer

in the class of interval *RSs*. This is the first class where a complete swap in the semantic value of a statement is acceptable (under certain circumstances) and yet the world, or our knowledge of it, is not assumed to be the meaningless random sequence postulated by non-monotonic *RSs*. This idea, that between the immobile eternal order and the complete chaos there is always space for finding complex, limited, uncertain but extremely useful patterns of regularity, is one of the main principles behind the adaptive *RSs*.

A much deeper understanding of the scope and role of the principle of Identity is not, anyway, the only return from our analysis. Just as interesting is the result by which a partial order is proved among the five classes of reasoning systems. It was known that deleting the rule of double negation transforms classical *RSs* into intuitionistic *RSs* (even if the process is not that simple, cf. Section 4.7.2), producing thus an order between them: having a general model against which to match several classes of reasoning systems, and doing so by the analysis of the most fundamental property of them all, Identity, is however a new standard, which shifts the form of competition among reasoning systems from the disputed argument to the formal proof.

One further gain is about the theory of homogeneity itself: our analysis of the Identity principle can be seen as a test of the strength of this theory. It is worth noting that, while the propositions themselves declare and prove the equivalence between the rules of identity and the respective assumptions of homogeneity, the rules are completely opaque, and none of the above analysis could have been suggested, let alone carried through, by the simple observation of the formal structures.

The point is that the theory of homogeneity makes absolutely explicit what the formalisation is made to disguise, that is, that no theory rests on itself, everything rests on some kinds of assumptions, and being aware of them can only lead to a better control of our own theories, and to an improved adaptability of those theories to the tasks they are created to face.

5.6 Interaction between RSs and their Environments

In order to serve a useful purpose, a reasoning system must have the machinery to interact with its environment. In this section we introduce the definitions of those structures we use to examine this interaction, modelling the environment and naming certain explicit behaviours that are of interest.

These definitions extend and expand the rudimentary examination of external sources of information which were used in Chapter 4, and led us to define non-derivable statements (Defn. 4.5.2).

An informal treatment of these definitions can be found in Section 5.7.

Definition 5.6.1

Let R be a reasoning system, with syntactic domain, Syn. Then $(Obs, <, o)$ is an *observation line*, iff Obs is a subset of wff, in R, totally ordered by $<$ and $\forall o' \in Obs \,.\, o < o'$. The elements of Obs are called *observations*.

Definition 5.6.2

If $ES = (\Xi_R, \sqsubset, S, match)$ is a valid exploration set for the reasoning system R, and $(T, <, t) \in \Xi_R$, then $T(match(T))$ is called the *perceived present in* T, $p(T)$.

Definition 5.6.3

If $ES = (\Xi_R, \sqsubset, S, match)$ is a valid exploration set for the reasoning system R, then the *life line* for ES, $l(ES)$, is the originated perceived progression: $(\{p(T) : (T, <, t) \in \Xi_R\}, <, p(S))$.

Definition 5.6.4

(i) If $ES = (\mathfrak{T}_R, \sqsubset, S, match)$ is a valid exploration set for the reasoning system R, and $O = (Obs, <, o)$ is an observation line then (ES, O) is a *potential interaction*. There is a natural order-preserving association of elements from $\mathfrak{T}_R$ with elements from Obs.

(ii) If $o' \in Obs$ is associated with $(T, <, t)$ then o' is *observed* between $(T, <, t)$ and its successor, $(T', <, t')$, in $\mathfrak{T}_R$.

(iii) If $o' = (\alpha^c)$, and $\alpha^d \mapsto T$ in perceived state $p(T')$, then o' is said to be *accepted* as α^d; it is *fully accepted*, if $c = d$.

(iv) The *acceptance set* for (ES,O), $A(ES,O)$, is defined as follows:

$A(ES,O) = \{\alpha_n^d : \alpha_n^c$ is observed and accepted as α^d between $n-1$ and n, for $n-1$ and n states in $l(ES)$, and state n is not closed$\}$.

Definition 5.6.5

A potential interaction, (ES,O), is an *interaction* iff $ND(T) \subseteq A(ES,O)$, where $l(ES) = (T, <, t)$.

It is an *active interaction* when $ND(T) \neq \emptyset$.

We will follow our previous ruling, and maintain that a reasoning system must accept any observations it does not specifically and explicitly modify or reject by its own formal specification.

Definition 5.6.6

If $ES = (\mathfrak{T}_R, \sqsubset, S, match)$ is a valid exploration set for the reasoning system R, then ES *collapses* if the life line for ES contains a closed state.

Definition 5.6.7

Let R be a reasoning system. Then R *reacts* after the sequence of observations, seq, if there is a second sequence of observations, seq', such that:

(i) if (ES, seq) is an interaction for R, then ES does not collapse;

(ii) if $(ES, seq.\alpha^{top}.\neg\alpha^{top})$, $(ES', seq.\alpha^{top}.\neg\alpha^{top}.seq')$, $(ES'', seq.\neg\alpha^{top})$ and $(ES''', seq.\neg\alpha^{top}.seq')$ are interactions for R, then:

$$A(ES', seq.\alpha^{top}.\neg\alpha^{top}.seq') - A(ES, seq.\alpha^{top}.\neg\alpha^{top}) \neq A(ES''', seq.\neg\alpha^{top}.seq') - A(ES'', seq.\neg\alpha^{top}).$$

Definition 5.6.8

Let R_1 and R_2 be reasoning systems. Then $R_1 \prec R_2$ iff for every sequence of observations, seq:

(i) if (ES_1, seq) and (ES_2, seq) are interactions for R_1 and R_2 respectively, then if ES_1 does not collapse, ES_2 does not collapse;

(ii) if R_1 reacts after seq then R_2 reacts after seq.

5.7 An Informal View of the Interaction Process

In Definition 5.6.1 we present the observation line, which will be used to model the sequence of information entering the reasoning system from its environment. Of course, the process of translation of the raw material that is received from the environment into comprehensible and useful statements is a complex one, but we ignore this problem for the present. For the purposes of this first analysis we assume that the translation has already been achieved.

As the reasoning system creates and recreates its perceived world - the view that is held by the system of the environment in which it operates - it perceives its own progress through the world. This progress is from perceived present to perceived present (Defn. 5.6.2), and we call the series of these perceived presents the life line (Defn. 5.6.3).

A potential interaction (Defn. 5.6.4) is obtained by pairing a valid exploration set with an observation line. The interaction can be seen as the process whereby the system creates its first view of the world, an observation is made and as a result, a new view of the world is created (which might be identical to the old view, if the observation is rejected). If the system accepts the observation, it appears in the perceived present immediately following the state in which the observation was first made, although possibly with a modified certainty (the use of this will be apparent when we discuss Adaptive *RS*s).

In Definition 5.6.5 we distinguish the true interactions from the absurd, so that a true interaction only finds non-derivable statements in its life line if it has observed them. An active interaction is one in which at least one non-derivable observation is accepted.

Collapse (Defn. 5.6.6) is a straightforward name for a particularly notable form of behaviour. It will be useful when we consider the behaviour of reasoning systems confronted with the test set by Definition 5.6.7, which is designed to consider whether a reasoning system can react to what is, at least intuitively, an apparent inconsistency in its environment. It will be seen later that most of the reasnning systems we have so far considered over-react in the face of this problem, while non-monotonic *RS*s do not react at all. When we construct the unforgiving *RS* and the adaptive *RS*, we will see the possibility of a more useful reaction.

5.8 Performances of the Reasoning Systems under Interaction

An informal explanation of the following result is presented in Section 8.9.

Proposition 5.8.1

Let (ES,O) be an interaction for the reasoning system, C, with $l(ES) = (S_c, <, s)$, then ES does not collapse iff $ND(S_c) = \emptyset$. (That is, there can be no non-derivable statement in the observation line).

Proof:

It has already been shown, in Theorem 5.4.1, that the validity of the identity rule for classical RSs within the $opp\ (S_c, <, s)$ is equivalent to the assumption $hom(S_c, P_\alpha^c, Q_c) = 1$, for each statement α in C.

If $\alpha_n \in ND(S_c)$ then α_{n-1} cannot hold, since otherwise α_n could be derived by the rule of identity. But then $hom(S_c, P_\alpha^c, Q_c) \neq 1$, and ES collapses. Therefore, if ES does not collapse, then $ND(S_c) = \emptyset$.

Conversely, if $ND(S_c) = \emptyset$, then every statement that appears in $l(ES)$ is derivable from the rules, so $l(ES)$ must be consistent for C. □

Note that any valid exploration set for a reasoning system, which starts in a consistent state and does not alter its perceived progression, must remain consistent. Thus if there are no non-derivable statements on the life line of the exploration set, then the reasoning system must remain consistent.

Proposition 5.8.2

Each of the five classes of reasoning systems, described in Section 4.7, except classical RSs, has a valid exploration set, ES, such that there is an active interaction, (ES,O), for which ES does not collapse.

By Proposition 5.8.1, there can be no active interaction for a classical RS which does not collapse.

Consider the following interaction:

For each integer, i, let $(T_i, <, t_{i,0})$ be defined as follows:

$T_i = \{ t_{i,j} : \text{for all integers } j \}$ so that $<$ is inherited from the order on the second index of $t_{i,j}$.

For $i = 0$ and for $i > j$, let $t_{i,j} = \{ \alpha^{bot} \mapsto T \}$ for uncertainty RSs and $t_{i,j} = \{ \sim\alpha \mapsto T \}$ for intuitionistic RSs.

For $1 \leq i \leq j$, let $t_{i,j} = \{ \alpha^{top} \mapsto T \}$, for intuitionistic RSs, and the same for uncertainty RSs, except $t_{1,1} = \{ \alpha^{top} \mapsto T, \alpha^{bot} \mapsto T \}$.

Let $ES = (\{ (T_i, <, t_{i,0}) : i \text{ an integer} \}, \sqsubset, (T_0, <, t_{0,0}), match)$ be a valid exploration set, with $\sqsubset$ inherited from the order on the index of T_i.

Let $O = (\{ \alpha_i^{top} : i \text{ an integer} \}, <, \alpha_1)$.

It is clear that (ES, O) is the active interaction in which α is not known before the origin and is observed and accepted between the origin and its successor. The life line of ES is trivially consistent for both of the reasoning systems.

The same example will serve for non-monotonic RSs, with the small modification that for $i = 0$ and for $i > j$, $t_{i,j} = \{ -\alpha^{bot} \mapsto T \}$, and, for $1 \leq i \leq j$, $t_{i,j} = \{ \alpha^{bot} \mapsto T \}$, except for $i = j = 1$, where $t_{1,1} = \{ \alpha^{top} \mapsto T, -\alpha^{bot} \mapsto T \}$

Finally, the example can be used for interval RSs, assuming $f(\alpha, t_{1,1}) = \omega$. □

Proposition 5.8.3

Of the five classes of reasoning systems described in Section 4.7, all except non-monotonic *RS*s react after the empty sequence of observations.

Proof:

In all the reasoning systems described, other than non-monotonic *RS*s, the acceptance of α^{top} leads to the inference that $\alpha^{top} \mapsto T$ in the immediately following state. Thus, the acceptance of $\sim\alpha^{top}$ in the next state will cause a contradiction to arise in that state, and the exploration set will collapse. This, in turn, implies that the reasoning systems, other than the non-monotonic *RS*, do not accept anything after the contradiction is observed.

However, if $\sim\alpha^{top}$ is fully accepted between the origin and the next state, then if no further observations are made the exploration set will not collapse (assuming, for classical *RS*s, that $\sim\alpha^{top}$ is fully accepted without collapse - thus, by Proposition 5.8.1, is derivable). It is therefore impossible for these reasoning systems to find an active interaction which repeats the same behaviour as this, while first fully accepting α^{top}. Non-monotonic *RS*s, on the other hand, do not react after the empty sequence.

For, suppose (ES, O_1), is an active interaction, for a non-monotonic *RS*, R, such that $l(ES)$ is consistent in R and $O_1 = (Obs, <, (\sim\alpha^{top}))$, where $\sim\alpha^{top}$ is observed and accepted between the origin of ES and its successor.

Then, if $ES = (\mathfrak{E}_R, \sqsubset, S, match)$, define $ES' = (\mathfrak{E}_R', \sqsubset, S', match')$ as follows:

$\forall (T, <, t) \in \mathfrak{E}_R \,.\, (T, <, t-1) \in \mathfrak{E}_R'$

and $(S', <, s') \in \mathfrak{E}_R'$ where:

$\forall s \in S.\, (s < p(S) \Rightarrow s \in S'$, inheriting the same order and with $s' > s)$
$\wedge (s > p(S) \Rightarrow (s - \{\alpha^{bot} \mapsto T\}) \cup \{\sim\alpha^{bot} \mapsto T\} \in S'$, inheriting the same order and with $s' < s)$

and $s' = p(S) \cup \{ \alpha^{top} \mapsto T \}$.

$match'(T) = match(T)$ and $match'(S') = [s']$.

Then (ES', O_g) is an active interaction for R, where:

$O_g = (Obs \cup \{ (\alpha^{top}) \}, <, (\alpha^{top}))$,

and α^{top} is observed and fully accepted between the origin of ES' and its successor and $\sim\alpha^{top}$ is observed and fully accepted between the second and third *opps* in ES', and, by construction, the acceptance sets for ES and ES' following the $|\alpha|$ observations must always be the same. □

Proposition 5.8.4

If O is an observation line and ES is a valid exploration set for an intuitionistic RS which does not collapse, then there is a valid exploration set, ES_M, ES_U or ES_I, for each of non-monotonic, uncertainty and interval RSs, respectively, which does not collapse, and for which the acceptance set, $A(ES_R, O)$, for each reasoning system, R, contains the acceptance set, $A(ES, O)$.

Proof:

If α^{top} is in the acceptance set, $A(ES, O)$, then it is not possible for $\sim\alpha^{top}$ to be derived from any of the other elements of the acceptance set, in case the life line of ES is inconsistent. Thus, $\sim\alpha^{top}$ cannot arise in the life line.

Therefore, an uncertainty reasoning system can be constructed for which $\sim\alpha$ maps to T until α^{top} is accepted, this being the same observation as was accepted from the observation set by the intuitionistic RS, and no contradiction can arise, since $\sim\alpha^{top}$ cannot be introduced as an observation in $A(ES, O)$.

The same is true of interval RS.

For non-monotonic *RS*s, that there is a valid exploration set in which all the states are consistent and α^{top} is fully accepted at the same state as in *ES* is trivial. □

Proposition 5.8.5

There is an observation line, *O*, and valid exploration sets, ES_u and ES_i, for uncertainty and interval *RS*s, respectively, which do not collapse, and such that $A(ES_u,O) = A(ES_i,O)$, but for which there is no valid exploration set, *ES*, for an intuitionistic *RS* which does not collapse, and for which $A(ES,O) = A(ES_u,O) = A(ES_i,O)$.

Proof:

Let

$$O = (\{(\alpha^c)_1, (\beta^c)_2, (\beta^c)_3, \ldots, (\beta^c)_{f(\alpha,t)+2}, (\neg\alpha^d)_{f(\alpha,t)+3}, (\beta^c)_{f(\alpha,t)+4}, \ldots\}, <, (\alpha^c)_1),$$

where the order $<$ is inherited from the index, and $d > c$.

Consider the exploration set, ES_u, for an uncertainty *RS*, with an uncertainty structure containing c and d, which starts with an origin in which all the states contain only "don't know" values and the rules, and which fully accepts all the observations in the above observation line. The exploration set does not collapse, since the greater certainty of $\neg\alpha$ in observation $f(\alpha,t)+3$ overrides the certainty of α.

Consider the exploration set, ES_i, for an interval *RS*, with origin *Orig*, in which all the states contain only "don't know" values and the rules, and $t = p(S)$, where $Orig+1 = (S, <, s)$, and which fully accepts all the observations in the above observation line. The exploration set does not collapse, since the persistency of α stops before observation $f(\alpha,t)+3$ is made.

Now, the acceptance sets for each of the interactions (ES_u,O) and (ES_i,O) are both active, with $A(ES_u,O) = A(ES_i,O) = O$.

Suppose that ES is a valid exploration set for an intuitionistic RS, which does not collapse and for which (ES,O) is an active interaction, with acceptance set $A(ES,O)$. Then, suppose $A(ES,O) = A(ES_u,O) = A(ES_1,O)$. It has already been shown that $A(ES_u,O) = A(ES_1,O) = O$, so $A(ES,O) = O$.

But then $l(ES)$ contains states for which α is true, (from the second state in $l(ES)$). Then, by the identity rule of intuitionistic RSs, α is true for all states subsequent to the first state for which $\alpha \mapsto T$. But, in the $(f(\alpha,t)+4)$th state, $\sim\alpha$ must be fully accepted, since it is an observation and $A(ES,O) = O$. This is a contradiction in intuitionistic RSs, and ES collapses, contradicting the assumption.

This concludes the proof of this result. □

Proposition 5.8.6

There is an observation line, O, and a valid exploration set, ES_u, for an uncertainty RS, which does not collapse, such that there is no valid exploration set, ES_1, for an interval RS, which does not collapse and for which $A(ES_u,O) = A(ES_1,O)$.

Proof:

Let $O = (\{(\alpha^c)_1, (\sim\alpha^d)_2, (\sim\alpha^d)_3, (\sim\alpha^d)_4, (\sim\alpha^d)_5, \ldots\}, <, (\alpha^c)_1)$, where the order $<$ is inherited from the index, and $d > c$.

Let ES_u be a valid exploration set for an uncertainty RS, with an uncertainty structure containing c and d, which starts with an origin in which all the states contain only "don't know" values and the rules, and which fully accepts all the observations in the above observation line. The exploration set does not collapse, since the greater certainty of $\sim\alpha$ in observation $f(\alpha,t)+3$ overrides the certainty of α. Therefore $A(ES_u,O) = O$.

Suppose that ES_1 is a valid exploration set for an interval RS, which does not collapse and for which (ES_1,O) is an active interaction, with acceptance set $A(ES_1,O)$. Then, suppose $A(ES_1,O) = A(ES_u,O)$. It has already been

shown that $A(ES_u, O) = O$, so $A(ES_l, O) = O$.

But then α is true in the second state of $l(ES_l)$, t, with persistency counter ω. Thus, α is true in in the third state of $l(ES_l)$, with persistency counter $f(\alpha, t)$, by the rules of identity for interval RSs. Then, if $A(ES_l, O) = O$, $\neg\alpha \mapsto \mathcal{F}$ in the third state of $l(ES_l)$, which, by the rule of Consistency for interval RSs, is inconsistent, contradicting the assumption that ES_l does not collapse. □

Proposition 5.8.7

There is an observation line, O, and a valid exploration set, ES_l, for an interval RS, which does not collapse, such that there is no valid exploration set, ES_u, for an uncertainty RS, which does not collapse and for which $A(ES_l, O) = A(ES_u, O)$.

Proof:

Let

$$O = (\{(\alpha^c)_1, (\beta^c)_2, (\beta^c)_3, \ldots, (\beta^c)_{f(\alpha,t)+2}, (\neg\alpha^c)_{f(\alpha,t)+3}, (\beta^c)_{f(\alpha,t)+4}, \ldots\}, <, (\alpha^c)_1),$$

where the order $<$ is inherited from the index.

Let ES_l be a valid exploration set for an interval RS, with origin $Orig$, in which all the states contain only "don't know" values and the rules, and $t = p(S)$, where $Orig + 1 = (S, <, s)$, and which fully accepts all the observations in the above observation line. The exploration set does not collapse, since the persistency of α stops before observation $f(\alpha, t) + 3$ is made. Therefore $A(ES_l, O) = O$.

Suppose ES_u is a valid exploration set for an uncertainty RS, with an uncertainty structure containing c, which starts with an origin in which all the states contain only "don't know" values and the rules, and which accepts all the observations in the above observation line.

Then, by the rule of identity for uncertainty RSs, α^c is true in the second

state of $l(ES_u)$ and in all subsequent states until overridden by a statement $|\alpha|$ with higher certainty. But there is no such statement internally derivable and there is no such statement in the observation line. However, $\sim\alpha^c$ is in the observation line, and since α^c is not overridden, the two statements must appear in the same perceived present in $l(ES_u)$, at the state when $\sim\alpha^c$ is accepted. This is a contradiction by the rule of Consistency of uncertainty *RSs*, which implies that ES_u collapses, contradicting the assumption. □

Proposition 5.8.8

Under the order of inclusion of possible acceptance sets, there is a partial order imposed on classical, intuitionistic, uncertainty and interval *RSs*, so that:

classical *RSs* < intuitionistic *RSs* < uncertainty *RSs* and interval *RSs*,

(the final pair being incomparable).

Proof:

By Proposition 5.8.1, the classical *RSs* can only perform inactive interactions, which all the other reasoning systems can do trivially, while Proposition 5.8.2 shows that all the other reasoning systems can actually perform active interactions.

Thus:

classical *RSs* < intuitionistic *RSs*, uncertainty *RSs* and interval *RSs*.

Proposition 5.8.4 proves that:

uncertainty *RSs* and interval *RSs* are not less than intuitionistic *RSs*.

By Proposition 5.8.5, combined with the above result,

intuitionistic *RS*s < uncertainty *RS*s and interval *RS*s.

Finally, the results of Propositions 5.8.6 and 5.8.7 show that the uncertainty *RS*s and interval *RS*s are incomparable in this order. □

All the reasoning systems explored, except the non-monotonic *RS*, react in precisely the same way when they do react. Thus, it is easy to see that the result of Proposition 5.8.8 implies, trivially, that:

classical *RS*s < intuitionistic *RS*s < uncertainty *RS*s and interval *RS*s,

while the last two are incomparable.

Since non-monotonic *RS*s do not react at all after the null sequence, for which none of the reasoning systems collapse (Prop. 5.8.3), it is clear that they do not succeed any of the other reasoning systems in the ordering. However, the non-monotonic *RS*s also do not precede any of the other systems, since they do not collapse after any sequence while the others each have a sequence that causes them to collapse.

Thus non-monotonic *RS*s are incomparable with the others in this ordering.

5.9 Analysis of the Comparison of RS's by Interactive Power

In the previous sections we present the formal definitions and propositions which form a second comparison of the classes of reasoning systems, this time from the point of view of the interaction with a second agent.

As we mentioned before, we do not see reasoning systems as abstract algebras, whose values reside in the beauty and richness of their structures: reasoning systems are primarily instruments by which we try to understand our view of the world and improve our control of it. Their mathematical properties are then to be matched against these yardsticks, rather then

The A.I. researcher, who tries to construct and use formal instruments which can simulate complex real life behaviours, has to add to the list of requirements which a reasoning system is expected to satisfy, the ability to deal with much more confused, corrupted and inconsistent information than any logician of the past generation would have ever had to face.

The term "real life problem" is in itself subject to confusion and misunderstanding, so it is necessary to build a formal frame in which the performance of different reasoning systems can be checked and measured or, at least, orderd according to some precise procedure. This is the role played by the definitions of Section 5.6. The Propositions 5.8.1 to 5.8.8 then use this formal frame to procede in the evaluation of bebaviours of the different reasoning systems.

Because of the range of problems that an A.I. reasoning system has to face, and also for reasons of theoretical generality, we have not considered any particular kind of application, but have concentrated on the way in which the simple existence of an interaction can affect the two more essential features of any reasoning system: its survival ability and its capability and range of reaction.

We have already pointed out how important it is to detect the conditions under which a reasoning systems collapses: it is interesting to note that, until recently, this problem was of almost exclusive concern to philosophers of mathematics, since the only practical users of a logic were the working mathematicians, perfectly able to organise the "collapsing and reconstructing" process in their mind, without any explicit formalisation. The need of a complete simulation in a machine has changed all that, sbedding light on the fact that the survival problem is absolutely inherent to the formal system itself, and that it has been possible to avoid the recognition of this feature only by a surreptitious introduction of human behaviour into the system. An analogous process has already taken place in linguistics, where components of the speaking process, considered automatic and trivial, have revealed an unexpected complexity once a proper simulation has been attempted.

This is an additional bonus from the parallel and dual processes of applying formal methods to A.I. research, and testing A.I. models against formal simulations of real life conditions, which should not be underestimated.

Another advantage of an abstract model for (elementary) interaction is that the nature of that interaction can be left completely undefined: it could be human to human, machine to user, machine to machine or entity to sensorial data. The same constraints about applicability, survival and reaction of the reasoning system would apply in each case.

Our analysis is then based on the apparently obvious principle of taking the logical rules seriously, and requiring that any mechanism for surviving an inconsistency be expressed consistently in the formalisation, not hinted at outside the system. Some of the results of this method may be disturbing, like that of Proposition 5.8.1, proving that the class of classical *RS*s can survive only if no active interaction takes place. On the other hand, it is only by applying principles and rules of each reasoning system to itself, and carrying the process to the extreme consequences, that we can learn the limits and strengths of each approach, and advance toward a better founded and more realistic model.

It must be said, however, that we do not intend our proofs about the limits of the existing classes of reasoning systems as a statement about the wisdom of the creation of their rel;ated logics, not only for the obvious reason of their role as necessary steps in the history of the field, but also because most of them had never been intended for use outside their original scope, inside which they maintain the same effectiveness. For example, classical logic was devised as an instrument to justify mathematical proofs, and in that sense it remains valid by our standards as well, since a proof (from known rules and already proved theorems) does take place in an immutable (though finite) world where every necessary piece of knowledge is already present and no active interaction is requested.

In the same way, the intuitionistic logic was constructed for formalising the process of increase of knowledge in mathematics (strictly, in the mind of each mathematician). Again, granting allowance for false theorems believed

true (it does happen even to the best mathematicians), and ignoring the problem of actual mathematical creativity, the logic satisfies its original requirements.

Giving to the reasoning systems examined their due credit, we can nevertheless claim that our analysis represents a useful clarification in three distinct directions: in the realm of pure logic, it shows what kind of hidden assumptions are present behind the apparently neutral face of some rules, and reveals limits of application and restrictions of philosophical meaning which can only contribute to a better understanding of the field; in the area of existing A.I. activity, it is hoped that our analysis will contribute towards slowing down the unfortunate existing trend of associating languages, programs and even implementations *tout court* with logics which where not designed for the purpose and are absolutely inadequate for it. Finally, having examined other systems under these constraints prepares the ground for the construction of our proposed class of reasoning systems, Adaptive *RS*s, which borrows heavily from the best scoring features of the other reasoning systems, trying at the same time to avoid inheriting into the same shortcomings.

As we have already stated, we believe that the formally expressed power of a reasoning system must be the yardstick against which the reasoning system is measured. The same principle has been extended in the construction of the frame needed in order to carry out the analysis (cf. Defn. 5.6.2, 5.6.3 and 5.6.6). The idea is that the same rules that a reasoning system applies to the world in order to make sense of it (in its perceived progression) are then applied to the reasoning system itself, in its passage from one view to the next. This concept is, at the same time, very natural, easily formalisable and controllable, and extremely effective: which thing, it must be said, speaks strongly in favour of the internal coherence of the systems.

The propositions can now be considered each in turn.

Proposition 5.8.1 states that the communication with a system based on a classical *RS* can only be one-way, in the sense that the system is prepared to be told only what it already knows. The apparent absurdity of this

conclusion, and the right historical perspective needed in order to understand it, have been discussed above.

It is interesting to note that *all* the results would hold even if the reasoning systems were provided with a possibility of refusing or modifying information. This possibility is not, however, conceived in any of the five classes considered, and rightly so, since this feature is a rather dangerous one, which can turn a reasoning system system into an obsessive solipsist, if applied without due precautions.

The central point is that information should be refused (or modified) only if the source is unreliable (or less reliable than another one), while the construction of a picture of the world, explaining and controlling most of the information, must be pursued. On top of all this, a refusal or modification of information must also produce a meta-change inside the reasoning system, so that a lesson is learnt out of what should be considered as an anomalous event. It is clear that none of the five classes of reasoning systems are in the least equipped to satisfy such a complex requirement, hence the impossibility of a last-moment insertion of the refusal feature. The class of adaptive *RSs*, on the other hand, has the formally expressed power to deal with this requirements, as it is shown in the next sections.

Proposition 5.8.2 simply shows that the other four classes of reasoning systems can survive an active interaction. As for several of these proofs, this is done by exhibiting one example of such interaction.

In Proposition 5.8.3 it is proved that non-monotonic *RSs* do not react, in the sense specified in Definition 5.6.7. This means that non-monotonic *RSs* can go through the same problem again and again and never learn anything from it. This concept is tightly linked to that of *ability to predict*. Predictions about the future are an essential feature of every cognitive system connected to the idea of science (other cognitive systems of a contemplative kind have been constructed, particularly in the East). On the other hand, a system which does not take notice of its own errors in forecasting the future can hardly be said to produce any prediction; at most, it holds apparently reasonable opinions, which it is paradoxically

prepared to relinquish on request but not to use in order to improve its own "beliefs".

The same proposition shows that all the other four classes do react, but in a very drastic manner. This could lead to the conclusion that a non-monotonic *RS* exhibits here a more reasonable behaviour, since at least it manages to survive. It is our contention, though, that this is not the case, because the two behaviours are incommensurable: survival is needed in order to continue to function, but on the other hand the central function to be saved is exactly being able to produce good predictions or, at least, learn from one's own mistakes.

The result proved in Proposition 5.8.4 highlights the order between intuitionistic *RS*s and the classes of uncertainty and interval *RS*s. It shows that the interactions, which can be performed without collapse by intuitionistic *RS*s, can be carried out with equal success by uncertainty and interval *RS*s. This result is complemented by Proposition 5.8.5, in which it is shown that the converse does not hold for either of the uncertainty or interval *RS*s in comparison with intuitionistic *RS*s. Together, these results show that uncertainty and interval *RS*s strictly succeed intuitionistic *RS*s in an order naturally induced by interactive power, as shown in Proposition 5.8.8.

That same order is further developed in the following two results (Propns. 5.8.6 and 5.8.7), which reveal the fact that uncertainty and interval *RS*s both have relative advantages and disadvantages, so are incomparable. The differences between the approaches of the two reasoning systems and their merits and demerits can be seen in the following example.

Suppose a machine equipped with an uncertainty *RS* is used as an investment advisor, and it is informed that, on a particular day, it is believed with a certain degree of certainty that the dollar is on an upward trend. If this certainty is not *top* then, when the investor asks whether dollars would be a wise investment, the machine will be cautious and suggest delay. If later the investor informs the machine that it is now certain that dollars are entering a crisis, then the machine will have been vindicated, and it can adjust its certainty about the trends of the dollar

accordingly.

Imagine the same situation had confronted an interval *RS* machine. To this machine, certainty is meaningless, so it would advise the investor that dollars are a good bet that day. If the dollar slumps, not only will the investor lose the investment, but the machine will be unable to adjust to the new information, and will enter a collapse.

Now imagine a rather different scene: if the investor informs the machine with an uncertainty *RS* that it is certain that the dollar is on an upward trend, then the machine will happily advise investment in dollars. Suppose, a month later, when the machine is still advising investment in dollars, the investor informs the machine that in fact dollars are suffering a decline - then the machine fails and collapses.

However, the interval *RS* machine fares rather better in this case, for, knowing that the persistency of trends in currency markets last only over a period of a few days, it stops advising the investor to buy dollars long before the investor learns of the dramatic fall in the dollars' worth. The machine willingly accepts the new information, and assigns a period of a few days persistency, during which it will treat dollars with a healthy disinterest.

It can be seen that the first machine has the advantage of the methodical researcher, which accumulates evidence and acts slowly, while the second machine recognises, or gambles on the existence of, (possibly short) patterns of regular behaviour, which, as everybody's daily experience confirms, often occur.

Proposition 5.8.8 proves that there is a partial order among four of the five classes, which is induced from the inclusion relation over acceptance sets. non-monotonic *RS*s have not been considered in this result because of their divergent behaviour in relation to reactions. As we pointed out before, the extreme reaction of collapsing and the absence of reactions at all are incomparable behaviours, since the opposite of both is clearly necessary.

It is interesting to note that the partial order so obtained is embedded in the one enforced by our analysis of assumptions of identity (Proposition 5.4.6). The embedding is that much more startling since the instruments used are very different, and so are the points of view embodied in them: on one hand, formal rules and homogeneity theory, expressing the view of the world embedded in each reasoning system; on the other hand, interaction, acceptances and reactions rendering the way in which those views can adapt to external contacts.

With this remark we conclude our comparative examination of five classes of reasoning systems: the model elaborated for this purpose can now be further developed in order to define the class of reasoning systems, adaptive *RS*s, which we want to introduce.

Chapter 6

A Model of Consistency Recovery for Adaptive Reasoning

6.1 Introduction

In this chapter we introduce a model for a basic part of what we call *adaptive* reasoning. We refer to this part as a *consistency recovery* mechanism. In fact, this is a shorthand for a rather complex behaviour, which involves the ideas of an "acceptable frequency of contradictions", and a "sufficient proportion of incoming information believed". The concepts of reasoning system, positive use of inconsistencies, capacity of reaction and survival, sources management, persistency, homogeneity, and individuals, all of which have been met before, are also used.

We have already stressed how capacity of reaction and survival must be two essential features of the way in which a reasonable model deals with inconsistencies.

In Section 6.3 we introduce the *unforgiving* model. This model is already stronger than the ones presented before, with the exception of interval *RS*, which could be, however, easily embedded in it. Propositions 6.5.1 and 6.5.2 prove the dominant position of the unforgiving *RS* in the order obtained above. The main function of this model is as an intermediary step toward the construction of the simplified model for adaptive *RS*, which we present in Sections 6.7, 6.8 and 6.9 (with a functional specification in Appendix A).

In Section 6.2 we prepare the ground for the definitions of the unforgiving and adaptive models.

6.2 Initial Structure

The following definitions are used to support the unforgiving and adaptive models presented in this chapter. An explanation of their use and meaning is given at the end of this section.

Definition 6.2.1

A *content* is an n-tuple $(n > 1)$ in which the first two coordinates are an *opp* and a set of formulae (possibly empty), for the same reasoning system, respectively.

Definition 6.2.2

Let $\mathfrak{L}$ be a set of contents of the same type, totally ordered by $\sqsubset$, with a minimum element under $\sqsubset$, *Orig*, and let *match* be the order-preserving embedding function as follows:

(i) $match: \{ \pi_1(opp): \exists l \in \mathfrak{L}.\, \pi_1(l) = opp \} \rightarrow \{ [s]: s \in \pi_1(\pi_1(Orig)) \}$;

(ii) $match(S) = [s]$, where $\pi_1(Orig) = (S, <, s)$.

Then $(\mathfrak{L}, \sqsubset, Orig, match)$ is a *life span*.

Definition 6.2.3

Let $(\mathfrak{L}, \sqsubset, Orig, match)$ be a life span.

Then the *link* for $(\mathfrak{L}, \sqsubset, Orig, match)$ is the function defined as follows:

(i) $link: \mathfrak{L} \rightarrow \mathfrak{L}$;

(ii) $\forall l \in \mathfrak{L}.\ link(l) = l+1$;

where $l+1$ is the immediate successor of l in the order $\sqsubset$.

Conversely, if $link: \mathfrak{L} \to \mathfrak{L}$, then $link$ *defines* $(\mathfrak{L}, \sqsubset, Orig, match)$, where $\forall l \in \mathfrak{L}.\ l \sqsubset link(l)$ defines $\sqsubset$, $Orig$ is the minimum element in $\mathfrak{L}$ in this order, and $match$ is constructed in the usual way.

Definition 6.2.4

Let $LS = (\mathfrak{L}, \sqsubset, Orig, match)$ be a life span, $O = (Obs, <, o)$ be an observation line, and (ES, O) be an interaction, where $ES = (\mathfrak{E}, \sqsubset, Or, match')$, in which:

$\mathfrak{E} \subseteq \{ \pi_1(cont) : cont \in \mathfrak{L} \}$, $\sqsubset$ is inherited from the order on $\mathfrak{L}$, Or is the *opp* in $Orig$, and $match'$ is defined in the usual way.

Then LS *contains* the interaction (ES, O), and each member of $\mathfrak{E}$ is called an *ob-state*.

Definition 6.2.5

Let $\theta = \{ k, k+1, \ldots, k+n-1 \}$ be a continuous sequence of n ob-states. An *θ-persistency* for the formula, α^c, $((\theta, \alpha, c)per)$ is a continuous subsequence of θ, $\{ m, m+1, \ldots, m+l-1 \}$, such that:

$$\neg P_{\alpha}{}^{c}(p(m-1)) \ \wedge\ \neg(\wedge\{ P_{\neg\alpha}{}^{c}(p(x)) : x \in \{ m, m+1, \ldots, m+l-1 \} \}).$$

$(\theta, \alpha, c)per$ is *headed* if $P_{\neg\alpha}{}^{c}(p(m+l))$ and

it is *topped* if $m+l-1 = k+n-1$.

In Definition 6.2.1 we introduce contents, which are the building block for the body of our reasoning systems. They form internal states, which keep

track of the current formation of the system. In fact, all the other reasoning systems would use the same basic structure for a complete description of their operation, but their internal structure is so much simpler than the models we present here, that there is no confusion created in having left out these implementation oriented details.

The life span that is defined in Definition 6.2.2 and link of Definition 6.2.3 are used in the machinery that drives these models. In fact, as will be seen later, the machinery is used explicitly only in the unforgiving model, to indicate the way in which it is achieved. The adaptive model would follow the same pattern, but be considerably more laden with detail, so it has been ommitted.

In Definition 6.2.4 we identify two pieces of terminology which will be convenient in the construction of the models, and Definition 6.2.5 is a technical construction, used in the specification of a function in the adaptive model. Its role is to give a name to an unbroken series of states, in a sub-sequence of a life line, with a particular property - that of satisfying P_a^{ϵ}.

6.3 The Unforgiving Model: Definitions

The following definitions are explained in Section 6.4, although a more detailled account of their meaning can be found in Section 6.6, after the model has been explored in Section 6.5.

The model divides into four categories of function: the object level function, the acceptance function, the meta-level function and the book-keeping functions. These will each be presented in a separate definition.

In all the following definitions $(C, <)$ is the uncertainty structure with $C = \{0, \ldots, top\} \subseteq \mathbb{N}$ and the order $<$ inherited from the natural numbers.

Definition 6.4.1

The object level rules for the unforgiving model are as follows:

(i) $\alpha_n^c \vdash \alpha_{n+1}^c$, $\qquad (c = max\{d : |\alpha|_n^d\})$;

(ii) $\alpha_n^c,\ \sim\alpha_n^c \vdash \blacktriangleright\{\alpha_n^c,\ \sim\alpha_n^c\}$, $\qquad (c \neq bot)$.

The symbol "$\blacktriangleright$" is called *shift*.

Definition 6.4.2

The object level functions for the unforgiving model are as follows:

(i) $Obj : \{opps\} \times \text{Form} \rightarrow \{opps\} \times \mathbb{P}(\text{Form})$;

(ii) $Obj(opp, \alpha^c) = (opp', \Theta)$;

where opp' is opp, extended by adding α^c to the statements that hold in $p(\pi_1(opp))$ and applying the transformation rules to build the entire perceived progression, if this does not lead to a shift. Then $\Theta = \emptyset$.

If the originated perceived progression obtained by the above procedure leads to a shift, with $\{\alpha_n^c, \sim\alpha_n^c\}$, then $\Theta = \{\alpha^c\}$ and opp' is built by removing from opp all those statements which include α^c or $\sim\alpha^c$.

Definition 6.4.3

The book-keeping functions for the unforgiving model are as follows:

(i) $down_k$: Sources $\rightarrow \mathbb{N}$;

(ii) sou_k : Form $\rightarrow$ Sources;

where k is an opp.

Definition 6.4.4

The interaction function for the unforgiving model are as follows:

(i) acc : Form × Sources × $\{ opps \}$ → Form ;

(ii) $acc\ ((\alpha^c)_k, u, k) = \alpha^d$;

where $d = max(\{ 0, c - down_k(u) \})$.

Definition 6.4.5

The meta-level function for the unforgiving model is as follows:

(i) ml: $\mathbf{P}(\text{Form}) \times \{ opps \} \rightarrow \{ down_k : k \text{ an } opp \} \times \text{Form}$;

(ii) $ml(\{\alpha^c\}, k) = (down_1, \alpha^0)$, if $sou_k(\alpha^c) = sou_k(-\alpha^c)$

$= undefined$, otherwise ;

where $\forall u \in \text{Sources} - \{ sou_k(\alpha^c) \}$. $down_1(u) = down_k(u)$
and $down_1(sou_k(\alpha^c)) = down_k(sou_k(\alpha^c)) + 1$.

(In fact, $link$ defined below will force $l = k+1$).

Definition 6.4.6

Let $O = (Obs, <, o)$ be an observation line. Let $\mathfrak{L}$ be a set of contents:

$\mathfrak{L} = \{ (opp, \Theta, sou_{opp}, down_{opp}) \}$.

The function $link'$ is then defined as follows:

(i) $link'$: $\mathfrak{L} \rightarrow \mathfrak{L}$;

(ii) $link'((k, \Theta, sou_k, down_k)) = (\pi_1(x), \pi_2(x), sou_{k+1}, down_{k+1})$,
if $\Theta = \emptyset$,

$= (\pi_1(y), \pi_2(y), sou_{k+1}, down_{k+1})$,
if $\Theta \neq \emptyset$,

where $x = Obj'(k, acc((\alpha^c)_k, u, k))$ and $((\alpha^c)_k, u)$ is observed between contents with *opps*, k and $k+1$ ("+" being in the order on contents), such that:

$$sou_{k+1} = sou_k \cup \{\alpha^c \mapsto u\}, \text{and } down_{k+1} = down_k;$$

and

where $y = Obj'(k, \pi_2(z))$, $z = ml(\Theta, k)$, $sou_{k+1} = sou_k$ and $down_{k+1} = \pi_2(z)$.

$link'$ defines the life span, $(\mathfrak{L}, \sqsubset, Orig, match)$, and contains the interaction, (ES, O), where $ES = (\mathfrak{E}, \sqsubset, Or, match')$, in which:

$\mathfrak{E} = \{k: (k, \Theta, sou_k, down_k) \in \mathfrak{L} \text{ and } \Theta = \emptyset\}$, $\sqsubset$ is inherited from the order on $\mathfrak{L}$, Or is the *opp* in $Orig$, and $match'$ is defined in the usual way.

6.4 Explanation of the Unforgiving Model Definitions

In Definition 6.3.1, the most significant transformation rules of the system are given. The most interesting part of this is the occurrence of shift, which is actually used as a message to the rest of the system that a contradiction has occurred and must be resolved.

The transformation rules are applied by the object level function introduced in Definition 6.3.2.

The book-keeping functions for the system are explicitly defined in Definition 6.3.3. These are functions which are indexed by the current opp, in order that they can be updated as the system regenerates its views of the world.

The observations are introduced to the rest of the system through the acceptance function (Defn. 6.3.4), which applies any downgrading of information that is necessary due to an unreliable source (or a source believed to be unreliable, at least).

The control of the system resides in the "meta-level" function introduced in Definition 6.3.5. This function is not actually as high in its observation and control of the system as several of the functions in the adaptive model. Nevertheless, the function is still the first one explicitly defined which moderates the behaviour of the entire system in the level below it, which is why we have granted it the name "meta-level".

Finally, in Definition 6.3.6, the complete mechanics of the system are exposed, by defining the link function which ensures the correct passage from internal state to internal state, and provides the communication link between all the other functions.

6.5 The Interactive Power of the Unforgiving Model

We now show how the unforgiving model for a reasoning system, *URS*, fits into the order we explored in Chapter 5. Although the unforgiving model has been defined with scope for several sources to be differentiated, we will assume throughout that there is only a single source. In an improved and more sophisticated model it would be desirable, indeed essential, to deal with contradictions between different sources, which is why provision has been left for recognising and recording the performance of several sources.

The reason we have not attempted to solve the problem of contradictions

between two sources is that to solve it satisfactorily requires a reasoning power which it is not the purpose of this work to consider. That this is the case can be seen in the following example: suppose one political party informs the electorate that it is both possible and desirable to defend the country which it proposes to govern without a nuclear arsenal, while an opposing party informs the electorate, with apparently equal conviction, that no country can consider itself adequately defended without a nuclear arsenal. It is not practical to completely remove both statements from those believed, so that one is left not believing one thing or the other, nor is it necessarily the case that both sources are equally unreliable. Thus, there cannot be any automatic decision about the relative merits of the statements made by the two parties, but a reasoned response based on other evidence and previous performance of the two sources must be made.

Reasoning power is not within the scope of this work, which is aimed at the problems of the most fundamental management of information, and the problem highlighted by the previous example is not approached in the unforgiving model, or later, in the adaptive reasoning system.

Proposition 6.5.1

There is no sequence of observations which causes the generated *ES* for the unforgiving model for a reasoning system to collapse.

Proof:

Suppose a formula, α^c, is mapped both to T and to F in some state in the life line of an exploration set for a *URS*. Then, by the standard rules of negation, $\alpha^c \mapsto T$ and $\sim\alpha^c \mapsto T$, which means that when the object level transformation rules are applied to the state, rule (ii) of Definition 6.4.1 will cause a shift to be generated.

By Definition 6.4.2, a new *opp* will then be created, in which α^c and $\sim\alpha^c$ are removed and the set, Θ, generated by *Obj*, will contain α^c.

Since it is assumed that the source is identical for the two formulae, α^c

and $\sim\alpha^{c}$, in the subsequent application of ml, the source will be downgraded and α^{0} (α^{bot}) will be generated as a replacement to be inserted in the next opp.

This final opp then created will then have a present state, in the life line of the exploration set, which is open.

Therefore, any state in the life line of an exploration set which is not open is transformed by the transformation rules of the URS, into an open state in the life line.

So, exploration sets for the URS can never collapse. □

Proposition 6.5.2

For any sequence of observations, seq, which contains fewer than $|C|-1$ contradictions, the Unforgiving Model based reasoning system reacts after seq.

Proof:

By Proposition 6.5.1, the URS does not collapse after seq.

Let seq' be the sequence of observations containing only β^{top} and let $(ES, seq.\alpha^{top}.\sim\alpha^{top})$, $(ES', seq.\alpha^{top}.\sim\alpha^{top}.seq')$, $(ES'', seq.\sim\alpha^{top})$ and $(ES''', seq.\sim\alpha^{top}.seq')$ be interactions for URS.

Then, $A(ES', seq.\alpha^{top}.\sim\alpha^{top}.seq') - A(ES, seq.\alpha^{top}.\sim\alpha^{top}) = \{\beta^{c}\}$, where $c < top$, while $A(ES''', seq.\sim\alpha^{top}.seq') - A(ES'', seq.\sim\alpha^{top}) = \{\beta^{c+1}\}$, provided the source has not been downgraded $|C|-1$ or more times in the entry of seq, and $A(ES''', seq.\sim\alpha^{top}.seq') - A(ES'', seq.\sim\alpha^{top}) = \{\beta^{c}\}$, otherwise.

Therefore, provided seq contains fewer than $|C|-1$ contradictions, the unforgiving model based reasoning system reacts after seq.

If seq contains $|C| - 1$ or more contradictions, then all observations will be downgraded to bot certainty, which will mean that URS fails to react thereafter. □

It is clear from the two Propositions proved above, that:

uncertainty $RS \prec URS$, and non-monotonic $RS \prec URS$.

However, if seq is a sequence of observations containing $|C| - 1$ contradictions, in which each pair of contradictory formulae are separated by a number of observations of α^{bot} which is greater than the persistency assigned to either of the contradictory formulae in an interval RS, then the interval RS will react after seq, by collapsing, while URS will not, as seen in the final point in the proof of Proposition 6.5.2. Therefore, the interval RS is incomparable, in this order, with URS.

6.6 The Final Position of the Unforgiving Model

The unforgiving RS is interesting for both its strengths and its weaknesses. The first thing we learn from it is how far away from a realistic model are the ones based on the five classes of logics considered, since they are easily beaten by a model so crude and unsophisticated as the unforgiving. Its main weaknesses come from two features: the inability to react to contradictions between different sources without collapsing, and the possibility of a pseudo-collapsed situation in which the system does not communicate with the external world any more.

The former problem is shared by our present model for adaptive RS: it would have been possible, of course, to devise some prepackaged technique to deal with the situation, but we believe that resolving the difficulties involved in deciding reasonably which source is responsible for a contradiction that has arisen, requires the use of reasoning power. While we do not see this as an insurmountable obstacle, even at the present stage of

research, it is clearly beyond the scope of this work. As mentioned before, it is our policy to introduce simplified, skeletonic models *only* if we expect them to survive in the more mature developments. At this stage, an algorithm for dealing with contradictions between different sources would have in fact played the opposite role, by suggesting a solution going in the opposite direction from the one we envisage.

The second problem with the unforgiving *RS*, on the other hand, highlights the central difference in a comparison with the adaptive one: the former model is prepared to refuse any contact with its sources in order to preserve its consistency, while the latter has an obligation to keep communicating with the world. Note that we used the expression "pseudo-collapse" because the unforgiving *RS* is still prepared to talk to new sources, and as such is not collapsed, even though it can look as if it is.

This brings us to the idea of tension between different motivations. It is clear that the "solution" of refusing all contact with a *substantial* part of one's environment as a way of avoiding contradictions is ridiculous, and the reason is that we want a model able to operate and interact with the reality in a sufficiently efficient way to be even partially comparable with a human behaviour in the same situation. This idea is explored in the adaptive reasoning system, which we introduce in the next section.

6.7 Adaptive Reasoning Systems

The structure of the adaptive reasoning system is so much more complex than the models we have presented hitherto, that it requires a rather different presentation. Here we give a complete picture of the way in which its parts mesh together and it controls both its interaction with its environment and its own progressive development of that control. The description will be fleshed out with as much detail as is useful and instructive, without hindering understanding. However, the specifications of the

particular functions themselves have been placed in Appendix A, together with a brief description of each individually.

In the previous section we discussed the ways in which the unforgiving model fell short of the demands of genuine interaction with a real environment. In the adaptive model we attempt to face these demands with a coherent strategy towards the maintenance of a balance between contradictions and accepted observations.

The ideas of acceptable frequency of contradictions and sufficient proportion of incoming information believed thus spring from the awareness of the fact that, given the changing nature of the world and the incompleteness and inaccuracy of the information we receive about it, we are bound, every now and then, to find ourselves in contradiction (that is, believing opposite data at the same time). The recovery of consistency can thus be only partial, for a limited period of time, unless we are prepared to completely stop our interaction with the external world.

This leaves us with two problems: how to balance the motivation towards consistency with that towards interaction (that is, amount of information believed and hence usable); and how to recognise the best possible balance in a particular situation and move toward it.

The problem of balance can be divided into two parts: first of all there must be a way of measuring the two motivations, and secondly these two measures must be comparable in a sensible way. A natural measure for the degree of interaction is the amount of information believed over the amount of information input in the same period (computed through the variations in the degrees of certainty). A system could then be said to tend towards an increase in interaction if, given a long enough period, this ratio increases over time.

The measure for consistency is constructed along the same lines, except that the periods of time are determined by the occurence of contradictions. Thus, we would say that a system tends toward consistency if the length of the periods (calculated in terms of observations, or inputs) between two

successive contradictions increases with time. We call this "convergence at level one". This trend is, unfortunately, a very difficult goal to achieve, unless we are prepared to make huge sacrifices on the interaction side. It may very well be, though, that considering periods containing no more than two contradictions is enough to realise the convergence, which would then be called "of level two". The process can clearly be reproduced for higher levels. It is nevertheless desirable to achieve the convergence at low levels (for very high levels, in fact, the concept tends to loose its relevance).

The problem of comparison is strictly connected to that of motivation. For example, a program simulating a working mathematician would consider consistency as an overall priority, while another acting as a consultant in an emergency ward should place more emphasis on the speed of response and the readiness for reaction. This side of the balance of motivation is clearly due to external factors, like the kind of work the system is expected to do. In the human case, these external reasons can be so deeply entrenched to be completely outside the control of any conscious mechanism, for example the gregarious needs which most of our species feels.

We have devised a balance constructed on "order priority", and assumed that, for the time being, we try first to minimise the level of convergence for inconsistencies (up to level two), and then we try to maximise the interaction for that level. Other levels of balance could be easily devised, and the functioning of the system is independent of the chosen balance of motivation.

Optimising these two measures is a hard problem, if arbitrary solutions are to be avoided, since very different elements are involved in the analysis, and it is almost impossible to devise weights from outside. This is because the weight with which different evidence can bear on an optimisation depends essentially on the kind of reality on which the system is working, and this is exactly the point which we want to keep totally flexible.

Our methodology is, first of all, to compare equals with equals, in the sense that we run "trial tests" of the system on an internal simulation, varing one parameter at a time. While of course the technique does not guarantee success, it is a very reasonable way of looking for improvements,

provided the length of the trial is sufficient. We will come back to this point later. The idea of a trial protects us against sudden, unjustified changes, since the old parameter is not abandoned until a new candidate has emerged with enough strength.

The problem with this approach is that the trial is expensive, even more so because only one or at most two parameters are allowed to vary at a time, and there is, of course, a very large number of possibilities for the variation. It follows that any randomised search has a very poor chance of succeeding.

We present a solution to this problem based on an "enquiry" model: evidence is collected, by specialised functions, which suggests a variation in some parameters. According to the *apparent* strength of the evidence (which could be very different from its real bearing on the case), and to the present general situation (which indicates to the *balancing* function, in control, a list of priorities), trials are run both on minimal and *suggested* changes of parameters. If, according to the balance of motivation explained above, the challenge is succesful, the new parameter is installed, and the evidence used is wiped out. If, on the other hand, the result has been negative for the candidate, the evidence can be reused, strengthened by new evidence, to make a stronger case.

The case-preparing functions have been devised following reasonable arguments, but the important thing is that they act in such a way that any decision can be subsequently reversed. The system thus achieves a very high degree of flexibility. Also, the information used for building the cases is all high-level, in the sense that it is not concerned with single instances but with trends. This way, we avoid prematurely linking the system to some possible environment or use, by embedding general enough analysis in the case-preparing functions, which is anyway reversable.

The performance of the balancing function, which is responsible for monitoring the general performance graph in the light of the motivation, for establishing priorities among cases, and for judging the winner in a trial, is then monitored in turn by another function. This controls the lenght of the trials, according to the following rule: if a decision to change a

parameters is subsequently completely reversed, then the trial is thought to be too short. If, on the other hand, the same conclusion as reached at the end of the trial could have been reached at a previous stage, then the duration of the trial is considered too long. This algorithm is a simple hill-climbing one, which for example does not take into account possible sudden *fractures* in the reality, which could justify the reversal of a decision.

In addition to the balancing of motivations, this model develops several new ways of organising its information - in particular, the use of circumscriptions to divide its view of the world into disjoint parts so that contradictions between formulae in different parts do not matter. This corresponds to a similar use of circumscriptions in human organisation of knowledge - for example, most people would happily accept the validity of the statement "cats don't bite" - until they are big-game hunting in central Africa. There is no contradiction between the apparently contradictory statements "cats don't bite" and "cats do bite", provided we recognise that the first is true within a circumscription containing the world of everyday animals that are encountered in Western Europe, while the second is true within the circumscription containing big cats in central Africa.

As we have pointed out already in the introduction (Chapter 1) to this work, there are several points where the use of reasoning techniques would improve the system: we think that the features illustrated here represent important lines of attack which we expect to be improved, but not completely replaced, in a more advanced version. When we recognise the situation to be otherwise, as for example is the case with the analysis of fractures in a source's behaviour (cf. *rpcalc* defined in Appendix A), we prefer simply to define the "socket" where the more powerful mechanism should plug in.

It is interesting to note that the cases where the consistency recovery mechanism cannot be sensibly developed any further without requiring reasoning power are those where information about the world is needed. On the other hand, the further we abstract from the individual data input in the particular situation, and the more we proceed toward an introspective

analysis in which the system works only on data which it has itself provided, the more the techniques we present acquire effectiveness, completeness and elegance. We consider this to be a clear signal of the fact that other essential units lie at the periphery of this model, such as an intelligent user interface, actions analysis and above all reasoning techniques.

An additional feature of the system is that it has action points, that is points in the certainty scale at which a belief is considered sufficiently reliable by the system for it to act upon. As we have already pointed out in the discussion of uncertainty *RS* (cf. Section 4.11), the existence of these points is essential to distinguish non-monotonic *RS* from uncertainty *RS* with infinite degrees of certainty (in the finite case, the action threshold can be trivially identified with the *top*). Even in the finite case, action points are useful to distinguish between absolute belief and a belief strong enough to act upon. This is particularly true when, as in our model, the data are organised in circumscriptions, which can then have different action points. The balancing function is also provided with a motivation to decide the weight of inconsistencies above the action points with respect to those below it (they are organised in two separate graphs), but it is very reasonable to assume that the former have much greater weight.

It is interesting to note that, under our homogeneity analysis, the circumscriptions are *individuals* (cf. Section 3.12).

The whole idea of convergence towards some balance point, which is used repeatedly within the construction of this model, rests on the basic hypothesis of homogeneity which we discussed in Section 2.5. The intuition behind this is that, if there is such a balance point in the reality in which the system operates (that is, there is a general point of view under which the reality is homogeneous enough to be predicted), then our techniques have some chance to find it. This hope is based on the reasons exposed in Chapter 2. If, on the other hand, there is no such a thing, then the system is likely to reveal this through its difficulty in finding a convergence at a meaningful level.

We would like to point out that, while several parts of the system, as described below, are computationally very expensive and should be cut to size before being implemented, there are no parts of the algorithm of greater than polynomial complexity (in fact, almost always linear), despite the specifications of some functions appearing to suggest an exponential complexity. In the case of the computation of best homogeneity and persistency for circumscriptions (cf. *hps* defined in Appendix A), this is achieved using Theorem 3.7.1. In other cases, as for the construction of the graphs used for performance monitoring (cf. *graph* defined in Appendix A), the low complexity is obtained by using an efficient data structure.

6.8 The Principal Features of the Adaptive Reasoning System

The design of the adaptive model can be analysed in terms of a few major concepts, namely source and domain identification, survival against all observations, reaction to inconsistencies, tendency towards stabilisation, modifications over data, sources and domains (both downgrading and upgrading), adaptation of its own reactions, tendency towards maximum information and the need for action.

Some of the mechanisms devised to carry out these tasks are interesting in their own right: we now examine them in more detail.

A general point about these mechanisms is that they have been created to avoid the arbitrarity of numerical values (that often do not correspond to anything) by instead embedding in them a qualitative argument.

The operation of the system is best described by examining first the acceptance functions. These, just as in the unforgiving model, are used to regulate the information which is fed by the sources. In particular, the principal function must decide what is the destination of a particular observation, and what certainty must be ascribed to it according to the current rating of the source providing that information.

The object level function is responsible for applying the transformation rules of the system to the formulae in the perceived states. If a contradiction is discovered, a major part of the machinery of the system is called into play. Firstly, the functions which resolve the responsibility for the contradiction are called. These must evaluate the liklihood of the source or the circumscription being the guilty party. Once decided, the guilty party is punished appropriately, by the punishing functions. The punishments that are used are downgrading for sources and cutting of the assumptions of homogeneity and persistence for circumscriptions. There is a third option, which is that the contradiction could be ascribed to a misunderstanding between the system and the source, so that the information should actually be placed in a different circumscription to that originally assigned.

The whole system is geared towards learning from the flow of contradictions, as has been emphasised many times. The first part of the machinery responsible for attempting to attain this goal is a "clock" which counts the observations that are being made in order to calculate the rate of flow of contradictions. The first part of the count is made using a special clock function, until the count exceeds the period for which an incoming contradiction would destroy the current picture of the flow of contradictions. Once this period is over, any subsequent time without contradictions appearing sees an increasing relaxation of the controls within the system that hold back attempts to increase the belief and trust of the system in circumscriptions, particularly at the borders of persistency or homogeneity, and the reinstatement of sources.

However, should a contradiction arise within the period of danger, a principal function responsible for tightening up the response of the machine to contradictions is called. This uses the evidence of the previous contradiction to decide how to increase the control of source or circumscription. The result is that the guilty party is much more severely punished and redeemed more slowly, in future contradictions attributed to the same problem.

Functions that play a central role in this group are *inconsistency* and *resolve* and *action*. The principal idea here is to make the system somehow aware of the assumptions it is making about the environment, and

it is fundamentally connected to the point made in Chapter 2 about the circular relation between assumptions, observations and reasoning. Since in this model there is no reasoning power in the abstract form, but only embedded in the consistency recovery functions, this circular relation simplifies to a binary one between assumptions and observations.

Parts of the system are constantly monitoring the performance of sources and circumscriptions, watching for opportunities to upgrade either of them, and trying to form bodies of evidence to suggest that one or other of them has been too severely dealt with by the activities of the inconsistency controlling side. A major weapon in composing this evidence is the existence of calls for action from outside the system, which request the provision of information above the appropriate action threshold for the circumscription in which the information resides. If such a call goes unsatisfied, a group of functions responsible for monitoring the failure of the system to act are called. The main role of the principal function in this group is to see if having believed a source more, or having trusted a circumscription more would have allowed the action after all. If so, evidence is presented for the appropriate change in the system to be made, in order that future requests might be met.

A large part of the machinery is devoted to the organisation and the evaluation of "trials". These take place when a sufficient body of evidence has been collected to suggest that a better set of values for the parameters controlling the behaviour of the system exists. A trial is then organised in which a second "shadow" system, identical to the first, except for all the trial control functions, is run in parallel to the first system, using the alternative parameters. At the end of an appropriate length of time, (again subject to internal control and adjustment) the trial is evaluated, based on the relative flow of contradictions and acceptance of information.

The central function in this group is obviously *balance*. The idea behind this mechanism is, as we pointed out before, to compare equals with equals, and to run trials, which are short simulations of the system inside it. This is an attempt to recreate the human ability to devise mental scenarios of a simplified world and "run" them, then translate the results in the real world and take actions based upon these results. According to some authors,

this is in fact the single most important characteristic of human mental behaviour.

There is a group of functions which have not been introduced in any of the previous models, principally led by *danger*, the purpose of which is to preempt any possible contradiction.

The preempting group is strongly related to the balancing group, since its task is related to predictions too. The difference is that, while the balancing group does its job by building a whole scenario, and then watching what would have happened and comparing it with the real interaction, in order to learn about the best way to deal with its environment, the preemptive ones are interested only in dangerous situations, which we do not want to be tried out at all, if it can be prevented. These functions do not build a whole scenario, but simply take some precautionary actions for a limited time. They are less sophisticated than the balancing ones, and less authoritative than the group containing *action*, *inconsistency* and *resolve*, but we believe that they have a very effective role to play. It must be remembered that the actions taken by these functions are transitory, and do not try to reflect a view of the world, but simply to avoid major predictable problems.

The main functions here are *danger* and *suspension*, which operate by watching any formulae which are overridden several times, changing semantic value back and forth and climbing the certainty scale. If there is a formula behaving in this way which it is estimated will cross the action threshold on the next occasion that an override occurs, and the override, it is estimated, will occur soon, then the formula is suspended. If the prediction turns out to have been accurate, then the consequences of the contradiction are not brought down on the whole system, but only on the responsible source.

The complete behaviour of the system is monitored by a complex series of data structures, which can be seen as a graph of performance. The operation of this mechanism has been described already, to some extent. The principal function itself is responsible for coordinating the collection and evaluation of all the performance details, relying on many book-keeping

functions to provide it with the necessary facts. The function itself can be found in Appendix A, as can the specifications and brief descriptions of all the other functions. It will be seen that the *link* function has not been written. This is because the flow of the machine is described in the details of each of the functions individually, while the technical detail of *link* is both unpleasantly complicated by the treatment of all the many cases that arise in the internal state of the system and, at the same time, not at all instructive.

6.9 Final Comments on the Adaptive Reasoning System

Having constructed and described the mechanisms of the adaptive reasoning system, as well as the principles on which it is based, it is of interest to consider the gains it embodies, over the reasoning systems we have considered in the earlier chapters.

Clearly the adaptive reasoning system is far more complex than any of its predecessors, which raises the obvious question - has the gain been worth the increased complexity? It is our belief that the system not only completely outstrips all the others we have considered, from the point of view of interactive performance, but it also contains several features which go beyond this basic requirement, such as self-improvement.

With the benefit of hindsight we can now see how naive was the attempt to consider the systems which were essentially logics in the role of reasoning systems. On the other hand, it is very difficult to imagine the adaptive reasoning system, we have proposed, in use as a logic, not only because of its complexity, which would make it totally impractical, but also since it contains exactly that thing which lacking makes logics too weak to be effective reasoning systems: the ability to observe and control its own behaviour and recover rationally from inconsistency.

Conclusion

As suggested in Chapter I, we see the contribution of this work being towards methodology as well as towards the issues it addresses directly. With respect to the methodological plane, we would like to enphasise the particular relationship between philosophy and A.I. that we have pursued. It is widely recognised that A.I., by its own nature, raises highly sensitive philosophical questions, such as "is processing the same thing as thinking?"; usually this kind of discussion is not found in papers at the forefront of A.I. research, especially nowadays, but are reserved for specialised occasions, and we believe this is wise.

There is another aspect to the relationship between A.I. and philosophy: understanding and analysing which philosophical hypotheses can support a view of a process which we want to model, or of a technique that we want to use. Symmetrically, a general view, held for reasons independent from the A.I. application, can prove to be the key for breaking free of well-known techiques and into a new ground. We have exploited both paths; an example in one direction is the need to compare systems using different logics, which has brought us to define the concept of reasoning systems. A typical and important example of the opposite way has been the stimulus provided by our pragmatic view of knowledge towards the recognition of the needs to distinguish and unify, from which homogeneity theory has evolved.

Moving onto the plane of contents, three areas can be identified where we think a contribution has been made: the theory of homogeneity, the definition and classification of reasoning systems, and the specification of the consistency recovery mechanism for the adaptive reasoning system.

In our experience, homogeneity theory has proved to be a very powerful tool and it is also, we think, not devoid of a certain elegance. As we

have pointed out already, we expect this theory to provide us with the theorethical basis on which to unify the different unstable reasoning techniques that we have identified.

The definition and classification of reasoning systems has put the controversy about which kind of logic is appropriate to simulate human reasoning on firmer foundations. This is a task which has been recognised more and more often recently (e.g. [CHO84] and [FA86]), but which has usually been addressed by an empirical comparison of features, very much dependent on the particular example chosen (sometimes implicitly) and on some intuitive interpretation. Our method, by contrast, explicitly declares any assumption used and formalises the framework in which the systems are to be tested. Furthermore, the tests are all run in a very formal manner, and the results proved. The results about ordering acquire a particular strenght from having been obtained through two very different techniques, one internal - the equivalence of the identity rules to assumptions of homogeneity - and one external - the interactive behaviours against some specified and controlled, but completely abstract, set of inputs. The abstraction that we have achieved allows us to claim that these results are completely domain independent, as long as the general conditions specified in the propositions are satisfied.

The specification of the consistency recovery mechanism for the adaptive reasoning system brings us to our own proposal in order to deal with the original problem of poor information. This follows naturally from the analysis carried out before, so that for example the need for the system to manipulate the data input before acceptance emerges clearly from the contraposition of reasoning systems that over-react and others that do not react at all.

The main ideas embedded in the adaptive reasoning system are the following: the need to identify sources and domains of discourse; the capability to transform single inputs (usually to reduce their scope); the possibility to act on sources and domains of knowledge, downgrading the reliability of the formers and weakening the predictive power embedded in the latters; the potentiality for reversing previous decisions about sources and domains; the existence of motivation towards an optimal interaction; the

capacity of constructing cases in favour or against each particular decison, and the existence of a trial mechanism by which to decide the cases.

The next stage in our research will be in two main directions. On the one hand, our work will concentrate on developing our understanding of each unstable reasoning technique, and to interpret and unify them through homogeneity theory. On the other hand, it will be concerned with connecting the consistency recovery mechanism described here to the reasoning techniques, in order to complete the former by "plugging in" the missing modules that, we believe, require additional reasoning power. Also, in a similar vein, we envisage the use of motivation, balancing techniques and case-preparing functions to provide a top level *reaction and adaptation* mechanism to control the interplay of the reasoning techniques.

The Functional Specification of the Consistency Recovery Mechanism for an Adaptive Reasoning System

$fst: \mathbb{P}(X \times \{ opps \}) \rightarrow X$

$fst(S) = \pi_1(x)$, where $\pi_2(x) = min(\{ \pi_2(y): y \in S \})$.

This function is used to deliver the member of the set S which occurred first, without its tagging opp.

$remf: \mathbb{P}(X \times \{ opps \}) \rightarrow \mathbb{P}(X \times \{ opps \})$

$remf(S) = S - \{ x: fst(S) = \pi_1(x) \}$.

This function removes the first element from S.

$lst: \mathbb{P}(X \times \{ opps \}) \rightarrow X$

$lst(S) = \pi_1(x)$, where $\pi_2(x) = max(\{ \pi_2(y): y \in S \})$.

Analogous to fst except delivering the last.

$reml: \mathbb{P}(X \times \{ opps \}) \rightarrow \mathbb{P}(X \times \{ opps \})$

$reml(S) = S - \{ x: lst(S) = \pi_1(x) \}$.

This function removes the last element from S.

$take: \mathbf{N} \times \mathbf{P}(X \times \{opps\}) \rightarrow \mathbf{P}(X)$

$take(n, S) = \emptyset,$ if $n = 0$ or $S = \emptyset$,

$= \{fst(S)\} \cup take(n-1, remf(S)),$ if $n \neq 0$ and $S \neq \emptyset$.

This function cuts the first n elements of S and makes a "list" from those.

$drop: \mathbf{N} \times \mathbf{P}(X \times \{opps\}) \rightarrow \mathbf{P}(X \times \{opps\})$

$drop(n, S) = S,$ if $n = 0$ or $S = \emptyset$,

$= drop(n-1, remf(S)),$ o.w..

This function drops the first n of the elements of S.

$indez: \mathbf{N} \times \mathbf{P}(X \times \{opps\}) \nrightarrow X$

$indez(n, S) = fst(drop(n-1, S)),$ if $|S| \geq n$,

= undefined, o.w..

This function produces the nth element of S.

All the previous group of functions are used in the manipulation and maintenance of the graph records, introduced later.

$exc: [0,1] \times [0,1] \times \mathbf{N} \rightarrow \mathbf{N}$

$exc(h, d, U) = min(\{k(1-h), k(h-1)+U, U-k, k\})$, where $k = d+(1-d)U$.

This function computes the number of exceptions that are compatible with a

given value of homogeneity, depth and size of world. The function is derived from a rearrangement of the result of Theorem 3.7.1.

$mp : \{ opps \} \times \mathbf{P}(\text{Form}) \times \mathbf{N} \rightarrow \mathbf{P}(\mathbf{N})$

$mp(k, \Phi, n) = \{ min(\{ |(\Theta,\alpha,c)per| : (\Theta,\alpha,c)per$ is headed, or $(\Theta,\alpha,c)per$ is topped and there is no headed $(\Theta,\alpha,c)per$, where $|\Theta| = n$ and $obs_k \in \Theta \}) : \alpha^c \in \Phi \}$.

This function computes the minimum persistency periods for each of the formulae in Φ, using a look-back over the last n ob-states.

$pcalc : \{ opps \} \times \mathbf{P}(\text{Form}) \times \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$

$pcalc\ (k, \Theta, n, ex) = min(mp(k, \Theta, n) - \Phi)$

where $|\Phi| = ex$ and $\forall v \in mp(k, \Theta, n)$.
$((\exists u \in \Phi. v \leq u) \Rightarrow v \in \Phi)$.

This function determines the best persistency possible, over the set of formulae Θ using look-back n, and allowing ex exceptions.

cg_k : Sources $\nrightarrow$ **N**

Book-keeper: storing the current grading for a source.

ud_k : Sources $\nrightarrow$ **N**

Book-keeper: storing the number of times a source has been upgraded and then downgraded.

sm_K : $\mathbb{P}$(Form)

Book-keeper: storing the set of observed formulae which were not accepted at the same certainty as they were entered.

pc_K : Sources $\rightarrow$ $\mathbb{N}$

Book-keeper: storing the number of potential contradictions a source would have been responsible for. This is computed with the aid of the record of the formulae, sm_K.

rp_K : Sources $\rightarrow$ $\mathbb{N}$

Book-keeper: keeps a record of the performance required of a source before it can be upgraded.

$rpcalc: C \times \mathbb{N} \times \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$

This function gives the required minimum performance before a source can be upgraded. Its value will be stored in rp_K. Its arguments are: certainty at which the contradiction for which the source has been downgraded occurred; ud_K, pc_K, cg_K (for appropriate source). This function is not specified in greater detail, because we believe that a more sophisticated method for deciding when a source has performed well enough to be upgraded is required, based on reasoning about possible fractures in the behaviour of a source.

cc_K : Sources $\rightarrow$ C

Book-keeper: recording the certainty at which the contradiction for which a source was last downgraded occurred.

sp_k : Sources $\rightarrow\!\!\!\!\!\!\bullet$ $\mathbb{N} \times \mathbb{N}$

Book-keeper: keeps record of the number of observations a source has delivered over the action threshold and over cc_k, and secondly the number of observations over cc_k.

wgt_k : Sources $\rightarrow\!\!\!\!\!\!\bullet$ $[0,1]$

Book-keeper: records the current weight given to inputs below cc_k, for each source, in measuring the performance of a source.

$wgtstacks_k$: Sources $\rightarrow\!\!\!\!\!\!\bullet$ $\mathbb{P}(\mathbb{N} \times \{ opps \})$

Book-keeper: keeps records of source weight values when they are upgraded and downgraded, in order that they can be restored if the source is upgraded.

$pfcalc$: $\{ opps \} \times$ Sources $\rightarrow \mathbb{R}$

$$pfcalc(k, source) = (1 - wgt_k(source)) . \pi_1(sp_k(source)) +$$

$$wgt_k(source) . \pi_2(sp_k(source)) - rp_k(source).$$

This function calculates the current performance of a source over and above its required performance for upgrading, using the weighting currently assigned to observations above and below cc_k for that source.

$Sources_k$: $\mathbb{P}$(Sources)

Book-keeper: keeps the set of current sources.

$ts : \{ apps \} \rightarrow \mathbb{P}(\text{Sources} \times \mathbb{R})$

$ts(k) = \{ (source, perf) : source \in Sources_k$
and $perf = pfcalc(k, source) > 0 \}.$

This function delivers performance values for sources to *trust*, pairing source with performance. Its name abbreviates "trust seeker".

$lkb_k : \text{Bool} \times \text{Circs} \nrightarrow \mathbb{N}$

Book-keeper: records the look-back parameter for a circumscription, which is a number of ob-states. This value is used to decide how far into the past the records of behaviour of a circumscription should be considered. Essentially, the further back the search goes, the more conservative and restrictive is the system.

$Circs_k : \mathbb{P}(\text{Circs})$

Book-keeper: records the names of current circumscriptions.

$minhom_k : \text{Circs} \nrightarrow [0,1]$

Book-keeper: stores the minimum homogeneity for a circumscription, which is the least value for which information so unreliable can be used in that circumscription for any purpose.

$depth_k : \text{Circs} \nrightarrow [0,1]$

Book-keeper: delivers the depth for a circumscription. (That is, the number of ob-states that must be taken together as the smallest size set under the screen).

$worlds_k(circ) = \{\alpha : circ(\{\alpha\})\}$.

Book-keeper: recording the set of formulae which belong to a circumscription.

fr_k: Form ⇸ Circs

Book-keeper: gives the world circumscription from which a formula has been drawn.

who_k: Form ⇸ Sources

Book-keeper: recording the source responsible for a formula.

obs_k: $\mathbb{N}$

Book-keeper: stores the number of ob-states that have passed by state k.

$filter$: $\{opps\} \times$ Circs $\times \mathbb{N} \rightarrow \mathbb{P}$(Form)

$filter(k, circ, n) = \{\alpha^c : \alpha^c \in worlds_k(circ)$ and
$\exists m \in \{opps\}. (obs_k - n \leq m \leq obs_k$ and m is a first moment for $\alpha^c)\}$.

$filter$ isolates within the set of formulae in a world circumscription, that subset which has appeared at any time within the look-back period, n.

$hcalc : \{opps\} \times \text{Circs} \times \mathbb{N} \rightarrow [0,1]$

$hcalc(k, circ, ex) = max(\{0, 1 - ex/x, 1 - (U - ex)/x, 2 - U/x\})$,

where $U = |filter(k, circ)|$
and $x = depth_{\mathbb{R}}(circ) - (1 - depth_{\mathbb{R}}(circ)).U$.

This function serves a similar purpose to exc delivering the best homogeneity that can be attributed to a circumscription given a specified number of exceptions.

$hps : \{opps\} \rightarrow \mathbb{P}(\mathbb{P}(\text{Circs} \times \mathbb{N} \times [0,1]))$

$hps(k) = \{\{(circ, pcalc(k, filter(k, circ, lkb_{\mathbb{R}}(circ)), lkb_{\mathbb{R}}(circ), ex), hcalc(k, circ, ex)) :$

$0 \leq ex \leq exc(minhom_{\mathbb{R}}(circ), depth_{\mathbb{R}}(circ), |filter(k, circ, lkb_{\mathbb{R}}(circ))|)\} :$

$circ \in Circs_{\mathbb{R}}\}$.

This function determines the homogeneity and persistency values for each circumscription, delivering a set of triples containing the name of the circumscription and a persistency and homogeneity value that can be achieved at present.

$times_{\mathbb{R}} : \text{Circs} \rightarrow\!\!\!\!\!\rightarrow \mathbb{P}(\text{TimeCircs})$

Book-keeper: storing the set of time-circumscriptions for formulae in a given world-circumscription. It is in this book-keeping function that the information generated by the previous function is indirectly stored.

$per : \text{TimeCircs} \rightarrow \mathbb{N}$

Book-keeper: recording the persistency of a time-circumscription in ob-states.

ha_k : Circs × TimeCircs → [0,1]

Book-keeper: maintaining records of the assumptions of homogeneity for time-circumscriptions in world-circumscriptions. This function and the last are the direct records accessible through the names kept in $times_k$.

to_k : Form → **N**

Book-keeper: keeps track of the time of origin of a formula (this is the origin given by the source) in ob-states.

$clock_k$: Bool → **N**

Says how many ob-states have passed since the alarm rang - that is, the number of ob-states beyond the earliest point at which a contradiction could have come without upsetting *inconsistency*. An ob-state is one a state in which an observation has been recorded.

$alarm_k$: Bool → **N** × **N**

The first value is the number of ob-states that have passed since the last contradiction, including state k, while the second is the number of ob-states which must pass before it is "safe" to have a contradiction.

$cert$: [0,1] → C

This function tells what the certainty is of a formula in a circumscription, with given homogeneity. It maps 0 to *bot* and 1 to *top*, and it is order preserving.

$incer : C \rightarrow [0,1]$

$incer(c) = max(\{ h \in [0,1] : cert(h) = c \})$.

This function "inverts" $cert$.

at_k : Circs $\nrightarrow C$

Book-keeper: giving the action threshold for each world circumscription.

par_k : Bool $\times$ Sources $\nrightarrow \mathbb{N}$

Book-keeper: records the current minimum performance and/or time required before a given source can be reinstated. The boolean value is used to distinguish between information about observations abve and below the action threshold.

$trust : \{ opps \} \rightarrow \mathbb{P}$(Sources)

$trust(k) = \emptyset$, if $clock_k(T) + wgt_k(\pi_1(x)).clock_k(F) = 0$,

$= \{ \pi_1(x) : x \in ts(k)$ and
$\pi_2(x) + clock_k(T) + wgt_k(\pi_1(x)).clock_k(F)$
$> par_k(\pi_1(x)) \}$, otherwise.

This function gives a list of sources which are to be upgraded at time k.

The following functions are all concerned with the maintenance of information in order to construct and develop the graph. The levels in the graph are forced to always follow a single trend upwards, then possibly a

trend downwards - they are not allowed to fluctuate up and down. This is prevented by keeping track of information when on a downward trend that will allow the reconstruction of the upper levels should the downward trend be halted. The boolean values in all the book-keeper functions is used to distinguish the graph of values above the action thresholds from that of values below the threshold.

lev_k : Bool $\rightarrow$ $\mathbb{N}$

Book-keeper: records the current level required in order to view the performance graph.

lvs_k : Bool $\rightarrow$ $\mathbb{P}(\mathbb{N} \times \{ opps \})$

Book-keeper: records the set of last values of counts between contradictions, together with the time at which each was evaluated.

cvs_k : Bool $\rightarrow$ $\mathbb{P}(\mathbb{N} \times \{ opps \})$

Book-keeper: keeps track of the set of all values of count since the trend became downward, together with their evaluation times.

$count$: Bool $\times$ $\{ opps \} \rightarrow \mathbb{N}$

$count(bl, k) = \pi_1(alarm_k(bl)) + clock_k(bl)$

This function computes the number of states that have passed from the last contradiction.

$trend_k$: Bool $\rightarrow$ { "u","d" }

Book-keeper: saves the current trend in changes between levels of

examination of the graph of performance.

$stack_n$: Bool $\rightarrow$ $\mathbf{P}((\mathbf{N} \times \mathbf{N}) \times \{ opps \})$

Book-keeper: recording the values of lev_n, the contradiction count, $conts_n$, and the state m in which it was recorded. This is used when unrolling a mistaken downward trend.

$down$: Bool $\times$ $\{ opps \} \nrightarrow \mathbf{N}$

$down(bl, k) = max(\{ n : \Sigma take(n, lvs_n(bl)) < count(bl, k) \}) - 1.$

(Assume $\Sigma\{\ \} = 0$).

This function is used to determine whether a downward move in lev_n is possible.

$process$: Bool $\times$ $\{ opps \}$ $\times$ $\mathbf{P}((\mathbf{N} \times \mathbf{N}) \times \{ opps \}) \rightarrow$

$\mathbf{N} \times \mathbf{P}((\mathbf{N} \times \mathbf{N}) \times \{ opps \})$

$process(bl, k, S) = (lev_n(bl) + 1, \emptyset),$ if $S = \emptyset$,

$= (\pi_1(lst(S)), reml(S)),$ if $check(conts_n, \pi_1(lst(S)), \pi_2(lst(S)), cvs_n \cup \{(count(k), k)\})$,

$= process(bl, k, reml(S))$, o.w..

This function is used in unwinding the stack when a downward trend is proved false.

$check : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \times \mathbb{P}(\mathbb{N} \times \{ opps \}) \rightarrow \text{Bool}$

$check\,(p,n,q,S) = T,$ if $|S| < n \geq p-q,$

$= (x < index\,(n,S)) \wedge check\,(p,n,q,remf(S)),$ if $|S| \geq n \geq p-q,$

$= check\,(p,n,q+1,remf(S)),$ o.w..

This function is also used in unwinding the stack.

$conts_k : \text{Bool} \rightarrow \mathbb{N}$

Book-keeper: keeps track of the number of contradictions having occurred by state k.

$start_k : \text{Bool} \rightarrow \mathbb{N}$

Book-keeper: records the contradiction number at which cvs_k starts.

$graph : \{ opps \} \rightarrow$
$(\text{Bool} \rightarrow \mathbb{N}) \times (\text{Bool} \rightarrow \mathbb{P}(\mathbb{N} \times \{ opps \})) \times (\text{Bool} \rightarrow \mathbb{P}(\mathbb{N} \times \{ opps \})) \times$
$(\text{Bool} \rightarrow \{ \text{"u","d"} \}) \times (\text{Bool} \rightarrow \mathbb{P}((\mathbb{N} \times \mathbb{N}) \times \{ opps \})) \times (\text{Bool} \rightarrow \mathbb{N})$

$graph\,(k) = (lev_{k+1}, lvs_{k+1}, cvs_{k+1}, trend_{k+1}, stack_{k+1}, start_{k+1}),$

where:

if $count_k(bl) > fst(lvs_k(bl))$ and $down(bl,k) = 0$ then:

$lev_{k+1}(bl) = lev_k(bl),$
$lvs_{k+1}(bl) = remf(lvs_k(bl)) \cup \{ (count_k(bl), k) \},$

$$
\begin{aligned}
cvs_{k+1}(bl) &= cvs_k(bl) \cup \{(count_k(bl), k)\}, \text{ if } trend_k(bl) = \text{“d”},\\
&= \emptyset, \qquad \text{o.w.},\\
trend_{k+1}(bl) &= trend_k(bl),\\
stack_{k+1}(bl) &= stack_k(bl),\\
start_{k+1}(bl) &= start_k(bl);
\end{aligned}
$$

if $count_k(bl) > fst(lvs_k(bl))$ and $down(bl,k) > 0$ then:

$$
\begin{aligned}
lev_{k+1}(bl) &= lev_k(bl) - down(bl,k),\\
lvs_{k+1}(bl) &= drop(down(bl,k)+1, lvs_k(bl)) \cup\\
&\qquad \{(count_k(bl), k)\},\\
cvs_{k+1}(bl) &= cvs_k(bl) \cup \{(count_k(bl), k)\},\\
trend_{k+1}(bl) &= \text{“d”},\\
stack_{k+1}(bl) &= stack_k(bl) \cup \{(lev_k(bl), conts_k(bl), k)\},\\
start_{k+1}(bl) &= conts_k(bl);
\end{aligned}
$$

if $count_k(bl) \le fst(lvs_k(bl))$ and $trend_k(bl) = \text{“u”}$ then:

$$
\begin{aligned}
lev_{k+1}(bl) &= lev_k(bl) + 1,\\
lvs_{k+1}(bl) &= lvs_k(bl) \cup \{(count_k(bl), k)\},\\
cvs_{k+1}(bl) &= \emptyset,\\
trend_{k+1}(bl) &= \text{“u”},\\
stack_{k+1}(bl) &= \emptyset,\\
start_{k+1}(bl) &= start_k(bl);
\end{aligned}
$$

if $count_k(bl) \le fst(lvs_k(bl))$ and $trend_k(bl) = \text{“d”}$ then:

$$
\begin{aligned}
lev_{k+1}(bl) &= \pi_1(process(bl, k, stack_k(bl))),\\
lvs_{k+1}(bl) &= drop(conts_k(bl) - start_k(bl)\\
&\qquad - lev_{k+1}(bl) + 1, cvs_k(bl) \cup \{(count_k(bl), k)\}),\\
cvs_{k+1}(bl) &= \emptyset, \quad \text{if } \pi_2(process(bl, k, stack_k(bl))) = \emptyset,\\
&= cvs_k(bl) \cup \{(count_k(bl), k)\}, \quad \text{o.w.},\\
trend_{k+1}(bl) &= \text{“u”}, \quad \text{if } \pi_2(process(bl, k, stack_k(bl))) = \emptyset,\\
&= \text{“d”}, \qquad \text{o.w.},
\end{aligned}
$$

$$stack_{k+1}(bl) = \pi_k(process(bl, k, stack_k(bl))),$$
$$start_{k+1}(bl) = start_k(bl).$$

This function is the heart of the graph maintenance. It is responsible for interpreting the information stored in the book-keepers and for updating them.

$valcirc,\ valsource: [opps] \times \text{Form} \rightarrow \mathbb{N}$

Functions giving the responsibilities of source and circ in a contradiction. These functions remain unspecified, since to allocate responsibility sensibly is a problem requiring reasoning. An unintelligent allocation can be based on the current "criminal records" of the suspects, together with the cost of losing either in terms of lost information, but this would be very unsatisfactory.

$CreateCirc_k$: Bool

Book-keeper: records true when a new circumscription record has been entered in the current state. It is assumed that when a contradiction occurs, this will only be true if the offered circumscription satisfies the problem.

$CircCreate_k$: Circs $\times$ $\mathbb{P}$(Form) $\times$ [0, 1] $\times$ C

Book-keeper: keeps track of the data for a circumscription that has just been created and awaits entry into the main files.

$PunishCirc: \{ opps \} \times \text{Form} \rightarrow$

$$(\text{TimeCircs} \times \mathbb{P}(\text{Form}) \times (\text{Circs} \times \mathbb{P}(\text{Form}) \times [0,1] \times C) \times \mathbb{P}(\text{Sources}))$$

$$PunishCirc(k, \alpha^c) = (\{\, tc : tc \in Times_k(fr_k(\alpha^c)),\ obs_k - to_k(\alpha^c) \leq per(tc) \text{ and } ha_k(fr_k(\alpha^c), tc) \geq incer(c) \,\},\ \{-\alpha^c\},\ (fr_k(\alpha^c), \emptyset, depth_k(fr_k(\alpha^c)), at_k(fr_k(\alpha^c))), \emptyset)$$

This function delivers the set of time-circs which must be destroyed, the formula which will be retained in the opp, a circumscription description (name, formulae, depth and action threshold), and an empty set of sources to be downgraded. α^c is the older formula in the pair in contradiction. The function is called when a circumscription has been found responsibe for a contradiction.

$PunishSource: \{ opps \} \times \text{Form} \rightarrow$

$$(\text{TimeCircs} \times \mathbb{P}(\text{Form}) \times (\text{Circs} \times \mathbb{P}(\text{Form}) \times [0,1] \times C) \times \mathbb{P}(\text{Sources}))$$

$$PunishSource(k, \alpha^c) = (\emptyset, \{\alpha^c\}, (fr_k(\alpha^c), \emptyset, depth_k(fr_k(\alpha^c)), at_k(fr_k(\alpha^c))), \{who_k(\alpha^c)\})$$

This function delivers the (empty) set of time-circs which must be destroyed, the formula which will be retained in the opp, a circ description (name, formulae, depth and action threshold), and a set of sources to be downgraded. α^c is the older formula in the pair in contradiction. This is called when a source is blamed for a contradiction.

$resolve$: $\{ opps \} \times$ Form $\rightarrow$

(TimeCircs $\times$ $\mathbb{P}$(Form) $\times$ (Circs $\times$ $\mathbb{P}$(Form) $\times$ $[0,1] \times [0,1] \times C$) $\times$ $\mathbb{P}$(Sources)) $\times$

(Circs $\times$ Sources $\times$ {"c","s"} $\times C$)

$resolve(k, \alpha^c) = ((\emptyset, \{ \alpha^c, \sim\alpha^c \}, \text{CircCreate}_k, \emptyset), (fr_k(\alpha^c), who_k(\alpha^c), \text{"c"}, c))$,

if $r = 1$ and $CreateCirc_k$,

$= (PunishCirc(k, \alpha^c), (fr_k(\alpha^c), who_k(\alpha^c), \text{"c"}, c))$,

if $r \neq 1$ or $\sim CreateCirc_k$, and

$valcirc(k, \alpha^c) > valsource(k, \alpha^c)$,

$= (PunishSource(k, \alpha^c), (fr_k(\alpha^c), who_k(\alpha^c), \text{"s"}, c))$, o.w.,

where $r = \pi_1(alarm_k(bl)) / \pi_2(alarm_k(bl))$, and $bl = (c \geq at_k(fr_k(\alpha^c)))$.

This function is responsible for controlling the punishment of offending circumscriptions or sources following a contradiction. α^c is the statement found to be in contradiction and is the original statement - that is, the one observed first of α^c and $\sim\alpha^c$.

$lkbinc_k$: Bool $\rightarrow$ $\mathbb{N}$

$parinc_k$: Bool $\rightarrow$ $\mathbb{N}$

Book-keepers: these two parameters control the action that *inconsistency* can take and are set by *balance*. The boolean values are to differentiate between records above and below the action threshold.

$inconsistency$: $\{ opps \} \times$ Bool $\times$ (Circs $\times$ Sources $\times$ {"c","s"} $\times C$) $\rightarrow$

$[0,1] \times \mathbb{N} \times [0,1] \times \mathbb{N}$

$inconsistency\,(k, bl, (circ, sou, tag, c))$

$= (incer(c), \lfloor r.\, lkbinc_k(bl) \rfloor, wgt_k(sou), 0),$

if $tag =$ "c",

$= (minhom_k(circ), 0, wgt_k(sou), \lfloor r.\, parinc_k(bl) \rfloor),$

if $tag =$ "s" and $bl = F$,

$= (minhom_k(circ), 0, wgt_k(sou) \cdot r, \lfloor r.\, parinc_k(bl) \rfloor),$

o.w.,

where $r = \pi_1(alarm_k(bl)) / \pi_2(alarm_k(bl))$.

This function takes action when inconsistencies occur too quickly. It delivers in the first part of the result, the new *minhom* value for *circ*, in the second part is the increment to *lkb* for *circ*. The third part is the new *wgt* value for *sou*, and the final value is the increment to the parole value, *par*, for *sou*. Its arguments are the boolean for whether or not the contradiction is above the action threshold, the circumscription, source, responibility tag and the certainty of the offending formula.

$Incpar_k$: Bool × Bool → $\mathbb{N} \times \mathbb{P}$(Sources)

Book-keeper: records the number of calls to *Incons* which have resulted in source being punished, and which sources have been hit. The first boolean is for trial or no-trial, the second is for above or below action threshold.

$Incacts_k$: Bool → $\mathbb{N} \times \mathbb{N}$

Book-keeper: records the number of successful calls to *action* that have resulted in a source being upgraded, and the total of the used budget on all these calls. The boolean is for trial, no-trial.

$Incactc_{\kappa}$: Bool $\rightarrow \mathbb{N} \times \mathbb{N}$

Book-keeper: records the number of successful calls to *action* that have resulted in a circ look-back being reduced, and the total of the used budget on all these calls. The boolean is for trial, no-trial.

$Inclkb_{\kappa}$: Bool $\times$ Bool $\rightarrow \mathbb{N} \times \mathbb{P}$(Circs)

Book-keeper: keeps track of the number of calls to *Incons* which have resulted in source being punished, and which circs have been hit. The first boolean is for trial or no-trial, the second is for above or below action threshold.

$actfail_{\kappa}$: Bool $\rightarrow \mathbb{N} \times \mathbb{N} \times \mathbb{N}$

Book-keeper: tracks the number of times action failed to act due to lack of budget, the maximum value of the failed budget in sources and in circs. The boolean is for trial, no-trial.

$trial_{\kappa}$: Bool

Book-keeper: true if there is currently a trial.

$trilen_{\kappa}$: $\mathbb{N}$

Book-keeper: saves the current length assigned to trials.

$torun_K$: **N**

Book-keeper: saves, when a trial is running, a count of how long there is to go - when it reaches 0 the trial stops.

$ProIncp$: Bool × { $opps$ } → **N** × **N**

$$ProIncp(bl,k) = (\pi_1(Incpar_K(T,bl)) - |\pi_2(Incpar_K(T,bl))| + lev_K(bl),$$
$$\lfloor (1 + |\pi_2(Incpar_K(T,bl))| / \pi_1(Incpar_K(T,bl))).parinc_K(bl) \rfloor)$$

This function is used to deliver the case for increasing $parinc(bl)$ (thus, there are two of these cases). The first value is the priority measure, while the second is the requested new value for $parinc_K(bl)$

$ProSource$: { $opps$ } → **N** × **N** × **N**

$$ProSource(k) = (\pi_1(Incacts_K(T)), \lfloor r.acts_K \rfloor, \lfloor r.parinc_K(T) \rfloor),$$

where $r = (1 - \pi_2(Incacts_K(T))/\pi_1(Incacts_K(T)))$.

This function delivers the case for reducing the parameters that relate to sources, so that they are not punished so severely and at the same time, $action$ cannot retrieve sources from such a severe punishment.

$ProIncl$: Bool × { $opps$ } → **N** × **N**

$$ProIncl(bl,k) = (\pi_1(Inclkb_K(T,bl)) - |\pi_2(Inclkb_K(T,bl))| + lev_K(bl),$$
$$\lfloor (1 + |\pi_2(Inclkb_K(T,bl))| / \pi_1(Inclkb_K(T,bl))).lkbinc_K(bl) \rfloor)$$

The same as for $ProIncp$ but working on look-back.

$ProCirc$: { $opps$ } → **N** × **N** × **N**

$$ProCirc(k) = (\pi_1(Incactc_k(T)), \lfloor r.actc_k \rfloor, \lfloor r.lkbinc_k(T) \rfloor),$$

where $r = (1 - \pi_2(Incactc_k(T))/\pi_1(Incactc_k(T)))$.

As for *ProSource*, but using circs instead of sources.

$$ProAct: \{opps\} \to \mathbf{N} \times \mathbf{N} \times \mathbf{N}$$

$$ProAct(k) = (\pi_1(actfail_k(T)), \pi_2(actfail_k(T)) + acts_k, \pi_3(actfail_k(T)) + actc_k)$$

This function is in favour of *action*, giving a case for the increase in *acts* and *actc*, using the apparent deficiencies in the current budget as a guide to what to ask for.

$$reduct: \{opps\} \times \text{Form} \to \mathbf{N} \cup \{-1\}$$

$$reduct(k, \alpha^c) = clock_k(T) + wgt_k(who_k(\alpha^c)).clock_k(F) + perf_k(who_k(\alpha^c)) - par_k(c \geq at_k(fr_k(\alpha^c)), who_k(\alpha^c)),$$

if $accert(cg_k(who_k(\alpha^c)) + 1, c) \geq at_k(fr_k(\alpha^c))$,

$= -1$, otherwise.

This function computes what reduction would be necessary in the performance required for a source to be upgraded, if the source were to be reliable enough now to act on.

$$accert: \mathbf{N} \times C \to C$$

This function takes the current grading of a source and a certainty value, and delivers the certainty which the system is prepared to assign to the appropriate formula, based on the current reliability of the source.

The following type is now used: BasicForm = $\{\alpha : \alpha^c \in \text{Form}\}$.

$action : \{opps\} \times \text{BasicForm} \to \mathbb{P}(\text{Sources} \times \mathbb{N}) \times \mathbb{P}(\text{Circs} \times \mathbb{N}) \times \mathbb{N} \times \mathbb{N} \times \text{Bool}$

$action(k, \alpha) = (\emptyset, \emptyset, 0, 0, F)$, if $\forall c \in C.\ (c \geq at_k(fr_k(\alpha^c)) \Rightarrow \alpha^c \notin sm_k)$,
and, if $\alpha^c \in p(k)$, and $to_k(\alpha^d)$ is defined, for some $d \geq at_k(\alpha^c)$, then:
$\forall n.\ (1 \leq n \leq lkb_k(fr_k(\alpha^c)) \Rightarrow pcalc(to_k(\alpha^d), filter(k, fr_k(\alpha^c), n), n,$
$exc(incer(at_k(fr_k(\alpha^c)), depth_k(fr_k(\alpha^c)),$
$|filter(k, fr_k(\alpha^c), n)|)) + to_k(\alpha^d) < obs_k$

$= (\emptyset, \emptyset, reduct(k, \alpha^c) - acts_k, 0, T)$,
if $\exists c \in C.\ (c \geq at_k(fr_k(\alpha^c)) \wedge \alpha^c \in sm_k)$ and
$reduct(k, \alpha^c) > acts_k$,

$= (\{(who_k(\alpha^c), reduct(k, \alpha^c))\}, \emptyset, 0, 0, T)$,
if $\exists c \in C.\ (c \geq at_k(fr_k(\alpha^c)) \wedge \alpha^c \in sm_k)$ and
$reduct(k, \alpha^c) \leq acts_k$,

$= (\emptyset, \{(fr_k(\alpha^c), lkb_k(fr_k(\alpha^c)) - n)\}, 0, 0, T)$,

if $\forall c \in C.\ (c \geq at_k(fr_k(\alpha^c)) \Rightarrow \alpha^c \notin sm_k)$,
and $\alpha^c \in p(k)$, and $to_k(\alpha^d)$ is defined for some $d \geq at_k(\alpha^c)$, so that:
$\exists n.\ (1 \leq n \leq lkb_k(fr_k(\alpha^c)) \Rightarrow pcalc(to_k(\alpha^d), filter(k, fr_k(\alpha^c), n), n,$
$exc(incer(at_k(fr_k(\alpha^c)), depth_k(fr_k(\alpha^c)),$
$|filter(k, fr_k(\alpha^c), n)|)) + to_k(\alpha^d) < obs_k$
and $lkb_k(T, fr_k(\alpha^c)) - n \leq actc_k$

$= (\emptyset, \emptyset, 0, lkb_k(T, fr_k(\alpha^c)) - n - actc_k, T)$, o.w..

This function is called when an action is not available immediately, to see if some source has entered the information needed, but it has been downgraded, or if the information has deteriorated in its circumscription. In either case the function checks if it has sufficient power to upgrade the

appropriate value. The function returns five values. The final value indicates whether there was any possibility for action. The first and second values indicate the source or circumscription which is affected and how much either look-back or par values must be altered. The third and fourth values are used when action might have been taken but the freedom to act is too restrictive. They show by how much action was short of the mark.

$caution : \mathbb{N}$

This is a motivation towards caution.

$act, know, talk : \mathbb{N}$

Motivations in each area.

$mot : \{$ "actcacts+", "lkbinc$T$+", "lkbinc$F$+", "actclkbinc$T$−",
"parinc$T$+", "parinc$F$+", "actsparinc$T$−", "nexttry" $\} \rightarrow \mathbb{N}$

This gives the "motivation" towards each of these values. In this system these are primitives, but in a more sophisticated system they would be provided by a higher (modifiable) function based on the more primitive motivation of the system.

$susp_{\kappa} : \mathbb{P}(\text{Form})$

Book-keeper: records the set of formulae being monitored for *danger*.

$over_{\kappa} : \text{BasicForm} \nrightarrow \mathbb{P}(\mathbb{N} \times C \times \mathbb{N})$

Book-keeper: keeps the overriding histories of formulae. The values stored

are the number of this override, the current certainty of the formula, and the ob-state in which it occurred.

$$extrap : \mathbf{P}(\mathbf{R} \times \mathbf{R}) \times \mathbf{N} \rightarrow \mathbf{N}$$

$$extrap(X, n) = m$$

where m is the first integer value for which the corresponding value in the best-fit line computed by extrapolating the values given in X exceeds n.

An algorithm for this can be found in [RA57], say.

$$danger: \{ opps \} \rightarrow \mathbf{P}(\text{Form} \times \mathbf{N})$$

$$\begin{aligned} danger(k) = \{ (\alpha^c, n) : \alpha^c \in \cup \{ worlds_k(circ) : circ \in circs_k \}, \\ extrap(\{ (\pi_1(x), \pi_2(x)) : x \in over_k(\alpha^c) \}, at_k(fr_k(\alpha^c)) = \\ max(\{ \pi_1(x) : x \in over_k(\alpha^c) \}) + 1, \\ \text{and} \\ (extrap(\{ (\pi_1(x), \pi_2(x)) : x \in over_k(\alpha^c) \}, at_k(fr_k(\alpha^c)) \\ - max(\{ \pi_1(x) : x \in over_k(\alpha^c) \})).caution = n \} \end{aligned}$$

This function decides whether there is any danger in an oscillating series of inputs converging on the action threshold for the appropriate circumscription. It delivers the formulae to be suspended and the time for which suspension is requested.

$$suspend_k : \text{Form} \nrightarrow \mathbf{N}$$

Book-keeper: records current times of suspension for formulae. They are decreased, if non-zero, until they reach zero, when they are eligible for readmission.

$trileve_{k}$: Bool $\rightarrow$ $\mathbb{P}$ ($\mathbb{N}$ $\times$ {*opps*})

Book-keeper: keeps records during the trials of the two graphs levels. True is used for the current graph, false for the dummy or on-trial graph.

$numtrials_{k}$: $\mathbb{N}$

Book-keeper: records the number of trials that have been carried out since previous reset.

$shorter_{k}$: $\mathbb{N}$

Book-keeper: records the sum of the required lengths of each trial.

$longer_{k}$: $\mathbb{N}$

Book-keeper: records the number of times an action previously made is subsequently undone (since last reset).

$done_{k}$: $\mathbb{P}$ ({ "actcacts+","lkbinc$T$+","lkbinc$F$+","actclkbinc$T$-",
"parinc$T$+","parinc$f$+","actsparinc$T$-" } $\times$ $\mathbb{N}$ $\times$ $\mathbb{N}$ $\times$ {*opps*})

Book-keeper: records what action has been taken by *balance*. Firstly to what, secondly, how much (possibly a pair of values) and thirdly, when.

mi_{k} : $\mathbb{N}$ $\times$ $\mathbb{N}$

Book-keeper: stores sum of certainties of input information and sum of certainties of accepted information. This is reset at the start of each trial.

$lentrial: \{opps\} \rightarrow \{1, 0, -1\}$

$lentrial(k) = 1$, if $longer_k > shorter_k / trilen_k$ and $numtrials_k > tests$,

$= -1$, if $shorter_k / trilen_k > longer_k$ and $numtrials_k > tests$,

$= 0$, o.w..

This function requests increases or decreases in the lengths of trials (implemented by *link* in the record of *trilen*).

$tests: \mathbb{N}$

The number of trials used to decide whether to increase or decrease the length of subsequent trials.

$tries_k: \mathbb{N} \times \{$ "actcacts+", "lkbinc$T$+", "lkbinc$F$+", "actclkbinc$T$-", "parinc$T$+", "parinc$F$+", "actsparinc$T$-" $\} \times \mathbb{N} \times \mathbb{N}$

Book-keeper: used to store current contenders information.

$tried_k: \{$ "actcacts+", "lkbinc$T$+", "lkbinc$F$+", "actclkbinc$T$-", "parinc$T$+", "parinc$F$+", "actsparinc$T$-" $\} \times \mathbb{N} \times \mathbb{N} \times$ Bool

Book-keeper: used to store the last trial record.

$value: \{opps\} \times \{$ "actcacts+", "lkbinc$T$+", "lkbinc$F$+", "actclkbinc$T$-", "parinc$T$+", "parinc$F$+", "actsparinc$T$-", "nextlry" $\} \rightarrow \mathbb{N}$

$$value(k,x) = mot(x).y, \qquad \text{if } x \neq \text{"nexttry"},$$
$$= mot(\pi_2(tries_k)).\pi_1(tries_k), \qquad \text{o.w.},$$

where:

$$y = \pi_1(ProAct(k)), \qquad \text{if } x = \text{"actcacts+"},$$
$$= \pi_1(ProIncl(T,k)), \qquad \text{if } x = \text{"lkbinc}T+\text{"},$$
$$= \pi_1(ProIncl(F,k)), \qquad \text{if } x = \text{"lkbinc}F+\text{"},$$
$$= \pi_1(ProCirc(k)), \qquad \text{if } x = \text{"actclkbinc}T-\text{"},$$
$$= \pi_1(ProIncp(T,k)), \qquad \text{if } x = \text{"parinc}T+\text{"},$$
$$= \pi_1(ProIncp(F,k)), \qquad \text{if } x = \text{"parinc}F+\text{"},$$
$$= \pi_1(ProSource(k)), \qquad \text{if } x = \text{"actsparinc}T-\text{"}.$$

This function is used by *win* in determining the most needful case for trial.

$$win: \{opps\} \rightarrow \{\text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"}, \text{"actclkbinc}T-\text{"},$$
$$\text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"}, \text{"nexttry"}\}$$

$$win(k) \in \{x \in \{\text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"}, \text{"actclkbinc}T-\text{"},$$
$$\text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"}, \text{"nexttry"}\}:$$
$$value(k,x) \text{ is maximal for } x \in \{\text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"},$$
$$\text{"actclkbinc}T-\text{"}, \text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"}, \text{"nexttry"}\}\}$$

This function decides which change to give trial to.

$$nexttry: \{opps\} \rightarrow \{\text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"}, \text{"actclkbinc}T-\text{"},$$
$$\text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"}\} \times \mathbb{N} \times \mathbb{N} \times \text{Bool}$$

$$nexttry(k) = (\pi_2(tries_k), av_1, av_2, F)$$

where av_1 and av_2 are the averages of the $\pi_3(tries_k)$ and $\pi_4(tries_k)$ with their appropriate corresponding value in the system, depending on $\pi_2(tries_k)$.

This function determines the next values for a trial if this is requested, from the last trial.

$$balance: \{ opps \} \rightarrow (\{ \text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"}, \text{"actclkbinc}T-\text{"}, \text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"} \} \times \mathbb{N} \times \mathbb{N} \times \text{Bool}) \times (\mathbb{N} \times \{ \text{"actcacts+"}, \text{"lkbinc}T+\text{"}, \text{"lkbinc}F+\text{"}, \text{"actclkbinc}T-\text{"}, \text{"parinc}T+\text{"}, \text{"parinc}F+\text{"}, \text{"actsparinc}T-\text{"} \} \times \mathbb{N} \times \mathbb{N})$$

$$balance(k) = (try, totry),$$

where:

$$\begin{aligned}
try = (&\text{"actcacts+"}, \pi_2(ProAct(k)), \pi_3(ProAct(k)), T), && \text{if } win(k) = \text{"actcacts+"},\\
= (&\text{"lkbinc}T+\text{"}, \pi_2(ProIncl(T,k)), 0, T), && \text{if } win(k) = \text{"lkbinc}T+\text{"},\\
= (&\text{"lkbinc}F+\text{"}, \pi_2(ProIncl(F,k)), 0, T), && \text{if } win(k) = \text{"lkbinc}F+\text{"},\\
= (&\text{"actclkbinc}T-\text{"}, 1, 1, T), && \text{if } win(k) = \text{"actclkbinc}T-\text{"},\\
= (&\text{"parinc}T+\text{"}, \pi_2(ProIncp(T,k)), 0, T), && \text{if } win(k) = \text{"parinc}T+\text{"},\\
= (&\text{"parinc}F+\text{"}, \pi_2(ProIncp(F,k)), 0, T), && \text{if } win(k) = \text{"parinc}F+\text{"},\\
= (&\text{"actsparinc}T-\text{"}, 1, 1, T), && \text{if } win(k) = \text{"actsparinc}T-\text{"},\\
= &\ nexttry(k), && \text{if } win(k) = \text{"nexttry"},
\end{aligned}$$

and:

$$\begin{aligned}
totry = &(\pi_1(ProAct(k)), \text{"actcacts+"}, 1, 1), && \text{if } win(k) = \text{"actcacts+"},\\
= &(\pi_1(ProIncl(T,k)), \text{"lkbinc}T+\text{"}, 1, 1), && \text{if } win(k) = \text{"lkbinc}T+\text{"},\\
= &(\pi_1(ProIncl(F,k)), \text{"lkbinc}F+\text{"}, 1, 1), && \text{if } win(k) = \text{"lkbinc}F+\text{"},\\
= &(\pi_1(ProCirc(k)), \text{"actclkbinc}T-\text{"}, \pi_2(ProCirc(k)), \pi_3(ProCirc(k))), && \text{if } win(k) = \text{"actclkbinc}T-\text{"},\\
= &(\pi_1(ProIncp(T,k)), \text{"parinc}T+\text{"}, 1, 1), && \text{if } win(k) = \text{"parinc}T+\text{"},\\
= &(\pi_1(ProIncp(F,k)), \text{"parinc}F+\text{"}, 1, 1), && \text{if } win(k) = \text{"parinc}F+\text{"},\\
= &(\pi_1(ProSource(k)), \text{"actsparinc}T-\text{"}, \pi_2(ProSource(k)), \pi_3(ProSource(k))), && \text{if } win(k) = \text{"actsparinc}T-\text{"},\\
= &\ tries_k, && \text{if } win(k) = \text{"nexttry"}.
\end{aligned}$$

This function is the principal function used in setting up trials. It decides to run trials specifying the values that are to be tried.

$endtrial : \{opps\} \rightarrow$ Bool

$endtrial(k) = F,$
if $act.lev_k(T) + know.lev_k(F) > act.lev_k'(T) + know.lev_k'(F) + talk,$
$= T,$
if $act.lev_k(T) + know.lev_k(F) + talk > act.lev_k'(T) + know.lev_k'(F),$
$= (\pi_2(mi_k)/\pi_1(mi_k) \geq \pi_2(mi_k')/\pi_1(mi_k'))$, o.w..

Here, the use of ' is to distinguish the dummy graph from the real one (unmarked). A true result indicates a successful trial for the dummy. Note that when the trial ends, if it fails then the value of $tries_k$ must be updated - if $tried_k$ ends with T then the value of $tries_k$ is updated to give it zero priority, since the first trial in the series failed. Otherwise, $tries_k$ is updated to he $tried_k$, but for the boolean value. This allows the next trial value (if a continuation of this trial) to be reduced. Also note that if the trial fails the values of all the records are updated from the true and false parts together (i.e. from the trial/no-trial parts put together) while if the trial is successful, only the ones not used by the Pro running a trial are updated. The other is reset.

lcf_k: $\mathbb{P}$(Form)

Book-keeper: recording low certainty formulae that are being maintained to check how dangerous they might have been had they been accepted into the *opp*. Formulae in lcf are all those with certainty lower than $cert(minhom_k(circ))$.

$infoc_k$: Circs $\times$ $C \rightarrow \mathbb{N}$

Book-keeper: delivering the number of times a formula in lcf has passed without contradiction (through its persistency period) and for which there have been no contradictions at a higher certainty, since start of count (that is, the count is zeroed for certainty c and higher, when a contradiction occurs at certainty c).

$belief: \{opps\} \rightarrow \mathbf{P}(\text{Circs} \times [0,1])$

$belief(k) = \{(circ, incer(c)) : circ \in circs_k, infoc_k(circ, c) \geq bel$ and
$\forall d \in C. infoc_k(circ, d) < bel\}$, if $clock_k(T) \neq 0$

$= \emptyset$, if $clock_k(T) = 0$.

This function decides which circumscritions are eligible for having their $minhom$ values raised.

$accept: \{opps\} \times \text{Form} \times \text{Sources} \times \text{Circs} \rightarrow$
$\mathbf{P}(\text{Form}) \times \mathbf{P}(\text{Form}) \times \mathbf{P}(\text{Form}) \times \mathbf{P}(\text{Form})$

$accept(k, \alpha^c, source, circ) = (E, F, G, H)$

where:

$E = \{\alpha^d\}, F = G = \emptyset$, if $d \geq cert(minhom_k(circ))$ and $\forall e \in C. \alpha^e \notin susp_k$,

$F = \{\alpha^d\}, E = G = \emptyset$, if $d \geq cert(minhom_k(circ))$ and $\exists e \in C. \alpha^e \notin susp_k$,

$E = F = \emptyset$, $G = \{\alpha^d\}$, if $d < cert(minhom_k(circ))$,

$H = \emptyset$, if $c = d$,

$H = \{\alpha^d\}$, if $c \neq d$,

$d = incer(minhom_{\kappa}(circ))$.

This function determines the acceptances for the system. The first set is destined for the object level (in fact, the opp), the second for suspension set, the third for lcf and the fourth for sm.

$object: \{ opps \} \rightarrow \{ opps \} \times \mathbb{P}(\text{Form}) \times \mathbb{P}(\text{Form})$

This function takes one opp and delivers the next using the axioms of the system. These are mostly the same as the axioms of the Unforgiving Model, except that there are white triangles in addition to the black triangles used there. These are used to indicate an override rather than a contradiction and it is important for *resolve* to learn about this. The formulae that are black triangled are in the second part of the result, while the white triangle formulae are in the third and final part. When contradictions or overrides occur between formulae entered by different sources, the funtion delivers nothing - the system collapses.

$ident: \mathbb{P}(\text{Form}) \rightarrow \mathbb{P}(\text{Form}) \times \mathbb{P}(\text{Form}) \times \mathbb{P}(\text{Form})$

This function applies the rule of identity to the formulae in its set using all the appropriate side issues that relate. Black and white triangles can be produced and it is the responsibility of *link* to ensure that the right functions are informed. The black and white triangle formulae are put in the second and third sets delivered.

References

[AD74] Ajdukiewicz K., *Pragmatic Logic*, Reidel 1974.

[AM85] Amarel S., *Expertise Acquisition and Theory Formation*, Machine Learning Conference 85, London, 1985.

[BK82] Bowen, K. A. and Kowalski R. A., *Amalgamating Language and Metalanguage in Logic Programming*, in [CL82]

[BL84] Brown J. S. and Lenat D. B., *Why AM and Eurisko Appear to Work*, A.I., 1984.

[BM77] Bell J. L. and Machover M., *A Course in Mathematical Logic*, North-Holland, 1977.

[BU83] Bundy A., *The Computer Modelling of Mathematical Reasoning*, Academic Press, 1983.

[BU84] Bundy A., *Meta-level Inference and Consciousness*, in [TO84].

[BU85] Bundy A., *Incidence Calculus: A Mechanism for Probabilistic Reasoning*, J. Automated Reasoning, 1, 1985.

[BY86] Bouchon B. and Yager R. R. (eds.), *International Conference on Information Processing and Management of Uncertainty*, Paris 30 June–4 July 1986.

[CH86] Chidgey J. R., *Relevant Tense Logic*, Temporal Logic and its Applications Conference, Leeds, 1986.

[CHO84] Choraqui E., *Computational Models of Reasoning*, in [TO84].

[CL82] Clark K. L. and Tarnlund S. A. (ed.), *Logic Programming*, Academic Press, 1982.

[CO83] Constable R. L., *Constructive Mathematics as a Programming Logic I: Some Principles of Theory*, in LNCS 158, Springer-Verlag 1983.

[DA79] Dalla Chiara M. L., *Logica*, Mondadori 1979.

[DAT73] Dalla Chiara M. L. and Toraldo di Francia G., *A Logical Analysis of Physical Theories*, Rivista del Nuovo Cimento, serie 2, Vol. 3.

[DAV80] Davis M., *The Mathematics of Non-Monotonic Reasoning*, A.I. 13, 1980.

[DI85] Dietterich T., *Machine Learning: Problems and Methods*, Machine Learning Conference 85, London, 1985.

[DL82] Davis R. and Lenat D. B., *Knowledge Based Systems in Artificial Intelligence*, Mc Graw-Hill, 1982.

[DU77] Dummett M., *Elements of Intuitionism*, Clarendon Press, 1977.

[FA86] Fargues J., *Logiques Non Classiques comme Bases pour une Typologie des Moteurs d'Inferences*, in [BY86].

[FE79] Feyerabend P., *Against Method: Outline of an Anarchist Theory of Knowledge*, Verso, 1979.

[FI69] Fitting M. C., *Intuitionistic Logic, Model Theory and Forcing*, North Holland, 1969.

[GA82] Gabbay D. M., *Intuitionistic Basis for Non-Monotonic Logic*, in LNCS 138, Springer-Verlag, 1982.

[GH86] Green Hall N., *Strategic Planning with Uncertain Values*, in [BY86].

[GL83] Gallaire H. and Lassere C., *Metalevel Control for Logic Programs*, in [CT83].

[GO84] Goldblatt R., *Topoi*, North Holland, 1984.

[GR86] Garigliano R., *A Non-Classical Inference Engine for Expert Systems*, in [BY86].

[GRL86] Garigliano R. and Long D., *Reasoning by Analogy: a Formal Model*, in Tenth Annual Conference on Microprocessor Applications, Strathclyde 8-10 Sep., 1986.

[HAA78] Haack S., *Philosophy of Logic*, Cambridge University Press, 1978.

[HAL86] Halpern J. (ed.), *Theoretical Aspects of Reasoning about Knowledge*, Kaufmann, 1986.

[HAU85] Haugeland J., *Artificial Intelligence - The Very Idea*, MIT Press, 1985.

[HAY77] Hayes P., *In Defense of Logic*, Proc. Ijcai 5, 1977.

[HEY56] Heyting A., *Intuitionism: an Introduction*, North Holland, 1956.

[HHT] Hansson A., Haridi S. and Tarnlund S. A., *Properties of a Logic Programming Language*, in [CL82].

[HO84] Hodges W., *Model Theory*, draft, 1984.

[HS84] Hoare C. A. and Shepherdson J. C. (ed.), *Mathematical Logic and Computer Programming*, Prentice Hall, 1984.

[KL52] Kleene S. C., *Introduction to Metamathematics*, 1952.

[KU62] Kuhn T. S., *The Structure of Scientific Revolutions*, University of Chicago Press, 1962.

[LA76] Lakatos I., *Proofs and Refutations - The Logic of Mathematical Discovery*, Cambridge University Press, 1976.

[LM70] Lakatos I. and Musgrave A. (eds.), *Criticism and the Growth of Knowledge*, Cambridge University Press, 1970.

[LU70] Lukasiewicz J., *Selected Works*, 1970.

[MC69] McCarthy J. and Hayes P. J., *Some Philosophical Problems from the Standpoint of A.I. Machine Intelligence* 4, (Meltzer and Michie eds.), Edinburgh University Press, 1969.

[MC80] McCarthy J., *Circumscription: a Form of Non Monotonic Reasoning*, in A.I. 13, 980.

[MD80] McDermott D. and Doyle J., *Non-Monotonic Logic I*, in A.I. 13, 1980.

[MI85] Michalski R. S. and Winston P. H., *Variable Precision Logic*, A.I. Memo, Artificial Intelligence Laboratory, MIT, 1985.

[ML82] Martin-Lof P., *Constructive Mathematics and Computer Programming*, 1982, in [HS84].

[NE81] Negoita C. V., *Fuzzy Systems*, Abacus Press, 1981.

[NO75] Noto A., *Le Logiche Non Classiche*, Bulzoni 1975.

[NU83] Nutter T., *What Else is Wrong with Non Monotonic Logics? Representational and Informational Shortcomings*, in Proc. to Fifth Annual Conference of the Cognitive Science Society, 1983.

[PE84] Pearl J., *Heuristics*, Addison-Wesley, 1984.

[PE86] Pearl J., *Probabilistic Reasoning Using Graphs*, in [BY86].

[PI86] Pitrat J., *Using Knowledge to Use Knowledge*, in [BY86].

[PO72] Popper K., *The Logic of Scientific Discovery*, Hutchinson, 1972.

[PR65] Prawitz D., *Natural Deduction*, Alwquist and Wiksell, 1965.

[RA57] Raisford H. F., *Survey Adjustments and Least Sqares*, Constable and Co. 1957.

[RE38] Reichenbach H., *Experience and Prediction. An Analysis of the Foundation and the Structure of Knowledge*, University of Chicago Press, 1938.

[RE180] Reiter R., *A Logic for Default Reasoning*, in A.I. 13, 1980.

[RL86] Des Rivieres J. and Levesque H., *The Consistency of Syntactical Treatments of Knowledge*, in [HA86].

[RZ86] Ras Z. and Zemankova M., *On Learning, a Possibilistic Approach*, Proc. of 1986's CISS, Princeton, 1986.

[SA85] Santambrogio M., *Generic and Intensional Objects*, Unpublished paper, 1985.

[SCA77] Schank R. C. and Abelson R. P., *Scripts, Plans, Goals and Understanding*, Erlbaum, 1977.

[SE85] Selfridge O., *Invited Talk*, Machine Learning Conference 85, London, 1985.

[SH76] Shafer G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SP86] Spiegelhalter D. J., *Probabilistic Reasoning in Predictive Expert Systems* in *Uncertainty in Artificial Intelligence*, (Kanal and Lemmer eds), North Holland, 1986.

[ST85] Steels L., *The Use of Inconsistent Logic in Learning*, Machine Learning Conference 85, London 1985.

[TO84] Torrance S. (ed.), *The Mind and the Machine*, Ellis Horwood, 1984.

[WI80] Winograd T., *Extended Inference Modes in Reasoning by Computer Systems*, in A.I. 13, 1980.

[ZE84] Zemankova M., and Kandel A. , *Fuzzy Relational Database - a Key to Expert Systems*, Verlag TUV Rheinland, 1984.