MODELLING CAUSAL REASONING

Sergio Brandano

Being a monograph submitted to the University of London
for the degree of Doctor of Philosophy.

MODELLING CAUSAL REASONING

Sergio Brandano

University of London
Queen Mary College

UNIVERSITY OF LONDON, SENATE HOUSE
Malet Street, London WC1E 7HU, England

The author certifies that this monograph and the research to which it refers are the product of his own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. The author is entirely responsible for study concept and design, acquisition of data, analysis and interpretation of data, drafting and critical revision of the manuscript.

Modelling Causal Reasoning/Sergio Brandano—pages 150, ~~word length 52793~~ —Includes bibliographical references. **Categories and subject descriptors:** F.1, F.4, I.2.3 [ACM CS 1998]. **Additional key words and phrases:** causal logic (science of causal reasoning), methods of research, classification of models, relations of equivalence and subsumption between models; causal calculators, problem independence, language independence.

Copy-edited and typeset by the author
Printed by Collis-Bird & Withey, London, England

Abstract

Although human causal reasoning is widely acknowledged as an object of scientific enquiry, there is little consensus on an appropriate measure of progress. Up-to-date evidence of the standard method of research in the field shows that this method has been rejected at the birth of modern science.

We describe an instance of the standard scientific method for modelling causal reasoning (causal calculators). The method allows for uniform proofs of three relevant computational properties: correctness of the model with respect to the intended model, full abstraction of the model (function) with respect to the equivalence of reasoning scenarios (input), and formal relations of equivalence and subsumption between models. The method extends and exploits the *systematic paradigm* [*Handbook of Logic in Artificial Intelligence and Logic Programming*, volume IV, p. 439-498, Oxford 1995] to fit with our interpretation of it.

Using the described method, we present results for some major models, with an updated summary spanning seventy-two years of research in the field.

ACKNOWLEDGEMENTS

> «Peer review is anonymous, and thus write your reviews as you like, abusing the author if you wish. Peer review comes once, and thus write your camera-ready copy as you like, changing the content if you wish. Nobody will complain. This is the standard practice.»
>
> (ZH, APRIL 1999)

Thus spoke my senior colleague and room-mate in London. I did not receive his words of wisdom. Years passed by, until April 2002, when a letter from Elsevier Science rejected my classification of a family of symbolic logics, being an invited paper. The reviewer displayed lack of expertise with the method of research and its known results, only to abuse the author by «suggesting», through rejection, «to skip the proofs» in favour of «test examples», being my paper full of proofs. Such review, dated April 18th, was followed by my comments on April 26th. After two months of silence, I had to call Elsevier's Secretariat. I was told that they received the letter, but did not manage to contact the editor, and thus suggested to submit elsewhere. This was my first submission to the very press which delivered one of Galilei's works from dogmatic Inquisition. Four hundred years later, to my formal appeal against dogma and abuse, the same press did not display the slightest hint of embarrassment. I had similar rejections, but none so upsetting. However, as my pre-press readers kept questioning the very method of research, there was no purpose in describing new results. I had to change my method of research, because I could no longer publish with it.

I found myself reading in the words of my anonymous pre-press readers, to understand what they meant. All I had from them was the «suggestion» to «skip the proofs» in favour of «test examples». I then looked for enlightenment elsewhere, reading more closely the classics on method, and collecting descriptions by other authors in my field. In September 2002 a report on Peer Review by the British *Parliamentary Office of Science and Technology* confirmed my feeling when stating that «some scientists believe that anonymity provides an opportunity for settling old scores and burying rival research». While reading from the British Parliament, I recollected the original words of wisdom by my colleague. I still did not receive them, and kept reading. While reading Kuhn, and a monograph on Scientific Controversies by Engelhardt and Caplan, I also learned that a possi-

ble closure happens by the natural death of one of the opponents. I began having Orwellian nightmares as soon as I merged the above with the «Publish or Perish» dogma which is forced upon us by the publishing business. I saw authors exploited as sheep by the editorial economic powerhouse. I saw rival authors intellectually murdered under the dogmatic and occlusive system of anonymous pre-press review. I saw authors becoming editors to grant themselves an upper level of fairness in peer review. I felt increasingly sick. Where they only nightmares? The filling of the relevant session with papers by the very editor of a recent IJCAI congress now seemed less than casual. Further, I had the certainty of being exploited when I realised that IEEE acquired the copyright of one of my works using the carrot-and-stick approach, namely by threatening me not to include my accepted work in the proceedings while offering broad distribution via their «not-for-profit» publishing system. I discovered, in fact, that several well-known book retailers were selling the IEEE proceedings using a highlight of my included paper as advertisement. I then realised that copyright is a certificate of property, and the assets of each known academic press consisted in millions of such properties, acquired royalty free using the carrot-and-stick approach. In so doing, it was no longer a surprise that the financial statement of academic publishers revealed multi-million-euro businesses, with Elsevier Science and other publishers now hitting the news because of their greed [21, 137, 3]. It also became clear to me that I was not free to set my own research agenda, because I could not pay for my own research via teaching, consulting and royalties. Indeed, still using the carrot-and-stick approach, authors are discouraged to teach by the low wages, and encouraged to apply for research grants, through yet another «peer review» system of research proposals. As money goes where anonymous peers want, the authors choose those research topics for which money is available, and the only money available is through such research grants. I realised that, with my independent work, I was moving against stream. Both the aim and method of my research were being opposed to at pre-press review. I also realised that authors of favoured research were still miserable, because their intellectual properties, namely copyright and patents, were acquired royalty free by the press and their funders respectively. Independent work leading to author owned copyright was impossible, due to the absence of both sufficient funds to produce it and of academic publishers who would pay royalties to the author to publish it. Independent work leading to author owned patents was also impossible, due to the unreasonable cost of patents. The legal system of patents has been invented in Italy as a means to help individual authors gain credit and profit from their useful inventions. However, as presently implemented, with its costs, the patent system moves against the individual authors and in favour of corporations, which

is contrary to the very aim for which the legal system of patents has been originally invented. To my eyes, the whole system was thoroughly engineered to control and exploit the work of researchers, ruled by the «carrot and stick» method of persuasion. The researchers did not earn the knowledge for themselves, and thus did not take any responsibility for it. I have also been plagiarised, both as researcher and consultant, and thus learned to beware of colleagues. As I turned out to be the sole author of my works, I have been refused a Lecturing chair by a committee who wanted a co-author. When listening to Apt's comments at Dijkstra's retirement banquet, I learned that nowadays, it is very rare to encounter researchers who put their own name on their own work only. I had confirmation of Apt's words when reading surveys on the ethics of coauthorship [69, 194, 193, 93]. Science turned out to be an Industry without Ethics. It became clear to me that I had to focus not only on my method of research but also on my method of conducting ethical business, to finally have credit and profit.

I fell ill. In September 2007, after years of trial and errors with various physicians and the side effects of their experimental therapies, I have been certified with a history of bipolar disorder in co-morbidity with anxiety and panic disorder. In my struggle for self control, I now had to reshape my whole life.

Once upon a time, I found myself surrounded by people inclined in truth fabrication; they were dogmatic, being always necessarily right, to pretend authority over me. They also were abusive, attacking me and denying my own vital sphere. In my struggle for understanding, I found myself observing that their truths did not correspond to my observations. I also found that they were contradicting themselves with their own actions. Their information was at times unsound or inconsistent. I despised the community I was living in, and felt a slave in it. I wanted to live in a nurturing community of peers, ruled by fairness and persuasion instead of force and conditioning. Using what was left of my self-esteem, I decided to conquer my own freedom by becoming a Scientist, rejecting the generations-old family business and its wealth. However, as a Scientist, I discovered that the scientific community is as dogmatic and abusive as the community I escaped from, and thus feel an unbearable pain. My quest for freedom shall continue.

This monograph is an exercise of Scientific research, being a report of independent and unbiased research in the field of human reasoning. It serves no trend, political or economic interest.

I am indebted to Edmund Robinson for a discussion on chapter 1, to Jon Rowson for reading the same chapter, to Murray Shanahan for reading the first result in chapter 5, and to Wilfrid Hodges for reading chapter 6.

I also acknowledge Leslie Lamport for useful explanations on his model, and Vladimir Lifschitz for useful explanations on language $\mathcal{A}$, $\mathcal{C}$ and $\mathcal{C}+$. Portions of the material have been presented at various fora in Europe, and gratitude is due to all those who have shared their comments. Scientific indebtedness is acknowledged by references to the Bibliography.

My most felt acknowledgements go to my physicians at the University Hospital of S.ta Chiara in Pisa, to my family and friends.

This work has been partly funded by the Institute through research, teaching and travelling funds, directly awarded to the author. The Institute had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

CONTENTS

PREFACE

Context of the problem

Four hundred years after Galilei and Bacon, supported by studies on the foundations of mathematics, references to the Aristotelian notion of truth and statistical evidence, Alfred Tarski defined the semantic notion of truth and referred to it as a mathematical model of human common sense reasoning [198, p. 341,360]. Tarski's explicit aim was not to formalise *a* notion of truth, but to formalise *the* notion of truth. Therefore, for any problem, rephrased as an Hilbertian *entscheidungsproblem* (decision problem), does its solution by Tarski's theory agree with a relevant history of natural and experimental facts? The interesting question is not what the theory cannot do [85, 37, 38, 209], but how well it models what we can do.

We are unable to solve all possible decision problems by pure reasoning, and Tarski's theory models this fact correctly, being both incomplete [202, p. 198] (there exist true sentences which are not provable) and undecidable [85, p. 607] [37, 38] (there exist sentences which cannot be proved or disproved within the theory, and adding such sentences as axioms to the theory does not solve the decision problem, because more similar sentences can be formulated in the extended theory). On the other hand, we perform common sense causal reasoning, but Tarski's theory fails to model this fact correctly; for example, it is unable to deduce the non-effects of actions, and adding axioms to the theory for all possible non-effects of actions violates Ockham's law of parsimony. This triggers the *frame problem* [134, §3.3]: given m features of objects in the environment and n actions to change their values, how do we avoid writing mn no-change axioms? The problem admits the obvious solution of linear complexity; the real challenge, however, is not to implement the brute force solution, but to model human reasoning, which avoids the no-change axioms altogether.

Tarski is the father of formal common sense reasoning, and solving the

frame problem served as a growth direction. The literature records various attempts to model causal reasoning, and in particular to solve variants of the frame problem. In the sixties, following relevant criticism on Turing's thinking machines [184], McCarthy proposed the philosophical, informal Calculus of Situations [123, 124, 134], with no solution to the frame problem. In 1972 [165] Sandewall solved the frame problem by introducing formal non-monotonic reasoning. Following Sandewall, various formalisms have been proposed as «a form of non-monotonic reasoning», including the *Closed World Assumption* inference rule and default reasoning in 1978 [159] [160, 161], predicate completion and negation as failure in 1978 [41, 42], abduction in 1979 [101, p. 241] [65, 94], McDermott-Doyle non-monotonic logic in 1980 [135], predicate circumscription in 1977–1986 [125, 126, 128], the Calculus of Events in 1986 [103], and the stable model semantics in 1988 [81]. This is the genesis of a field where each researcher delivered extensions to existing theories or at least one new theory of causal reasoning. In 1999 [171] the relevant community consisted of three-hundred researchers worldwide. See Gabbay et al. [74], Shanahan [180] and Reiter [164] for the most recent surveys.

Already in 1987, Shoham observed that «the non-standard nature of the various systems and their diversity has made it hard to gain a good understanding of them and to compare among them». He then proposed to consider any suitable compositional model-theoretic semantics as base logic, including Tarskian and modal logic, and extend it to non-monotonic reasoning using a model-preference criterion [185, 186, 187, 121]. He named the resulting logic as *preference logic*. Shoham's approach turned out to be a generalization of predicate circumscription [126, 115], being it itself a generalization of predicate completion [162]. The works of Bossu and Siegel [23], of Halpern and Moses [88], of Doyle [60, 34] and variants of Reiter's default logic [161] were subsumed under Shoham's framework. Further work led to Sandewall's filtering technique [166], also used in Shanahan's circumscriptive axiomatisations of the Calculus of Events [180, 103]. In 1993 [168] Sandewall observed that the various models available to date were the result of informal arguments, rapidly followed by refutations via counter-examples. This suggested the need for more systematic research, and a method was presented and used therein, namely the formal assessment of

the range of correct applicability of *Discrete Fluent [Preference] Logics* [169].

The *Handbook of Logic in Artificial Intelligence and Logic Programming* [174] refers to the informal arguments approach [46] as the «classical paradigm», and refers to the formal assessments approach [169] as the «systematic paradigm».

Statement of the problem

According to the history of Science, and in particular to Kuhn [105, 104], any scientific paradigm has the following three characteristics: **(i)** it is a «fundamental scientific achievement, one which includes both a theory and some exemplary applications to the results of experiment and observation»; **(ii)** it is an «open-ended achievement, one which leaves all sorts of research still to be done»; **(iii)** it is an «accepted achievement, in the sense that it is received by a group whose members no longer try to rival it or to create alternatives for it; instead, they attempt to extend and exploit it in a variety of ways». Neither the «classical paradigm» nor the «systematic paradigm» qualify as scientific paradigm.

On the «classical paradigm», some researchers contend explicitly that *the field is constrained by appeals to intuition* [20, 180]. Most researchers follow and go along with this constraint, using it as if it were the standard scientific method. Appeals to intuition, however, are everything but a systematic classification of scientific evidence. Therefore, the «classical paradigm» meets only the third of the above requirements to qualify as scientific paradigm.

On the «systematic paradigm», some researchers try to rival it [20, 180, 183] or to ignore it deliberately [47]. Although originally expected that «much more will happen in this area in the years to come» [174, p. 472], the number of formal assessments in the literature is hard evidence suggesting that the community has not received this approach. With a single exception [170], no report of research by the «systematic paradigm» has ever been recorded in a journal.

Original contribution of this work (summary)

We describe the method of research that we have endeavoured to use, as resulting from the analysis of the above problem (chapters 1-2). We then

describe and discuss the results obtained using this method (chapters 3-7).

CHAPTER 1 Although human causal reasoning is widely acknowledged as an object of scientific enquiry, there is little consensus on an appropriate measure of progress. Up-to-date evidence of the standard method of research in the field shows that this method has been rejected at the birth of modern Science.

CHAPTER 2 We describe an instance of the standard scientific method for modelling causal reasoning. The method allows for uniform proofs of three relevant computational properties: correctness of the model with respect to the intended model, full abstraction of the model (function) with respect to the equivalence of reasoning scenarios (input), and formal relations of equivalence and subsumption between models. The method extends and exploits the «systematic paradigm» [*Handbook of Logic in Artificial Intelligence and Logic Programming*, volume IV, p. 439–498, Oxford 1995] to fit with our interpretation of it. An updated summary of results is included, spanning seventy-two years of research in the area. The results show the adequacy and broader scope of the method, in a humble search for a coherent thematic view.

CHAPTER 3 *Background:* Classical Mechanics is a renown mathematical model of causal reasoning that no mathematical logic has yet succeeded to emulate. What are its epistemological and ontological characteristics? *Methods:* We used the systematic paradigm as gold standard. *Findings:* According to the present taxonomy, the relevant family of epistemological and ontological characteristics is *Ksp-RAdCi*. *Interpretation:* The result shows the distance and growth direction of classified logics with respect to the target class.

CHAPTER 4 *Background:* The Horn-clause fragment of Tarskian first-order logic, or angelic fragment of positive logic programming, is a renown theory of both human reasoning and general purpose computation. A thirty years old orthodoxy, based on «appeals to intuition», insists that positive logic programming is inadequate as a model of causal reasoning and, specifically, as a solution to the frame problem. We classified and compared this theory with two of its non-monotonic extensions, namely, Answer Set Programming (ASP) and Abductive Logic Programming (ALP). *Methods:* We used the systematic paradigm as gold standard. *Findings:* The compos-

itional model-theoretic fixpoint semantics of positive logic programming belongs to the class *Ksp-IAd*; we divided the reasoning model from its original object language, and used *K-IA*'s own language for the classification. The correctness result implies the full abstraction of this theory with respect to the equivalence of causal reasoning scenarios. *Interpretation:* The formal classification rejects the named orthodoxy. The comparisons show that part of the original problem-solving power has been lost in the non-monotonic extensions. In the case of ALP, the epistemological characteristics improved from *Ksp* to full *K*, but the ontological characteristics regressed from *IAd* to *IbsAd*. In the case of ASP, the ontological characteristics regressed from *I* to *Ibs*. This suggests a growth direction for both theories.

CHAPTER 5 *Background:* The Calculus of Events is a model of causal reasoning available in various Circumscriptive axiomatisations (CCE), all based on the Horn-clause fragment of Tarskian first-order logic. The design of this model was «constrained by appeals to intuitions», and thus its literature does not include any assessment result. We aimed to fill this lacuna. *Methods:* We used the systematic paradigm as gold standard. *Findings:* The collection reduced to five models: Boolean CCE belongs to the class *Ksp-IbA*; our redesigned Continuous CCE belongs to *Ksp-RA*; Discrete CCE is belongs to *Ksp-IA*; Abductive CCE is not correctly applicable to *K-IA* $\setminus$ *Ksp-IA*; Concurrent CCE is not correctly applicable to *Ksp-RACi*. The correctness of Boolean, Discrete and Continuous CCE depends on careful coding of the reasoning scenarios, and thus we devised a technique to synthesise and verify those scenarios. The full abstraction of each model with respect to the equivalence of reasoning scenarios holds as corollary of its correctness result. *Interpretation:* The results show a general limitation of the CCE models to the family *Ksp* of epistemological characteristics. This answers negatively to the open question whether there can be some cases of incompletely known initial state where CCE behaves correctly. Further, no CCE model proved adequate for solving problems in the full class *K-RACi*, especially in its fragment *K-RACi* $\setminus$ *Ksp-RA*. This suggests a precise growth direction, namely the integration of Concurrent and Abductive CCE into a single model. The technique to synthesise and verify the scenarios answers to an open question by Shanahan. The whole work answers to an open question by Sandewall.

CHAPTER 6 *Background:* We know that positive logic programming is correctly applicable to *Ksp-IAd* with a language shift. Its classification shows the model's ability to *simulate* the game semantics for *K-IA* when the environment has a fixed strategy. We described and assessed its non-simulative version with relevant extensions. The theory models our own approach to causal reasoning. *Methods:* We used the systematic paradigm as gold standard. *Findings:* The new model belongs to the class *K-RACi* of epistemological and ontological characteristics. The result implies the full abstraction of the model with respect to the equivalence of reasoning scenarios. *Interpretation:* The comparison of the model with former classified models, spanning seventy-two years of research in the field, it shows its strictly broader range of correct applicability. The model is correctly applicable to a superclass of Classical Mechanics. Among the subsumed models are the Answer Set Programming, Abductive Logic Programming, two variants of McCarthy's Calculus of Situations and the Calculus of Events in its various circumscriptive axiomatizations.

CHAPTER 7 *Background:* We classified the epistemological and ontological characteristics of Turing's model of a man in the process of computing a real number. *Methods:* We used the systematic paradigm as gold standard. *Findings:* Turing's automatic machines belong to the class *Ksp-IAd*, Turing's choice machines belong to the class *Ksp-IA*, Turing's computing machines belong to the class *Ksp-IbAd*. *Interpretation:* Turing's automatic machines are equivalent to Positive Logic Programming, the computing machines are subsumed by the Circumscriptive Boolean Calculus of Events, and the choice machines are equivalent to the Circumscriptive Discrete Calculus of Events.

The table at page 63 summarises the results.

We conclude the present monograph with notes and comments on theories by McCarthy, Allen, Kuipers and Shults.

1

METHODS (PART I)

1.1 Introduction

It is the paramount duty of any scientific community to foster the fundamental understanding of its subject and sustain the critical assessment of its literature. The usual fulfilment of this duty consists in a three-step process: describe the method of research, describe the results obtained using this method, discuss these results and show in what respects they advance the study of the subject. Any scientific community is defined by its peers, and their independent critical judgement is its driving force. What is most compelling is not whether independent critical judgement is allowed by authority, but whether we understand it and choose to exert it, as a free act of the individual when making a moral commitment. As Galilei has it, «in questions of science, the authority of a thousand is not worth the humble reasoning of a single individual». As Bacon has it, the absence of individual commitment leads to addiction «from prejudgement and upon the authority of others; so that it is a following and going along together, rather than consent», as «true consent is that which consists in the coincidence of free judgements, after due examination» [18, I-63,77]. The good scientist does not seek for truth in authority; they rather seek for internal coherence of theories and their agreement with natural and experimental facts. Scientific truth is neither absolute nor arbitrary, but always relative to the natural and experimental history. The standard scientific method, namely Galilei's experimental method, is so successful that any fair resolution of controversies consists in a direct appeal to available facts and to rigorous reasoning about those facts [18, II-10 111][63]. However, there seems to be much space for disagreement. An editor laments that forty-five experts reviewing exactly the same paper reported extreme judgements, from unacceptable to excellent, thus raising the problem of the reproducibility of anonymous pre-press review [207, p. 19]. After twenty-five years of ser-

vice, another editor describes peer review as «an ineffective lottery prone to bias and abuse [87]» [190, 215]. According to the British *Parliamentary Office of Science and Technology*, «some scientists believe that anonymity provides an opportunity for settling old scores and burying rival research» [4]. The persistence of editorial disorder suggests that either Baconian peer review needs to be improved, or anonymous pre-press review fails to implement it. Any peer should be able to use any relevant instance of the standard scientific method for verifying its results. In practice, however, peers gather in schools of method, and their community has no binding code of ethics whenever an author and an anonymous pre-press reader belong to rival schools [4, 2, 104]. As different schools use different methods, editorial control has the moral duty to be independent of specific schools. The general demand is to describe both the method of research and its results in sufficient detail to allow verification by any peer reviewer. The resulting scientific literature must ultimately reflect proven science. To ease the closure of scientific controversies, Frege invented the *begriffsschrift*, a method to detect and eliminate any jumping to conclusions, to proceed step-by-step by preventing anything intuitive from penetrating the Cartesian long chains of elegant thoughts, where «elegant» means «ingeniously simple and effective» in persuasion. The *begriffsschrift* is widely acknowledged as the founding work of symbolic logic.

Following the seminal work by Tarski [198] on modelling common sense truth, and by Turing [209] on modelling actions and change of a human calculator, a community of logicians aimed at modelling common sense causal reasoning, to explore farther than the computational limits of Turing machines. After much research in science departments, does the literature reflect proven science or only the opinion of individuals? What is the standard method for measuring scientific progress in the field? Our aim is to collect evidence of what this method *is*, discuss it, and consider how it might be improved.

1.2 Evidence

We expect the standard method of research in the field to emerge from Baconian consent. A selection of the collected views of individuals is arranged in a dialogue, from Frege to McCarthy. We also describe the stand-

ard structure of reports on the field, because aims and methods influence the structure of reports.

FREGE: «To prevent anything intuitive from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free of gaps. [The] demand is not to be denied: every jump must be barred from our deductions. What is so hard to satisfy must be set down to the tediousness of proceeding step-by-step.» (1879 [71, p. 5-6], 1884 [72, p. 102-3])

TURING: «All arguments which can be given are bound to be, fundamentally, appeals to intuition, and for this reason rather unsatisfactory mathematically. [The] arguments which I shall use are of three kinds: **(a)** A direct appeal to intuition. **(b)** A proof of the equivalence of two definitions (in case the new definition has a greater intuitive appeal). **(c)** Giving examples of large classes of numbers which are computable.» (1936 [209, p. 249])

TARSKI (ON LOGICAL TRUTH): «The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion. We must then characterize this notion precisely enough to enable anyone to determine whether the definition actually fulfils its task. [As] far as my own opinion is concerned, I do not have any doubts that our formulation does conform to the intuitive content of that of Aristotle. [Some] doubts have been expressed whether the semantic conception does reflect the notion of truth is its common sense and everyday usage. [In] spite of all this, I happen to believe that the semantic conception does conform to a very considerable extent with the common sense usage—although I readily admit I may be mistaken. [I] believe that the issue raised can be settled scientifically, though of course not by a deductive procedure, but with the help of the statistical questionnaire method. As a matter of fact, such research has been carried on, and some of the results have been reported at congresses and in part published [142]. [I] was by no means surprised to learn (in a discussion devoted to these problems) that in a group of people who were questioned only 15% agreed that 'true' means for them 'agreeing with reality', while 90% agreed that a sentence such as 'it is snowing' is true if, and only if, it is snowing. Thus, a great majority of these people

seemed to reject the classical conception of truth in its 'philosophical' formulation, while accepting the same conception when formulated in plain words.» (1944 [198, p. 341,360,374])

TARSKI (ON LOGICAL CONSEQUENCE): «The sentence X follows logically from the sentences of the class K if and only if every model of the class K is also a model of the sentence X.» (1936 [201, p. 417]) «I have the impression that everyone who understands the content of the above definition will admit that it captures many intuitions manifested in the everyday usage of the concept of following.» (1936 [203, p. 186]) — «Since we are concerned here with the concept of logical, i.e. *formal*, consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of objects to which the sentence *X* or the sentences of the class *K* refer.» (1936 [201, p. 414-5])

TURING: «The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any unproven conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.»[1] — «I propose to consider the question, 'Can machines think?' The [question] can be described in terms of a game which we call the 'imitation game'.» (1950 [212, p. 422, 433])

MCCARTHY: «Consider the problem of designing a machine to solve well-defined intellectual problems. We call a problem well-defined if there is a test which can be applied to a proposed solution. In case the proposed solution is a solution, the test must confirm this in a finite number of steps. If the proposed solution is not correct, we may either require that the test indicate this in a finite number of steps or else allow it to go on indefinitely.» (1956 [122])

DIJKSTRA: «When an automatic computer produces results, why do we trust them, if we do so? What measures can we take to increase our confidence that the results produced are indeed the results intended?» (1965

[1] Turing's praise of conjectures and refutations in 1950 will find support by Lakatos in 1961 [107], Kuhn in 1962 [104], Popper in 1963 [152], and a formal model by Sandewall in 1972 [165].

[55]) «We must not forget that it is *not* our business to make programs: it is our business to design classes of computations that will display a desired behaviour. [Today] a usual technique is to make a program and then to test it. But: program testing can be a very effective way to show the presence of errors, but it is hopelessly inadequate for showing their absence.» (1972 [56])[2]

SANDEWALL: «The introduction of the Unless operator is a drastic modification of the logic. It even violates the extension property of almost all logical systems, which says that if you add more axioms, everything that used to be a theorem is still a theorem.» (1972 [165, p. 199])[3]

MCCARTHY: «The intuitive idea of circumscription is as follows: We know some objects in a given class and we have some ways of generating more. We jump to the conclusion that this gives all the objects in the class. Thus we circumscribe the class to the objects we know how to generate. [...] Circumscription is not deduction in disguise, because every form of deduction has two properties that circumscription lacks—transitivity and what we may call monotonicity.» [125]

SHOHAM: «The non-standard nature of the various systems and their diversity has made it hard to gain a good understanding of them and to compare among them. We propose a unifying framework for nonmonotonic logics, which subsumes previously published systems. The basic idea behind the construction is the following. In classical logic $A \models C$ if C is true in all models of A. Since all models of $A \wedge B$ are also models of A, it follows that $A \wedge B \models C$, and hence that the logic is monotonic. One gets a nonmonotonic logic by changing the rules of the game, and focusing on only a subset of those models, those that are 'preferable' in a certain respect. In the new scheme we have that $A \models C$ if C is true in all *preferred models* of A, but $A \wedge B$ may have preferred models that are not preferred models of A. In fact, the class of preferred models of $A \wedge B$ and the class of preferred

[2]Dijkstra's argument reminds us of Locke: «it is one thing to show a man that he is in an error, and another to put him in possession of truth» [120, IV.7,§11].

[3]Sandewall's 1972 formalisation of non-monotonic reasoning for solving the frame problem had numerous followers, including Reiter's *Closed World Assumption* rule in 1978 [159], Clark's predicate completion in 1978 [41, 42], McCarthy's predicate circumscription in 1977-80 [126, 162, 115], Shoham's general notion of model preference in 1986 [185, 186, 187, 121], Kowalski and Sergot's calculus of events in 1986 [103], and Shanahan's circumscriptive calculus of events in 1995-7 [179, 180].

models of A may be completely disjoint.» (1987 [186])

HANKS AND MCDERMOTT: «Nonmonotonic formal systems have been proposed as an extension to classical first-order logic that will capture the process of human 'default reasoning' or 'plausible inference' through their inference mechanisms, just as *modus ponens* provides a model for deductive reasoning. But although the technical properties of these logics have been studied in detail and many examples of human default reasoning have been identified, for the most part these logics have not actually been applied to practical problems to see whether they produce the expected results. We provide axioms for a simple problem in temporal reasoning which has long been identified as a case of default reasoning, thus presumably amenable to representation in nonmonotonic logic. Upon examining the resulting nonmonotonic theories, however, we find that the inferences permitted by the logics are not those we had intended when we wrote the axioms, and in fact are much weaker. This problem is shown to be independent of the logic used; nor does it depend on any particular temporal representation. Upon analyzing the failure we find that the nonmonotonic logics we considered are inherently incapable of representing this kind of default reasoning.» (1986 [89]) «It almost certainly is the case that nonmonotonic inference is idiosyncratic and domain dependent.» «The relationship between these logics and human reasoning is not well understood.» «We can no longer engage in the logical 'wishful thinking' that led us to claim that circumscription solves the frame problem [127].» (1987 [90])[4]

DAVIS: «The basic approach used here, as in much of the research in automating commonsense reasoning, is to take a number of examples of commonsense inference in a commonsense domain, generally deductive inference; identify the general domain knowledge and the particular problem specification used in the inference; develop a formal language in which this knowledge can be expressed; and define the primitives of the language as carefully and precisely as possible. Only occasionally is there any dis-

[4]From 1972 to 1986, the correctness of non-monotonic solutions to the frame problem has been conjectured using tests and appeals to intuition. In 1986 Hanks and McDermott refuted the method using a counter-example. The community awarded the paper, but persisted with the old method; the counter-example has been adopted as a positive test, known as Yale Shooting Problem, and more similar tests have been adopted as benchmarks [113]. A survey of this development in the decade 1987–1997 is available from Shanahan [180].

cussion here of the algorithms, the control structures, or the organization of data that would be used in an actual reasoning system. There is almost no discussion of domain-independent techniques.» (1990 [46, p. 4])

BROOKS: «Traditional Artificial Intelligence has tried to tackle the problem of building artificially intelligent systems from the top down. It tackled intelligence through the notions of *thought* and *reason*. These are things we only know about through introspection. [Recently] there has been a movement to study intelligence from the bottom up[.] Some of this work is based on engineering from first principles[.]» «The idea is to first build a very simple complete autonomous system, and test it in the real world.» (1991 [32, p. 88,134-5])

SANDEWALL: «There has been much research in recent years on methods for reasoning about actions and change, and on finding solutions to the so-called frame problems. New variants of non-monotonic logics for common sense reasoning have been proposed, only to be quickly refuted by counter-examples. According to the standard research method in the area, the evidence in favour of a proposed logic should consist of intuitive plausibility arguments and a small number of scenario examples for which the logic is proven (or claimed) to give the intended conclusions and no others. Clearly there is a need for more systematic results, where a proposed logic is verified for a whole class of reasoning problems and not only for single examples.» (1993 [168]) — «There are sometimes 'clashes of intuitions' where different informers make incompatible statements about what are the admissible common sense inferences for a given example. This means that the empirical data for the research are unreliable. Also, since the informer is usually the researcher himself, there is always a danger that the researcher will be influenced by the particular theory he has developed [208]. A major reason for the uncertainty is that it is difficult to delimit the 'general domain knowledge' that has been used for a particular inference.» «In practice, the systematic methodology and the example-based methodology have to be used together. The argument for a systematic methodology is not intended to imply that common-sense examples are useless, but only that they are not sufficient as a basis for developing a logic of common-sense reasoning.» (1994 [169, p. 64,69])[5]

[5]Sandewall moves along points **(b)** and **(c)** of Turing's method, criticizing but still ac-

GABBAY: «We still have conceptual problems with the subject. There are lots of nonmonotonic mechanisms. It is not clear how they fit into a coherent thematic view. I believe, however, that it is possible to present a good framework. It must be possible. After all, these logics are supposed to analyse human practical reasoning. We humans are more or less coherent so there is something there! [We] are developing the necessary tools for modelling human reasoning. However, [we] should be careful not to give too much emphasis to the tools. [Developing] and comparing formal nonmonotonic systems should serve the study and analysis of human practical reasoning.» (1994 [75, p. v])

SANDEWALL AND SHOHAM: «The **classical paradigm** or *paradigm of common-sense examples* [has] been concisely summarised by Davis[.]» «The methodology of common sense examples has resulted in a somewhat chaotic development; logics, examples, and counter-examples, have been confronted, and it has not always been clear which property of a logic was to be given the credit for its success, or the blame for its failures. It has not even always been clear what was a success or a failure in terms of the proposed reasoning examples. The most important weakness, however, is that various logics have been proposed to be 'solutions to the frame problem' just on the basis of intuitions and a few examples, only to be very quickly refuted by the arrival of more examples. The **systematic paradigm** [169] attempts to structure the problem area a bit better. [We] expect that much more will happen in this area in the years to come.» (1995 [174, p. 441,468,492-3])

BELL: «An experimental justification of a formal theory of common sense causal reasoning, consists of intuitive plausibility arguments for the theory in question, together with a demonstration of how it succeeds in representing scenario examples. [The] experimental approach is, after all, just standard scientific method. [By] contrast, formalist justifications, for example [118, 114, 95, 96], establish relationships between formal theories of common sense causal reasoning. Sandewall criticizes justifications of this kind because the results they provide are entirely 'within the framework of logic'; the results relate pairs of theories but do not attempt to

cepting point (a). In moving along point **(c)**, he implicitly agrees with Dijkstra on classes of computation.

square either of them with our intuitions about common sense causal reasoning. He then proposes his own formal game-theoretic trajectory semantics and proves a series of theorems relating his logic to them. However, as he does not attempt to justify his trajectory semantics by appealing to our intuitions about causal reasoning, he is, in effect, repeating the formalist move. Any justification [must] (ultimately) relate the intuitive notion and the formal theory. [...] Our intuitions about common sense causal reasoning are even less clear than those about effective computability.» (1995 [20, §5])

SANDEWALL: «[I] have [noticed] that it is apparently much easier to get articles published if they use situation calculus. This may possibly be due to notational chauvinism [on] the side of some reviewers: If one really believes that (e.g.) the situation calculus is the best answer to all problems, then why accept a paper from someone that hasn't seen the truth? — If our research area is going to conserve an older approach to such an extent that essential new results can't make it through the publication channels, then the whole area will suffer.» (1997 [173])

SHANAHAN: «The style of work I have adopted emphasizes the individual example. It can be argued that an approach to the frame problem can only properly be assessed by establishing formal correspondence with a more abstract formal framework that encompasses a large class of examples. According to this argument, appeals to intuition through single example scenarios are suspect. Without denying the importance of establishing correspondences between different formal frameworks, I would contend that the field is constrained by appeals to intuition until its formalisms are deployed in the design of working systems. After all, any abstract framework that is used to assess an approach to the frame problem is itself open to assessment. How do we know that the abstract framework itself is correct? Only by appealing to our intuitions in an examination of its performance on a judiciously chosen set of representative examples.» (1997 [180, p. xv,xvi])

DIJKSTRA: «[The] notion of a 'convincing argument' immediately raises the question 'convincing to whom?': if the audience to be convinced is sufficiently gullible, the argument can be gloriously defective! It became the task of the mathematical community to cultivate the scepticism against

which the quality of the 'convincing argument' would be checked. [The] ultimate consequence of the adoption of the postulational method is the inadequacy and subsequently the irrelevance of the consensus model[,] as the latter reflected a form of interaction with the community which was needed to compensate for the major shortcomings of informal intuitive reasoning. [Regrettably] this evolution seems to be ignored [and] a sort of feel-good mathematics is promoted in which rigorous arguments don't seem to play a role[.] This is a tragedy.» (1998 [57])

SANDEWALL: «Shanahan adopts, and argues for the example-based methodology, where approaches to formalizing actions are explained and motivated through a small number of scenario examples, and where an existing approach can also be refuted by showing an example where it does not provide the intended conclusions. Some of the research in this area uses another, *systematic* approach, where one attempts to characterize the properties of various known approaches. The example-based and systematic methodologies need not be mutually exclusive. It is perfectly possible, and in fact advisable, to use them together. [It] is a pity that Shanahan has not taken the opportunity to include any assessment results in his book, even informally. [Shanahan's] omission in this respect is even more surprising in view of the book's subtitle: 'A Mathematical Investigation of the Common Sense Law of Inertia'. Assessment results have a reasonable generality and require a certain degree of mathematical investigation. The plentiful formal propositions in his book, by contrast, are only statements about one particular toy example at a time. [There] is one single 'abstract framework' that has been used extensively for [the assessments], namely state transition systems and various generalizations of them. In particular, state transitions systems were introduced in the early nineties for defining the underlying semantics in both the Features and Fluents approach (Sandewall) and in action languages such as $\mathcal{A}$ (Gelfond and Lifschitz). They are also a widely used framework in other branches of engineering and in computer science, and they are much more transparent than the modern, relatively complex logics for the frame problem. Therefore, it is not a circular exercise to assess logics for actions and change relative to a transition-system framework. On the contrary, doing so provides a better understanding of when the logic works correctly and when it doesn't, and

it helps relate research in this area to neighboring disciplines. Shanahan's strict adherence to only one of the two methodologies is therefore unfortunate. These critical comments illustrate that, not surprisingly, the field of actions and change is characterized by a multiplicity of approaches and less than total agreement about methodology.» (2000 [172])

DIJKSTRA: «[Most] books are recommended for being intuitive instead of formal, for being chatty instead of crisp, for being vague and sloppy instead of rigorous. [This] disrespectful, almost contemptuous attitude has not only affected teaching and publishing, it has affected the academic research agenda as well[.]» (2000 [58]) «The *Concise Oxford Dictionary* gives as one of the meanings of 'elegant': 'ingeniously simple and effective'. Why has elegance found so little following? Elegance has the disadvantage, if that's what it is, that hard work is needed to achieve it, and a good education to appreciate it.» (2001 [59])

REITER: «What is available is more like a Tower of Babel than a unifying representational and computational formalism. To be fair, this state of affairs is the natural outcome of disciplines organized by their applications. We all solve problems that arise in our own, sometimes narrowly circumscribed fields of specialization[.]» (2001 [164, p. 1-2])

MCCARTHY: «The engineering approaches to [Artificial Intelligence] regard the world as presenting certain kinds of problems to an agent trying to survive and achieve goals. It studies directly how to achieve goals. The **logical approach** [to Artificial Intelligence] is a variety of the engineering approach. A logical agent represents what it knows in logical formulas and infers that certain actions or strategies are appropriate to achieve its goals. [...] The logical approach also has the advantage that when we achieve human-level AI we will understand how intelligence works. Some of the evolutionary approaches might achieve an intelligent machine without anyone understanding how it works. [...] Human-level intelligence is a difficult scientific problem and probably needs some new ideas. These are more likely to be invented by a person of genius than as part of a Government or industry project.» (2003 [132, p. 76ff])

The above dialogue on method describes the views of individuals. To see which of the views has been received by the community, we shall now

describe the structure of reports on the field. Raw evidence of it is plentifully available.

For any given «leading formalism in the field», a research report consists in evidence of its failure through a test, or benchmark problem, the description of an ad-hoc extension, and appeals to intuition to support the claim that the extended formalism solves the problem. The report is divided in essentially four sections. The first section describes the problem in plain English and discusses relevant known information. The second section describes the principles of adequacy for the proposed solution. The third section describes the formalism and uses its language to represent the given problem. The fourth section literally claims «the formalisation demonstrates that sophisticated kinds of common sense reasoning [can] be captured by the reasoning model», and concludes with «a few words of comparison» with selected works.

In general, aims and methods of research influence the structure of reports, and reduce them to three broad categories: theoretical, experimental, and developmental [91]. If one reads the «appeals to intuition» and the «few words of comparison» as fulfilment of the given principles of adequacy, then the standard structure of research reports on the field matches the structure of developmental (engineering) papers.

We find it appropriate to close the above dialogue by quoting from pre-press review, where the reviewers share without inhibitions, believing their text to be anonymous and off the record. The following is a passage from a paradigmatic pre-press review in our own record. Editorial control implies editorial accountability [4].

C. Bettini and A. Montanari (ed.): «The paper suffers from the lack of examples, without which the subtleties of the formal assessment are difficult to understand intuitively. I'd suggest to skip the proof.» (Elsevier Science, 2002)

The synthesis of diverse views outlines three schools of method. According to a first school, appeals to intuition must be barred from the literature; tests can only show the presence of errors, when refuting conjectures. According to a second school, progress is measured by informal assessments of correctness with respect to individual tests, moving along

point **(a)** of Turing's method, criticising and rejecting points **(b,c)**; this is «just standard scientific method» by some, and «a variety of the engineering approach» by others. According to a third school, progress is measured by formal assessments of correctness with respect to classes of tests, moving along points **(b,c)** of Turing's method, criticizing but still accepting point **(a)**. Although there are three rival schools of method, the structure of reports follows the second school. Authority, namely pre-press review, it is biased, contending through rejection that the field is constrained by appeals to intuition.

1.3 Discussion

The evidence shows the absence of Baconian consent on method. In our discussion, we shall separate the people from the problems, and thus we shall not argue about individual positions. We shall rather focus on individual interests, to identify opportunities for a resolution of the controversy, and insist on using objective criteria, to avoid arbitrary judgement. We resolved to use a single objective criterion, namely the consistency of individual interests with the standard scientific method [66, 18]. The reason for this is plain: any local paradigm of scientific research must serve the general interests of science.

1.3.1 On the first school

The first school of method knows the limitations of theoretical truth [92, 85, 86, 43]. Both proof theory and model theory are affected, and subsequent research reports no objective progress on this front, showing that the community is still affected by the Gödelitis. Indeed, there is a gap on the fundamental adequacy of a strictly theoretical method in conducting scientific research. The community insists in its purely theoretical method while ignoring its original, scientific aim of modelling human reasoning. By word of Tarski himself, the very concept of logical consequence is not influenced by empirical knowledge, and doubts have been expressed whether the semantic conception of truth reflects the Aristotelian notion of truth. Although we have no record of such doubts, and thus are unable to verify them, we have doubts of our own, expressed as follows.

The Aristotelian notion of truth conforms to common sense, it is the

core method of experimental enquiry [8, III.10][79, I], and Tarski spent twenty-five years to formalise it within mathematical logic, leading to the model-theoretic definitions of logical truth and logical consequence in 1953 [204]. In 1944, Tarski described his aim in plain English (we quoted it at p. 18). To pursue his aim, Tarski used the Cartesian method [52, regula XII] [54, part II] [198, p. 352-3]. The resulting work benefits from it, but also inherits its weakness. The Cartesian method is similar, although not identical, to Caesar's art of war *divide et impera*. Descartes retained Caesar's approach to structure a complex problem into its simpler components, but replaced the physical solution to the simplest problem with a sentence that he judged true beyond doubt: «cogito ergo sum» [54, p. 18-9 32–34]. The observation that «our senses sometimes deceive us» misled Descartes away from the experimental method. Descartes himself demonstrated the weakness of his method when claiming: «it is evident that vacuum does not exist» [53, XVI]. Despite the inadequacy of the Cartesian method for experimental research, the method is still in the highest regards for theoretical research. Tarski used the Cartesian method, both to define the notion of a sentential function in formalised languages and to obtain a definition of satisfiability, where he replaced the experimental facts with context free meta-level objects, and most notably, he did so as a proposed solution to the liar paradox. Although the Tarskian notion of truth conforms to the Cartesian notion, Tarski did not display awareness of it; on the contrary, he believed explicitly that it conforms to that of Aristotle (p. 18). To qualify in that sense, however, the meta-level objects should rather be experimental facts, like objects of the physical environment, or properties of objects, or relations between objects, as resulting from perception and reasoning, but Tarski did not develop a model of critical perception. Properties and relations can change in time, and thus critical perception requires performing actions that lead to context dependent experimental facts, but Tarski did not develop a model of causation. Finally, modelling common sense (causal) reasoning involves modelling a context dependent notion of truth (relative truth), but Tarski modelled a notion of truth for context free sentences (absolute truth).

A comprehensive programme for a relevant extension of classical, Tarskian logic, aiming at Tarski's intended goal of modelling the Aris-

totelian notion of truth, should then include the following: the threefold extension of classical logic to include a materially adequate model of perception, causation, and context dependent truth. This identifies the research problem with more precision than the frame problem [116], and states the inadequacy of purely theoretical methods of research in this field. *We remark the difference between the above programme and McCarthy's aim to merely use classical logic to axiomatize knowledge of actions and their effects. It is clear to us that classical Tarskian logic is fundamentally inadequate to model human causal reasoning, due to its Cartesian notion of truth.*

1.3.2 On the second school

The *Oxford Dictionary* defines «intuition» as «the immediate apprehension of an object by the mind without the intervention of any reasoning process» [189]. The «classical paradigm» is «constrained by appeals to intuition», and thus it demands for «immediate apprehension» of a formal model of human reasoning «by the mind», through a written test, but «without the intervention of any reasoning process». This leads to the absurd conclusion that to peer review a formal model of human reasoning we would have to suspend critical judgment.

To avoid the absurd, the «classical paradigm» should have been described in terms of a more appropriate word than «intuition», to describe correctly the meaning acquired by habit, such as: «intuitive plausibility arguments» without «formalist justifications». The following key words match semantically [189]: **(i)** *assumption*: «a belief or feeling that something is true, although there is no proof of it»; **(ii)** *conjecture*: «a ground or reason for conclusion (not amounting to demonstration)»; **(iii)** *opinion*: «judgement resting on grounds insufficient for complete demonstration», «distinguished from knowledge».

The «classical paradigm» turns out to be the exact converse of the standard scientific method, because Science is «clear and certain mental apprehension», «knowledge as opposed to belief or opinion» [189], and it does not consist of assumptions, as also stated by Ockham's Law of Parsimony. Key references supporting this conclusion are Galilei [77, 78] [79, I][77, 78] [79, I] and Bacon [18, I-65-6] at the birth of modern science, but also Plato [149, 98a] and Aristotle [8, III.10] [10, VI.10] [9, I,2,71 b ff.] [12, I,100 a 27]

[13, I,33,89a 38;I,31,87b 27;I,2] [11, VI,2,1027 a 20;VII,6,1031 b 5]. Bacon's *Novum Organum* summarises and still popularizes both the standard scientific method and the principles of peer-review, standing as the manifesto of the Royal Society of London and any academy of science since 1660 [191]. The «classical paradigm» moves away from the very principles that have been fostering modern science in the past four centuries, and it does so with no explicit and well-argued reason. By rejecting proofs in favour unsupported claims, it is also cunningly eluding the proper role of conjectures, as explained by Turing himself.

The rejection of the method raises doubts on the literature obtained using this method. The «classical paradigm» comes with the following question, usually at pre-press review: «As we know well, there are several leading formalisms in the field (e.g. the situation calculus [...]), so what is the substantial reason for introducing another [formalism] instead of working with the existing ones?» This classical question has its classical answer from Galilei, Bacon, and four hundred years of European academy. The «substantial reason for introducing another formalism instead of working with the existing» «leading formalisms» is that writing scientific models is what scientists do. We are free from the orthodoxy that *you shall have no other formalisms before the leading formalism*, because the paramount duty of Science is to model Nature, not to acknowledge authority. The adjective «leading», when referred to the «classical paradigm», it is an adjective unsupported by evidence: it is indeed the case that the «classical paradigm» does not use any objective measure of progress for the systematic analysis and comparison of relevant formalisms. Its «leading formalisms» are *not* the measure of scientific progress, via fair classification and comparison, but the measure of the ability of their school to bury rival research with dogmatic and abusive pre-press review. The anonymous pre-press reviewers abuse their role by conditioning authors to use a non-scientific method when conducting scientific research. The use of editorial control to gain competitive advantage, preventing the work of rival schools to appear on the record, it is less than fair play. Under the *publish or perish* dogma, the act of conditioning by rejection is an intellectual crime.

It certainly is an attractive quality to show respect for the past, with all its embarrassing mistakes, but we must beware of its devious influence. Em-

barrassment occurs when people's opinions are mixed with their egos, and thus independent critical judgement triggers defence mechanisms such as authority and prejudgement. Proper scientific work is objective, dispassionate, not bound to anyone's intuition, or ego, or authority. As Cicero has it, «anyone is liable to make mistakes; but no one persists in making them, except the unwise person» [40, Philippica XII]. The «classical paradigm» shows its respect for the past by compulsively persisting in the embarrassing mistakes, forcing the community to commit more of them, through dogmatic and abusive pre-press review.

1.3.3 On the third school

According to the «systematic paradigm», scientific progress in the field is measured by formal assessments of correctness with respect to a state transition semantics. The father of the «systematic paradigm» acknowledges the engineers Ljung and Nebel for the research problem and inspiration for the research method respectively [169, p. vi], but leaves a gap in knowledge on the fundamental adequacy of an engineering method in conducting scientific research.

From our standpoint, Science and Engineering have different aims and methods, and are therefore distinct. We read the «systematic paradigm» as a Linnean method of binomial nomenclature for the classification of models based on their epistemological and ontological characteristics. These characteristics are stated as constraints on a state transition system which mimics the common sense causal inferences *step-by-step*. The criterion of classification consists in proving soundness and completeness of the model (its observational behaviour) with respect to the intended model (the state transition semantics). Based on our reading, we use this method to arrange the collection of raw data in tables, which fulfills the second point of the standard scientific method, as described by Bacon.

1.3.4 Pathologies

In modelling human behaviour, it is important for us to be aware of pathologies, to see non-trivial errors in past research and the directions to avoid them in our own models. The field of Psychiatry discriminates between two disorders that are expecially compelling in our research: neurosys and

psychosis, or perceptual bias and inferential bias [1].

Any agent by the Cartesian method is psychotic. The agent reasons and acts upon its internal representations only, displaying lack of awareness of the external environment. We identify as examples of psychotic agents, Tarski's semantic theory and Nilsson's SHAKEY [144]. McCarthy's circumscriptive calculus of situations has the «jumping-to-conclusions» inferential bias, which is a symptom of clinical paranoia [22].

Any agent by Brooks's method [31, 32, 33] is neurotic. Given the agent's lack of internal representations, and thus of memory, the agent would repeat certain actions over and over, displaying the symptoms of obsessive-compulsive disorder [67].

Both methods lead to incomplete models of human behaviour, all requiring human assistance. By comparison, any good agent by Aristotle's method would model both perception and reasoning, to use them properly, thus requiring less human assistance in Cognitive Robotics.

1.3.5 Conclusion

We collected evidence of the standard method of research in the subject, and conclude that the community has not yet reach Baconian consensus. There are indeed three schools of method. Methodological limitations of the first school include inferential bias through the failure to model perception. Methodological limitations of the second school include inferential and perceptual bias, and the more general failure to measure progress objectively. The third school, in our interpretation, fulfills the second point of the standard scientific method.

2

METHODS (PART II)

2.1 Introduction

We are interested in modeling human causal reasoning. Our motive is plain curiosity, with no specific application in mind. Our aim is to foster the fundamental understanding of the subject via the systematic arrangement of raw data, through analysis of its structure. To pursue this aim, we need a suitable instance of the standard scientific method for the collection, classification and comparison of data. We outline this method as follows.

COLLECTION OF RAW DATA Language is the medium of thought; it makes human reasoning observable. The raw data, therefore, is plentifully available. The relevant history, both natural and experimental, consists of explicit reasoning and observation of salient behaviour. Test examples are *not* suitable as raw data, because they describe problems, not the reasoning behind their solution. Existing theories are suitable as raw data, because they are, albeit in part, an explicit and concise description of the reasoning of their respective authors. We shall use corpus-based grammars of natural language [157] to fasten the collection to experimental evidence of relevant linguistic structures, such as the use of time points and continuous change in progressive tenses.

CLASSIFICATION OF COLLECTED DATA We read Sandewall's «systematic paradigm» [174, 169] as a suitable method of binomial nomenclature based on epistemological and ontological characteristics of collected data. The binomial nomenclature reminds us of Linnaeus, and the denotation of characteristics and subcharacteristics reminds us of Berzelius (with subscripts instead of superscripts). These characteristics are stated as constraints on a state transition system which mimics the causal inferences step-by-step. The criterion of classification consists in proving soundness and completeness of each theory (its observational behaviour) with respect to the state transition semantics (the intended models). Our intuitions are *recollected*

by reading the classification, formally proved by structural induction on the length of the causal chains. We revise and extend this classification method to solve a number of open problems. The resulting method is independent of specific models. To describe time points and continuous change in progressive tenses, the state transition system is redefined in terms of lattices and continuous fluents. The method encompasses models of perception (function of reality-testing), causation (including continuous change, concurrency of actions, cause-and-effect chains, delayed effects of actions, causal qualification and structure-based ramification), and context-dependent truth. The given notion of perception renders the local notion of truth *fair to facts*.

COMPARISON OF CLASSIFIED DATA We compare the classified data by assessing equivalence and subsumption relations, formally proven by set-theoretic inclusions. If $\mathcal{M}$ is the set of classified data, the structure $(\mathcal{M}; \subseteq)$ is a partially ordered set. We adopt this structure as the coherent thematic view of which we were in search.

In what follows, we describe the above method in detail, namely, the general method for classifying and comparing data (§2.2.1) and specific epistemological and ontological characteristics for the classifications (§2.2.2). Our design is *not* to teach the method which each should follow, but only to describe the one we have endeavoured to use. In the last part of this report, we present and discuss an updated summary of the results obtained using this method.

2.2 Method

Note on the register. The word «model» has different meanings in different contexts and fields of research. The Oxford English Dictionary [189], for example, records fifteen core meanings and about twenty-five submeanings of this noun. In general, scientific models are theories which aim at explaining what Nature does or may be made to [18, aph. I-10]; they aim at describing Nature and are the results of scientific discovery. Engineering models are descriptions of objects made by humans, that is, objects that do not exist in Nature and are the result of human invention. Classical, Tarskian, mathematical logic is a scientific model of human reasoning. For

any set Γ of well-formed formulae of classical logic, a «model of Γ» is a certain mathematical structure S such that all formulae in Γ are true in S, as defined by Tarski in 1953 [204] [198, p. 352-3]. «Model theory» is the branch of meta-mathematics announced by Tarski in 1954 [199]. Preferential logics [186, p. 389] are the result of associating a preference relation on interpretations to *any* base logic with compositional model-theoretic semantics. Classical and modal logic are an example of such base logics. In Sandewall's monograph [169, p. 185,176-177,132-134], and in the following def. 2.2.1, we read «classical models» as «models of a compositional model-theoretic semantics» rather than «models of classical logic», because the specific work is more general than classical logic and uses base logics other than classical logic. In the same references, «intended models» are formal models by the transition-system describing the epistemological and ontological characteristics of discrete-fluent preferential logics. We shall use the same noun to mean either scientific theories or semantic models, and let the reader disambiguate from the context.

2.2.1 How to classify and compare models

Definition 2.2.1 (Correctness) «If Γ is a set of propositions, we write $[\![\Gamma]\!]$ for the set of classical models for Γ and $\Sigma(\Gamma)$ for the set of intended models. The research problem is how to obtain $\Sigma(\Gamma)$ in terms of operations on formulae in Γ. [Rather] than allowing the definition of $\Sigma(\Gamma)$ to rely on intuitions that are independently applied for each particular choice of Γ, [the method] requires a formal definition of $\Sigma(\Gamma)$ as the set of intended models. The task for any proposed nonmonotonic logic is then to correctly identify this $\Sigma(\Gamma)$. In fact one needs not just one but several such functions Σ_i. [Sandewall] [168] defines Σ using an *underlying semantics* based on a notion of possible trajectories that characterize the various ways of performing each action. This may be understood as a way of having an auxiliary language with only its semantics but without any syntax. [It] is important that the definition of the function Σ shall [capture] our notions of common sense. [Once] one or more Σ_i have been defined, one can analyse some of the nonmonotonic logics that have been described above, or some new ones. [Each logic is] characterized by its own model-selection function S. [Each] logic must then be formulated as a function S from a set

of axioms to a set of models, and one can ask the following question: For a given logic S and a given Σ_i, does the logic obtain exactly the intended conclusions, i.e. is $S(\Gamma) = \Sigma_i(\Gamma)$ for all Γ? [In] order to characterize the range of applicability of each S it is natural to first define a common and fairly broad base logic, i.e. a logic in the conventional sense of a syntax and a Tarskian semantics.» [174, p. 445,468-470]

The above criterion of correctness admits the existence of an object language and a meta-language. The object language is the preferential logic. The meta-language is a formalized language; its syntax is identical to the syntax of the object language, its semantics is the *underlying semantics*, and its models are defined using the same structure of interpretations of the object language. The equivalence of preferred models and intended models is the correctness criterion. The evidence of this intimate relation between the two languages is available in Sandewall's monograph, where a base logic is used to define both S and Σ [169, p. 201,239].

This intimate relation between object language and meta-language is essential to Sandewall's notion of correctness. The assessment of correctness standardizes the object language; it must use the same syntax and structure of interpretations of the meta-language. Factual evidence of this is available in Sandewall's monograph, where the PGM entailment [169, p. 243] is a reformulation of a model by McCarthy [127, 128], an instance of the GMON entailment [169, p. 242] [174, p. 485,486] is a reformulation of a model by McCarthy [127, 128], and the OCM entailment [169, p. 243] is a reformulation of a model by Kautz [97]. The base logic for both OCM and PGM is the discrete-fluent logic DFL-1. The base logic for GMON is the discrete-fluent logic DFL-2. The reformulation is also instrumental to examine any question of interpretation that may occur during the assessment of an imprecise language; the cited work by McCarthy is evidence of this [169, p. 242-3].

The same intimate relation between object language and meta-language raises serious difficulties. In this field, any model embodies the history and identity of its community, which would rather reject the assessment method than reformulate the model. Further, the assessment of correctness does not necessarily succeed immediately, because the assessment aims at identifying Σ_i, reformulating the model at each attempt. The reformulation may also change the model, like an inaccurate diagnostic instrument whose

usage changes the very properties it aims at diagnosing.

We solved the above problems by lifting the need for reformulations. We introduced and used this solution to assess Logic Programming, the Circumscriptive Calculus of Events [28] and the Calculus of Fluents [26]. We describe the solution in the following definition. We perfect the method introducing the term «classification» of models, which occurs neither in Sandewall's monograph [169] nor in the *Handbook* [174].

Definition 2.2.2 (Classification) Let Γ be any set of sentences in the meta-language (reasoning problem). Let S be any causal logic (reasoning model), defined as a function from a set of sentences to a set of semantic models. We refer to the language of S as the object language. Let T_1 and T_2 be functions whose respective purpose is to translate the syntax of the meta-language into the syntax of the object language, and the structure of interpretations of the meta-language into the structure of interpretations of the object language. We say that S is *correct* for Γ if and only if $S(T_1(\Gamma)) = T_2(\Sigma_i(\Gamma))$, that is, for any intended interpretation I, $T_2(I) \in S(T_1(\Gamma))$ if and only if $I \in \Sigma_i(\Gamma)$. The correctness of the causal logic is expressed as equivalence between the logic and the underlying semantics. The equivalence is formally proved by structural induction on the length of the causal chains, comparing the observable behaviour of causal steps in $S(T_1(\Gamma))$ with those in $T_2(\Sigma_i(\Gamma))$. The correctness criterion reduces to $S(\Gamma) = \Sigma_i(\Gamma)$ if T_1 and T_2 are the identity function. The *range of applicability* of S, or problem-solving power, is the class of all Γ such that S is correct for Γ. The *classification* of S is the formal assessment (proof) of the range of applicability of S.

The above criterion of correctness is more demanding than Milner's *bisimulation* [175]. The standard bisimulations compare the observable behaviour of a computational step in one system with that of another, essentially checking the labels of transitions. In our theory, the observable behaviour must correspond at any given point in time, including the times between transitions (inertia) and during transitions (discrete or continuous change).

The classification of a model is its *certificate of correct applicability*. When applied to any reasoning problem in its class, the model always gives the

correct (intended) set of conclusions. There are uncountably many reasoning problems in each non-trivial class, and thus the assessment holds true with respect to the general characteristic properties defining this class. In practice, we use the above definition as follows: given a reasoning model and a reasoning problem, we say that the model solves the problem *only if* the problem belongs to the class for which the model is provably correct.

The cost of introducing a new model is the number of fair and accurate comparisons with all pre-existent models. The following definition reduces this cost to the single classification of the new model, with the added value of a result whose validity holds up to future additions into the set of classified models. Using the same definition, the task of assessing equivalence and subsumption relations among models reduces to deciding set-inclusions between their respective range of applicability, and thus, once classified, detailed knowledge of the models is not a requirement for their formal comparison.

Definition 2.2.3 (Equivalence and Subsumption) Let S_i and S_j be classified models of causal reasoning. We define a binary relation, $S_i \subseteq S_j$, which satisfies the property of reflexivity, antisymmetry and transitivity. If S_i and S_j have the same class, we write $S_i = S_j$ and say that S_i and S_j are equivalent. If $S_i \subseteq S_j$ and $S_i \neq S_j$, we write $S_i \subset S_j$ and say that S_j subsumes S_i. Let $\mathcal{M}$ be the set of all classified models. The structure $(M; \subseteq)$ is a partially ordered set.

Our ability to offer equivalent answers to equivalent questions in equivalent contexts is supporting evidence of coherence. We describe this ability as the property of *full abstraction* of the reasoning model (function S) with respect to the equivalence of reasoning problems (input). The term was inspired to us by Levi [76] and Milner [136].

Definition 2.2.4 (Full abstraction) Let Γ and Γ' be sets of sentences in the meta-language. Let T be a translator. Let $=$ be the usual set-theoretic identity. We define the following relations.

- Equivalence of Γ and Γ' with respect to the set of intended models:
  $\Gamma \approx \Gamma' \Leftrightarrow \Sigma(\Gamma) = \Sigma(\Gamma')$

- Correctness of S with respect to $\approx$:
  $\Gamma \approx \Gamma' \Leftarrow S(T(\Gamma)) = S(T(\Gamma'))$
- Full abstraction of S with respect to $\approx$:
  $\Gamma \approx \Gamma' \Leftrightarrow S(T(\Gamma)) = S(T(\Gamma'))$

Proposition 2.1 For any given model S and intended model Σ, the correctness of S with respect to Σ implies the full abstraction of S with respect to $\approx$.

Proof The proposition holds true by mere transitivity of $=$. q.e.d.

As observed by Milner, proving full abstraction of formal models is non-trivial. In the above case, by proposition 2.1, this proof is a corollary of classification. Proving full abstraction for non-classified models falls back to Milner's observations.

The wide variety of existing models has made it hard to gain a good understanding of them, to compare among them, and fit them into a coherent thematic view [186] [75, p. v]. We solve this open problem by deemphasising the importance of individual models. We proceed as follows: (step 1) we collect models as raw data, clearly stated and free of appeals to intuition; (step 2) we arrange the collection in tables, according to shared epistemological and ontological characteristics (def. 2.2.2); (step 3) we compare the classified data with each other by assessing equivalence and subsumption relations (def. 2.2.3). This describes an instance of the standard scientific method in its usual reading of data collection, data analysis and data interpretation, as concisely described by Bacon [18, aph. II-10], and it aims at meeting Bacon's and Frege's advice to proceed free of gaps in the production and verification of results [18, aph. I-65, I-66] [72, p. 102,103] [71, p. 5,6].

The standard scientific method of research suggests a uniform structure for its reports, namely, four standard sections with a clear understanding of their individual contribution and mutual relation.

- The first section describes the research **aims** in context. Examples of application do *not* qualify as scientific aims.
- The second section describes the **method** of research, being a relevant instance of the standard scientific method. If the method is well described in a published paper or standard text then a reference to the source will

be sufficient. Any new extension used in the report must be described explicitly. The criterion for a well-written section on method is that a reasonably knowledgeable colleague could reproduce the results after reading the description.

- The third section presents the **results**, as obtained using the above method. In our field, the results consist of classifications (def. 2.2.2).
- The fourth section is the **discussion**, namely, the interpretation of the results with respect to the original aims, and the comparison between these results and former relevant work (def. 2.2.3).

We separate the results from the discussion because it helps preserving the objectivity of the results. Cautious «appeals to intuition» are allowed in the discussion. Different people may have different interpretations of the results, but the results themselves must appeal to facts and to rigorous reasoning about those facts. Given this division, a «test example» may only be described in the discussion, although proper examples of application are best presented in engineering papers.

Throughout the above description, we used the formal notion of intended model Σ_i. We shall now describe this notion, followed by three instances as used in various reports when classifying and comparing models. This concludes the description of the method.

2.2.2 Epistemological and ontological characteristics

In the above definitions, Σ_i describes the epistemological and ontological characteristics of a class. These characteristics are stated as constraints on a state transition system which mimics the observational behaviour of the human mind when reasoning about actions and their effects in the environment.

Informally, the structure of the transition system consists of a directed graph whose nodes are *states*, whose arrows are labelled by *actions*, and *causal laws* justify the state transitions. The graph is rooted in the initial state σ_0, which may either be unspecified or partially specified. There is at least one state σ_t for each time point $t > 0$, with a single incoming arrow and at least one outgoing arrow. The predecessors of any state σ_t, $t > 0$, are linearly ordered and define a *causal chain* or path $(\sigma_0, \ldots, \sigma_t)$ in the state transition system. The set of causal chains is the set of intended models

(def. 2.2.1). We use state transition systems because of their standard structure and theorem-proving technique. We prove correctness (def. 2.2.2) by structural induction on the length of the causal chain. In so doing, we follow Bacon's recommendation to use induction. Such proofs are uniform and free of gaps or appeals to intuition, as advocated by Frege.

Formally, Σ_i is defined in terms of a board-game between the physical *environment* and a Freudian *ego* (the one of the three parts of the mind that connects a person to the physical environment) [73], and it mimics the causal inferences step-by-step. We use a binomial nomenclature of the form *E-O*, to represent the epistemological and ontological characteristics respectively.

We shall now describe three specific Σ_i by successive extensions, namely *K-IA*, *K-RA* and *K-RACi*. In *K-IA*, the environment is a discrete description of an imaginary environment whose dynamics obeys to temporal inertia and boolean causal laws. *K-RA* extends *K-IA* to the case of continuous time and continuous change. The extension to the continuum solves an open problem formulated in [169, p. 293-4] [174, p. 493]. *K-RACi* extends *K-RA* to the case of non-interleaving concurrency of independent actions. In *K-RACi*, the environment is close to Newtonian Mechanics, and the given notion of truth renders the environment realistic. The mature version of *K-IA* appeared in print in [169]; the concise definition in this section is our interpretation, revision and simplification of the original, with precise references for side-by-side comparison. The first version of *K-RA* and *K-RACi* appeared in print in [25]; their definition in this section is the mature version, as used in previous works to classify the Circumscriptive Calculus of Events [28, 29] and the Calculus of Fluents [26].

Definition of the class *K-IA*

We describe a game for two, played on a board where each piece is moved according to specific rules. The board represents the environment. The two players are the agent's mind, or «ego», and the environment itself. The agent's body is an object in the environment. It is useful to picture the two players in front of the board, with a set of pieces for each player (def. 2.2.6). The game has two sets of rules that must be obeyed. The first set ensures that the configuration of the board is correct at each stage in the

development of the game (def. 2.2.8). The second set describes how the pieces can be used by each player (def. 2.2.10). The players take turns. As the games unfold, the state transitions are recorded (def. 2.2.11). The set of all correct state transitions is the *set of intended models* (def. 2.2.12).

Definition 2.2.5 (Features and Fluents) Let $\mathcal{O}$ be a finite collection of names representing objects of the environment, as perceived by the agent. Let $\mathcal{F}$ be a finite collection of names representing **features**, i.e. perceivable properties of objects or perceivable relations among objects. We specify objects as parameters for features, for example $on(block_A, block_B)$. Features have values at points in time. Let $\mathcal{T} = \mathbb{N} \cup \{0\}$ be the domain of time points, and let the symbols $=$ and $<$ represent the usual relations on natural numbers. Let $\mathcal{V}$ be a discrete domain of values for features. An *observation* is an element of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \mathcal{V}$, for example $(0, on(block_A, block_B), true)$. A *state* of the environment at time τ is a set σ_τ of elements of $\mathcal{S} = \mathcal{F} \times \mathcal{V}$, for example $\sigma_\tau = \{(f_1, v_1), \ldots, (f_n, v_n)\}$. The agent's body is an object in the environment; it has individual properties, as perceived by the agent itself, and relationships with other objects in the environment. A distinctive ability of the agent is to change the value of features by performing actions. The actions start and end at time points. Let $\mathcal{E}$ be a finite collection of names for actions. We specify objects as parameters for actions, for example $move(obj, (x, y, z), (x', y', z'))$. Features change over time. We call **fluent** any feature traced over time.

Definition 2.2.6 (Configuration of the board) We define a *board* with five places the game is played on, formally represented as the five-tuple $(\mathcal{B}, M, H, \mathcal{P}, \mathcal{C})$, where

- $\mathcal{B} = [0, n_\mathcal{B}] \subset \mathcal{T}$, where $n_\mathcal{B}$ represents the present time.
- M is a mapping which assigns values in $\mathcal{B}$ to some or all the temporal constant symbols, and values in $\mathcal{O}$ to all the object constant symbols.
- $H : \mathcal{B} \to \mathcal{S}$ is the *history*, where $\mathcal{S} \subset \mathcal{F} \times \mathcal{V}$.
- $\mathcal{P} \subset \mathcal{B} \times \mathcal{B} \times \mathcal{E}$ is the *past-action set*.
- $\mathcal{C} \subset \mathcal{B} \times \mathcal{E}$ is the *current-action set*.

For any time point $t \in \mathcal{T}$ and feature $f \in \mathcal{F}$, the state of the environment at t is $H(t)$, and the value of f at t is $H(t, f)$. We define *configuration of the*

board any instance of its formal representation. [169, p. 18]

Definition 2.2.7 (Perception) «Let $\mathcal{S}_M$ be a material-level state domain, and $\mathcal{S}$ an image-level state domain. A *perception function* is a mapping *Perc* from finite histories for $\mathcal{S}_M$ to finite histories for $\mathcal{S}$ which satisfies the following characteristics: *(1)* the parameter and value are histories over the same time period; *(2)* if H is a history over $[0,t]$, $s < t$, and $H_{[0,s]}$ is the restriction of H to the period $[0,s]$, then $Perc(H)_{[0,s]} = Perc(H_{[0,s]})$.» [169, p. 5]

Definition 2.2.8 (Correct revision of the board) Let J and J' be configurations of the board. We say that J' is a *correct revision* of J iff the following conditions hold:

- $\mathcal{B} \subseteq \mathcal{B}'$. If $b \in \mathcal{B}' \setminus \mathcal{B}$ then is $b > n_{\mathcal{B}}$.
- $M \subseteq M'$.
- The restriction of H' to $[0, n_{\mathcal{B}}]$ is equal to H.
- $\mathcal{P} \subseteq \mathcal{P}'$. If $(s,t,A) \in \mathcal{P}' \setminus \mathcal{P}$ then both $t = n_{\mathcal{B}'}$ and $(s,A) \in \mathcal{C} \setminus \mathcal{C}'$.
- If $(s,A) \in \mathcal{C} \setminus \mathcal{C}'$ then either $s = n_{\mathcal{B}}$ or $(s, n_{\mathcal{B}'}, A) \in \mathcal{P}'$.
  If $(s,A) \in \mathcal{C}' \setminus \mathcal{C}$ then $s = n_{\mathcal{B}}$. [169, p. 19]

Definition 2.2.9 (Causal Laws) A *causal law* describes the patterns of features' change over a time period, as resulting from the execution of an action. We write causal laws using the following three-step process.

STEP 1 We collect a natural and experimental history of the action that we want to model. This collection represents the action's observational behaviour.

STEP 2 We write the action's observational behaviour in a table, like a Wittgenstein's truth-table in propositional logic states the truth-value of a sentence on a case-by-case basis. We do so as follows. Let $A \in \mathcal{E}$ be the name for the action that we are modelling. Let $\sigma_s = H(s)$ be a state of the environment at time s. We define the function $Infl : \mathcal{E} \times \mathcal{S} \to \wp(\mathcal{F})$, such that $\wp(\mathcal{F})$ is the power set of $\mathcal{F}$ and $Infl(A, \sigma_s) \in \wp(\mathcal{F})$ is the set of those features whose value changes if the action A starts at time point s. We define the function *Trajs* from $\mathcal{E} \times \mathcal{S}$, such that $Trajs(A, \sigma_s)$ is a set of trajectories for $Infl(A, \sigma_s)$. A trajectory in $Trajs(A, \sigma_s)$ is a finite nonempty sequence of partial states $(\sigma_{s+1}, \ldots, \sigma_t)$, where $t \geq 1$, assigning values to

features in $Infl(A, \sigma_s)$ only. For the given action name A in $\mathcal{E}$, varying σ_s over all possible states in $\mathcal{S}$, the elements $(\sigma_s, Infl(A, \sigma_s), Trajs(A, \sigma_s))$ form a table stating the action's observational behaviour. We shall refer to it as the set-theoretic description of the action A. Varying A in $\mathcal{E}$ and σ_s in $\mathcal{S}$, the resulting elements form tables stating the complete observational behaviour of the environment. We shall refer to the collection of these tables as the set-theoretic description of the environment. [169, p. 75, 80–2, 170]

STEP 3 For a given table, as resulting from step 2, we interpret its information to identify all patterns, and then we write a logical formula to represent these patterns. This formula is the causal law that models the action. In the present definition, the causal law represents the information in the table like a formula in *full disjunctive normal form* represents a Wittgenstein's truth-table in propositional logic. We shall now describe how to read the table and write the causal law.—The action may have preconditions, intended as specific values of certain features at the time when the action starts. As we read the table, we represent any relevant statement of type «the feature f_j has have value v at time s» by a formula of type $[s]f_j = v$. Let *Antecedent* be a logical conjunction of all such formulae.—The action extends over a time period, where features are influenced. We then represent the statement «the feature f_j is occluded (prevented from being seen, i.e. neither true nor false) during $(s, t]$ and is explicitly assigned to the value w at time t» by the formula $[s, t]f_j := w$. We represent unoccluded change by a finite conjunction of $[\tau]f_j := w$ assignments, where $\tau \in (s, t]$. Let *Consequent* be a finite non-empty conjunction of such formulae.—Let s and t be temporal variables, and let M be the mapping by definition 2.2.6. We write the *causal law* for the action A as follows:

$$[s, t]A \Rrightarrow Antecedent \rightarrow Consequent$$

We represent alternative results by disjunctive statements:

$$[s, t]A \Rrightarrow \bigvee_{i=1}^{n} Antecedent_i \rightarrow Consequent_i$$

We read the causal law as follows: if the action A starts at time $M(s)$, ends at time $M(t)$, and the formulae in $M(Antecedent_i)$ are true in the state

$\sigma_{M(s)}$, then the influenced features change according to the formulae in $M(Consequent_i)$. [169, p. 128,136,137,238,170]

Example 2.2.1 Let consider a simple action A, with preconditions, duration and alternative results, where the value of influenced features is unknown during its execution. The features f_1 and f_2 are expected to be true at the time when A starts. Both features are occluded during the execution of the action. The first possible result makes both features false. The second possible result makes f_1 false and f_2 true. Thus, the action A is nondeterministic with respect to f_2 and deterministic with respect to f_1. These are given raw-data, intended as a natural and experimental history, carefully and precisely collected from experimental observations of an action performed in the environment (step 1). Generally, we have plenty of raw-data; the purpose of the table $(Infl, Trajs)$ is to reduce and arrange such raw-data with method and order (step 2). The following table states the observational behaviour of the simple action A.

σ_s	$Infl(A, \sigma_s)$	$Trajs(A, \sigma_s)$
$\{\ldots, (f_1, T), (f_2, T), \ldots\}$	$\{f_1, f_2\}$	$\{(\{(f_1, F), (f_2, F)\}),$ $(\{(f_1, F), (f_2, T)\})\}$
$\{\ldots, (f_1, T), (f_2, F), \ldots\}$	$\varnothing$	$\{(\varnothing)\}$
$\{\ldots, (f_1, F), (f_2, T), \ldots\}$	$\varnothing$	$\{(\varnothing)\}$
$\{\ldots, (f_1, F), (f_2, F), \ldots\}$	$\varnothing$	$\{(\varnothing)\}$

The first row of the table reads as follows. Let σ_s be a state of the environment at time point s, where $\sigma_s = \{\ldots, (f_1, T), (f_2, T), \ldots\}$, f_1 is true, f_2 is true also, and other elements (feature,value) in $\mathcal{S}$ may be known explicitly but are not relevant (column 1). When starting the action A in any such time point s, the action A influences (changes the values of) the features f_1 and f_2 only (column 2). The respective values of f_1 and f_2 change as stated in the table, according to two different trajectories (column 3). We then interpret the table to form the causal law (step 3). We obtain the following logical formula:

$$
\begin{array}{rcl}
[s,t]A & \Rrightarrow & (\mathit{Antecedent} \rightarrow \mathit{Consequent}_1) \vee \\
 & & (\mathit{Antecedent} \rightarrow \mathit{Consequent}_2) \\
\text{where} & & \\
\mathit{Antecedent} & \equiv & [s]f_1 = T \wedge [s]f_2 = T \\
\mathit{Consequent}_1 & \equiv & [s,t]f_1 := F \wedge [s,t]f_2 := F \\
\mathit{Consequent}_2 & \equiv & [s,t]f_1 := F \wedge [s,t]f_2 := T
\end{array}
$$

Definition 2.2.10 (Rules of the game) We define the function $\oplus\colon \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{S}$ such that $\sigma^\star = \sigma \oplus \sigma'$ is the state where $\sigma^\star(f) = \sigma'(f)$ if $\sigma'(f)$ is defined and $\sigma^\star(f) = \sigma(f)$ otherwise. Given a state σ and a trajectory $h = (\sigma_1, \sigma_2, \ldots, \sigma_t)$, we define the abbreviation $\sigma \rhd h$ for $(\sigma \oplus \sigma_1, \sigma \oplus \sigma_2, \ldots, \sigma \oplus \sigma_t)$. If H is a history over $\mathcal{B} = [0,t]$ and $h = (\sigma'_1, \sigma'_2, \ldots, \sigma'_k)$ is a trajectory, then $H' = H \rhd h$ is the updated history over $\mathcal{B}' = [0, t+k]$, where

$$
H'(s) = \begin{cases} H(t) \oplus \sigma'_i & \text{if } s = t + i > t \\ H(s) & \text{if } s \leq t \end{cases}
$$

$H \rhd (\varnothing)$ updates H from t to $t+1$ so that $H(t+1) = H(t)$, where $(\varnothing)$ is the null trajectory.—Let $(\mathcal{B} = \{0\}, M, H(0), \mathcal{P}, \mathcal{C})$ be an initial configuration of the board, where $H(0)$ is a nondeterministically chosen initial state of the environment. Let the players take turns.

- The *ego* can do one of the following at its turn:
  - start a single new action: $\mathcal{C}' = \mathcal{C} \cup \{(s, A)\}$, $s = n_{\mathcal{B}}$
  - end a running action: $\mathcal{C}' = \mathcal{C} \setminus \{(s, A)\}$, $\mathcal{P}' = \mathcal{P} \cup \{(s, t, A)\}$, $t = n_{\mathcal{B}}$
- The *environment* can do one of the following at its turn:
  - if $(s, A) \in \mathcal{C}$ then
    $\mathcal{B}' = \mathcal{B} \cup \{s + k\}$
    $H' = H \rhd h$, where $h = (\sigma'_1, \sigma'_2, \ldots, \sigma'_k) \in \mathit{Trajs}(A, H(s))$
    $\mathcal{P}' = \mathcal{P} \cup \{(s, s+k, A)\}$
  - if $\mathcal{C} = \varnothing$ then
    $\mathcal{B}' = \mathcal{B} \cup \{n_{\mathcal{B}} + 1\}$[6]

[6]There is an inconsistency between the general definition of image-level world in [169, p. 20] and its specialisation in [169, p. 76]; the former definition requires $\mathcal{B}' = \mathcal{B}$ and the latter

$$H' = H \rhd (\varnothing)$$
$$\mathcal{P}' = \mathcal{P}.$$

We use the single term *action* to refer both actions with definite duration and actions with indefinite duration. When executing an action with indefinite duration, k is a variable in $\mathcal{T}$ and its definite value will be decided by the ego. No action with definite duration can be ended by the ego. [169, p. 19,20,74,76,78]

The last part of def. 2.2.10 implements the property of *strict inertia*: *no feature changes its value except under the explicit effect of an action, within the duration of the action*. The frame problem is the problem of describing in a compact form what does not change when an action is performed [134]. The frame axiom is built in the property of strict inertia by requiring $\sigma^\star(f) = \sigma(f)$.

The game is nondeterministic, because of the nondeterministic initial configuration of the board and the nondeterministic selection of the trajectory h from *Trajs*. The environment, therefore, is a player with non-fixed strategy.

Definition 2.2.11 (Reasoning problem, scenario or narrative) A *scenario* is a description of the environment and its state transitions. A scenario in the class K-IA is a five-tuple

$$\Upsilon \equiv (K, \mathcal{O}, (IA, \textsc{law}), \textsc{scd} \cup \textsc{tc}, \textsc{obs})$$

whose components are as follows:

- $\textsc{obs} \subset \mathcal{T} \times \mathcal{F} \times \mathcal{V} = \mathcal{H}$ is a finite set of observations.
- $\textsc{scd} \subset \mathcal{B} \times \mathcal{B} \times \mathcal{E}$ is the $\mathcal{P}$ component of the board at the end of all games. Temporal variables are allowed, universally quantified over $\mathcal{B}$.
- TC is a possibly empty set of temporal constraints for SCD members.
- LAW is a set of *causal laws* defining the actions in SCD.
- IA is the family of characteristics regarding the ontological nature of the environment. The environment must enjoy the property of *strict inertia*

requires $\mathcal{B}' = \mathcal{B} \cup \{n_\mathcal{B} + 1\}$. We prefer the latter requirement, because the former would stop the flux of time between actions, thus failing with the Hanks-McDermott problem [90].

and time points must be non-negative integer numbers (I). Actions may have alternative results (A). There are no concurrent actions.

- K is the agent's epistemological constraint, meaning that knowledge about actions is explicit, accurate and complete. K is defined as follows. (1) each element of SCD is specified in terms of action symbol, parameters, starting and ending times; (2) there are no additional actions besides those specified in SCD; (3) each causal law in LAW must explicitly specify all the features it may influence (a void *Consequent* is not allowed), together with a set of preconditions for its applicability (a void *Antecedent* is allowed); (4) for every $A \in \mathcal{E}$ and $\sigma_s \in \mathcal{S}$, it is $Trajs(A, \sigma_s) \neq \varnothing$ and there is a causal law in LAW that defines how to perform A when started at time point s. [169, p. 8,25,26,35,42,44,160 and §8.2]

Definition 2.2.12 (Intended models) Let Y be a scenario in *K-IA*. The *set of intended models* for Y is as follows:

$$\Sigma_{K\text{-}IA}(\mathrm{Y}) = \{(M, H) \mid (\mathcal{B}, M, H, \mathcal{P}, \mathcal{C}) \in Mod(\mathrm{Y})\}$$

where the set $Mod(\mathrm{Y})$ is as follows: (1) construct the unique environment that is precisely specified by (IA, LAW) in Y, then select an arbitrary ego and an arbitrary initial configuration of the board; (2) generate all possible configurations of the board as correct revisions resulting from games between the ego and the environment; (3) restrict this set of configurations to those where the set of actions $\mathcal{P}$ is exactly the set of actions specified by SCD in Y, and where all formulae of SCD $\cup$ TC and OBS in Y are satisfied. [169, p. 46,180,182-185]—For any $(t, f, v) \in \mathcal{H}$ and model set $\Sigma(\mathrm{Y})$, we say that (t, f, v) is true in $\Sigma(\mathrm{Y})$ if and only if it exists an intended interpretation (M, H) such that $(\mathcal{B}, M, H, \mathcal{P}, \mathcal{C}) \in Mod(\mathrm{Y})$ and $(f, v) \in H(t)$.—We use the binomial nomenclature *K-IA* to represent the described family of epistemological and ontological characteristics.

Monotonicity and non-monotonicity are formal properties of formal reasoning methods. For a given formal language $\mathcal{L}$ and consequence relation Cn, Cn is monotonic iff for any set of premises Γ and for any sentence α in $\mathcal{L}$, $Cn(\Gamma) \subseteq Cn(\Gamma \cup \{\alpha\})$. Cn is non-monotonic otherwise. Monotonicity is based on the scientific ideal that *truth is absolute*. Non-monotonicity is based on the common experience in science that *truth is defeasible*.

Proposition 2.2 $\Sigma_{K\text{-}IA}$ is non-monotonic.

Proof SCD $\subseteq$ SCD$'$ does not imply $\Sigma_{K\text{-}IA}(\mathrm{Y}) \subseteq \Sigma_{K\text{-}IA}(\mathrm{Y}')$. [169, p. 194] q.e.d.

The full class $K\text{-}IA$ allows for qualitative uncertainty in the form of incomplete initial states and nondeterministic effects of actions. The IA family of ontological characteristics constrains the environment to features whose value persists unchanged unless there is a sequentially executed action that changes them explicitly. This is the classical case since the time of the STRIPS [68] and TWEAK [36] planning systems. In addition to this classical case, IA includes actions with extended duration, metric time, conditional effects, incomplete specification of the order of the actions and nondeterminism [169, p. 25].

We can define sub-classes of $K\text{-}IA$ by imposing constraints upon the epistemological and ontological characteristics in definition 2.2.11 [169, p. 28,42,208]. The class $Ks\text{-}IA$, for example, is defined as $K\text{-}IA$ with the additional constraint of complete knowledge about the initial state of the environment, consisting in a complete initial set of elements $(0, f, v)$ of OBS, all elements being relevant to the specific reasoning problem at hand. The class $K\text{-}IA \setminus Ks\text{-}IA$ is the class of all $K\text{-}IA$ scenarios whose initial state is either partially or completely unspecified, i.e. some or all the features are neither true nor false. To allow neither true nor false sentences is to meet both Prior's recommendation in [155] [156, ch. VIII] and the common sense experience that only a fraction of all possible aspects of the environment falls within the scope of personal knowledge. The class $Kp\text{-}IA$ restricts $K\text{-}IA$ to the case of no observations about any later state than the initial one, i.e. the class of pure prediction problems. The class $Kr\text{-}IA$ restricts $K\text{-}IA$ to the case of pointwise consistency retaining scenario, i.e. scenarios Y such that $[\![\mathrm{Y}]\!] \neq \varnothing \Rightarrow \Sigma_{K\text{-}IA}(\mathrm{Y}) \neq \varnothing$. Examples of constraints upon the ontological characteristics are the following. Ad restricts A to the case of deterministic actions, Is restricts I to single time-step actions, and Ib restricts I to boolean features, i.e. $\mathcal{V} = \{T, F\}$. We can also combine constraints; for example, $Ksp\text{-}IAd = Ks\text{-}IA \cap Kp\text{-}IA \cap K\text{-}IAd$.

Definition of the class *K-RA*

The notion of time in *K-IA* is Aristotelian[7]. In fact, the following holds by definition 2.2.10: **(i)** when an action is started by the ego, i.e. the current-action set $\mathcal{C}$ is not empty, then the environment increases the current *now* by k time points, where k is the duration of this action; **(ii)** when no action is started by the ego, i.e. the current-action set $\mathcal{C}$ is empty, then the environment starts, executes and ends a null action whose trajectory is the null trajectory $(\varnothing)$ of duration 1, so to increase the current *now* by a single time point. The causal laws justify the state transitions, and the state transitions justify the flux of time. Actions may have reversible effects on feature values without compromising the direction of time; in fact, $b \in \mathcal{B}' \setminus \mathcal{B}$ implies $b > n_{\mathcal{B}}$ by definition 2.2.8.

In pursuing the aim of amending *K-IA*'s notion of time to the continuum case, we meet the following problems.

The first problem is that no action can be physically started and ended such that its duration is infinitesimal and its repeated execution is the image-level justification to continuous time. The environment could just beat time by increasing the current *now* by an infinitesimal quantity $\varepsilon > 0$, only to lose a continuum of time points between the *now* $n_{\mathcal{B}}$ and $n_{\mathcal{B}} + \varepsilon$. *The problem of beating the continuous time consists in the non-existence of the successor function for real numbers.*

The second problem is the well-known paradox of Zeno (Elea, Italy, 489 B.C.), for which the trajectory of a running action can not describe the continuous change during the scheduled time interval. *The problem of describing the continuous change consists in the non-existence of the successor state function.* Furthermore, if we use continuous domains for both feature values and time points, then the set-theoretic description of the environment (definition 2.2.9) results in a continuum of possible initial states, together with a pair $(Infl(A,\sigma), Trajs(A,\sigma))$ for each initial state σ and action A, and a continuum of ordered partial states for each trajectory in $Trajs(A,\sigma)$,

[7] The available literature dates back to Aristotle the notion of time as measurable order of change [7, IV.11]. Time is the measure of change according to a before and an after. Time evolves because of change, and thus the absence of change causes the end of the world. Aristotle assumed the world eternal. The emphasis on the irreversibility of change, due to the second principle of thermodynamics was later made by Reichenbach [158] for which time has a growth direction.

but no such an infinite amount of information can be stored explicitly in a propositional table (see example at p. 46).

There are additional open problems. The task of scientific modelling requires evidence of the object to be modelled. When the object is human causal reasoning itself, relevant evidence is given by its linguistic representations, and thus, the collection of a natural and experimental history ranges from *plain* sentences in natural language, such as newspaper articles, to formal models themselves, such as Newtonian Mechanics. In its simplest form, we reason about actions and their effects on objects using transitive verbs in a natural language [157, p.149-50, ch. 4]. Some of us represent time and actions by tenseless verbal languages [44, p. 50] which can be included in the corpus. When taking transitive verbs as evidence, how do we model them? Different authors are entitled to their views, although the resulting models may not be the most general ones. For example, McCarthy's 1969 model is limited to the *perfect aspect* of non-concurrent actions with implicit time; it does so using LISP-like sentences of type $(action_2, (action_1, state_0))$ [134]. Allen's 1983 model uses explicit time periods, preserving McCarthy's ontological limitation to the *perfect aspect* [5]. Kowalski's 1986 model includes the *progressive aspect* of discrete non-concurrent actions with explicit time points; it does so using Horn clauses to model the verbs *hold* and *do* [103]. Recent models go beyond those ontological limitations. The underlying semantics must encompass all the available models. This is an ongoing process where research leads to more general underlying semantics.

We solve the above problems introducing the notion of *partial fluent*, being a function of time. Like a transitive verb in a natural language, a partial fluent describes an action when the action affects an object. Like a vector in Newtonian Mechanics, a partial fluent describes a force exerted on objects. In doing so, partial fluents are versatile objects. In summary, we define *K-IA*'s converse approach using partial fluents, where the notion of change is built upon the notion of time: a master clock beats time and partial fluents describe continuous change. We redefine trajectories as a continuum of ordered partial states; their set-theoretic description is a characteristic function that defines the partial states implicitly, by selection from a common domain. This stores an infinite amount of information in a finite system.

Rather than working on a case-by-case basis with all the infinite elements of the set, we work directly with the definition of the set.

We shall now proceed by amendment to *K-IA*.

Definition 2.2.13 (Features and Fluents) We amend def. 2.2.5 as follows. $\mathcal{T} = \mathbb{R}^+ \cup \{0\}$ is the domain of time points; the time structure is the classical structure for the non-negative real numbers, where the Dedekind-Cantor axiom of completeness holds true. Let $clock^0$ be a strictly monotonic rising function with no parameters. We distinguish between two types of clock: the computable function and the oracle (for example, the function which reads the clock pulse of a computer). In the first case, the agent is rationally aware of the flux of time, to know exactly how the time is beaten. In the second case, the agent is still aware of the flux of time, but has no given knowledge of how the time is beaten. Let $\mathcal{V}$ be an unsorted domain for features[8]. We call *partial fluent* any function φ such that $\varphi\colon \mathcal{T}^3 \to \mathcal{V}$ and $Domain(\varphi) = \{(s,\tau,t) \in \mathcal{T}^3 : s \leq \tau \leq t\}$. Partial fluents are single-feature trajectory descriptors during time periods. We distinguish between four types of partial fluent. *Internal and static*: a computable function whose definition is known to the agent and it does not change passing time. *Internal and dynamic*: a computable function whose definition is known to the agent and it can only change according to internal criteria, for example, machine learning. *External of the first type*: a computable function whose definition is known to the agent and it can only change according to external criteria. *External of the second type*: a partial function whose definition is not necessarily known to the agent.

Example 2.2.2 (Aristotelian clock)

$$clock = \begin{cases} 0 & \text{initial value} \\ clock + duration(h) & \text{if } \mathcal{C} = \{(n_{\mathcal{B}}, E)\} \text{ and } h \in Trajs(E, H(n_{\mathcal{B}})) \\ clock + 1 & \text{if } \mathcal{C} = \varnothing \end{cases}$$

[8] We may define $\mathcal{V} = [0,1] \subset \mathbb{R}$, but this would restrict the classification to fuzzy logics [216, 217]. We may define $\mathcal{V} = \mathbb{R}$, but this would not allow for problem domains with many-sorted features. We may allow explicitly for many-sorted features, but this would restrict the classification to many-sorted logics. On the other hand, many-sorted logic reduces to unsorted logic [62, p. 277] [139, p. 483], hence the choice for an unsorted domain for features.

An instance of the Aristotelian clock is the Linnaean Floral Clock [119, pg. 274–276], where time is beaten by the blooming of flowers at different seasons of the year; to perceive the change, the definition uses an external partial fluent of the second type. In the above examples, time is built upon a notion of change. A different example, where time is an oracle and free from any notion of change, it is our Continuously Branching Clock [30], where time can be pictured as a tree with a continuum of branches, obtained by removing the total-order relation in the axiomatisation of non-negative real numbers.

Definition 2.2.14 (Causal Laws) We amend def. 2.2.9 as follows. Let $A \in \mathcal{E}$ be an action, let σ_s be any state of the environment such that $Infl(A, \sigma_s) \neq \varnothing$, and let n be the number of alternative results of A. We associate partial fluents to features using elements $(\tau, f, \varphi(s, \tau, t)) \in \mathcal{T} \times \mathcal{F} \times \mathcal{V} = \mathcal{H}$, meaning that at time τ the feature f has value $\varphi(s, \tau, t)$, where $t - s$ is the duration of an executed action A such that $f \in Infl(A, \sigma_s)$. Let $\varphi_{1j}, \ldots, \varphi_{ij}, \ldots, \varphi_{nj}$ be the partial fluents that are associated to each feature f_j in $Infl(A, \sigma_s)$. The i-th trajectory of A, for $i = 1, \ldots, n$, is the non-empty and totally-ordered set $(T_i(A, \sigma_s); \leq)$, where

$$T_i(A, \sigma_s) = \{(\tau, f_j, \varphi_{ij}(s, \tau, t)) \in \mathcal{H} \,:\, f_j \in Infl(A, \sigma_s)\}.$$

The set $Trajs(A, \sigma_s)$ is the set of all possible trajectories of the action A started at time point s. The *table* $(Infl, Trajs)$ is the set-theoretic description of the action. We shall now represent the information in the table by a logical formula. We represent antecedents and consequents of influenced features by partial fluents. We represent occlusion by discontinuities. We represent the natural language sentence «the feature f_j is expected to have value v at time $\tau = s$, it is occluded at time $\tau \in (s, t)$ and is explicitly assigned to the value w at time $\tau = t$» by the partial fluent

$$\varphi_j(s, \tau, t) = \begin{cases} v & \text{if } \tau = s \\ \text{undefined} & \text{if } \tau \in (s, t) \\ w & \text{if } \tau = t \end{cases}$$

We associate partial fluents to features using the first-order formula

$$S_{ij} \equiv \forall\, \tau \in [s,t] \subset \mathcal{T}\,.\,[\tau] f_j = \varphi_{ij}(s,\tau,t)$$

A *causal law* is a formula of the form

$$[s,t]A \Rrightarrow \bigvee_{i=1}^{n} \Big(\bigwedge_{j=1}^{m} S_{ij} \Big)$$

where the formula $[s,t]A$ is expanded into a formula in Full Disjunctive Normal Form, that is, into a disjunction of conjunctions of S_{ij} formulae, each of which corresponds to the feature f_j in the action alternative i. The expressiveness of causal laws depends directly on the expressiveness and type of partial fluents in it.

Example 2.2.3 We re-write the causal law in example 2.2.1 as $[s,t]A \Rrightarrow \bigvee_{i=1}^{2}(\bigwedge_{j=1}^{2} S_{ij})$, where the partial fluents in S_{ij} are defined as follows:

$$\left| \begin{array}{ll} \varphi_{11}(s,\tau,t) = \begin{cases} T & \text{if } \tau = s \\ T \vee F & \text{if } \tau \in (s,t) \\ F & \text{if } \tau = t \end{cases} & \varphi_{12}(s,\tau,t) = \begin{cases} T & \text{if } \tau = s \\ T \vee F & \text{if } \tau \in (s,t) \\ F & \text{if } \tau = t \end{cases} \\ \\ \varphi_{21}(s,\tau,t) = \begin{cases} T & \text{if } \tau = s \\ T \vee F & \text{if } \tau \in (s,t) \\ F & \text{if } \tau = t \end{cases} & \varphi_{22}(s,\tau,t) = \begin{cases} T & \text{if } \tau = s \\ T \vee F & \text{if } \tau \in (s,t) \\ T & \text{if } \tau = t \end{cases} \end{array} \right|$$

We observe that, in the above *matrix of partial fluents*, it is $\varphi_{11} = \varphi_{12} = \varphi_{21}$, and thus we re-write the matrix as follows:

$$\left| \begin{array}{ll} \varphi_{11}(s,\tau,t) & \varphi_{11}(s,\tau,t) \\ \varphi_{11}(s,\tau,t) & \varphi_{22}(s,\tau,t) \end{array} \right| = \left| \begin{array}{ll} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right|$$

We can describe concisely actions with numerous features and alternative results. Any element a_{ij} of the matrix is the descriptor of the feature j in the action alternative i. Each row of the matrix describes an action alternative. Each column of the matrix describes the corresponding feature in each action alternative. This embodies our view that causal laws

should be stated clearly and concisely, and be suitable for database storage and update, away from any awkward and error-prone process where formulae are handwritten, parsed, reduced and compiled before usage. Ideally, the agents should be able to observe an action and infer its causal law autonomously, and thus models of causal reasoning should rather include inductive-learning techniques [140, 141] than reduction algorithms. We consider the matrix of partial fluents a step forward in this direction.

Definition 2.2.15 (Perception) When writing causal laws, we use external partial fluents of the second type as perception function, to sense the physical environment as often as desired, then we verify the correspondence between the physical environment and its abstract representation. The resulting notion of truth is Aristotelian.—When we associate a partial fluent φ_{ij} to the feature f_j in the action alternative i, we are associating a specific sensor with a specific feature. The partial fluent φ_{ij} hides the hardware; its definition depends on the specific hardware and its machine language. The partial fluent φ_{ij} reads the raw data from its sensor, interprets the raw data, and uses the interpreted data as value for its corresponding feature f_j. Action alternatives are allowed, and thus φ_{ij} may be defined to offer different interpretations of the same raw data. If one wishes to perform these interpretations explicitly, φ_{ij} can be defined as the null interpretation, to use the raw data as direct value.

The resulting approach differs from both the traditional *perception-thought-action* model [33, p. viii] and the *perception-action* model by Brooks [33, p. xi]. It differs from the traditional model because the *perception* subsystem is here part of the *action* subsystem: to perceive means to act in order to perceive, and actions are controlled by the reasoning subsystem. It differs from Brooks's model because agents here act in order to perceive, selecting what to observe, and for how long (we retain control of sensors on resource-bounded agents and use them when needed, for as long as necessary). No such ability is possible in Brooks's model, because it lacks the reasoning subsystem. Those perceptions that in a biological system may not involve a deliberate action, such as feeling heat through the skin, they are formally described as «internal» actions.

Example 2.2.4 (Aristotelian truth) «Such appears to be the truth, judging from theory and from what are believed to be the facts about them; the facts, however, have not yet been sufficiently grasped; if ever they are, then credit must be given rather to observation than to theories, and to theories only if what they affirm agrees with the observed facts.» (Aristotle [8, III.10])—Let $\varphi_1, \varphi_2, \varphi_3$ be partial fluents such that φ_1 changes the environment, φ_2 perceives the environment, and φ_3 verifies the correspondence between them, namely between theoretical truth (values of featured in the logical environment, as resulting from reasoning) and observed experimental facts (values of features in the physical environment, as resulting from perception). The feature f_1 has an *internal* value (φ_1 is an internal static partial fluent), and f_2 has an *external* value (φ_2 is an external partial fluent of the second type). The corresponding causal law is as follows:

$$[s,t]A \Rrightarrow \bigwedge_{i=1}^{3} \forall\, \tau \in [s,t] \subset \mathcal{T}\,.\,[\tau]f_i = \varphi_i(s,\tau,t)$$

where

$$\varphi_3(s,\tau,t) = \begin{cases} 1 & \text{if } [\tau]f_1 = [\tau]f_2 \\ 0 & \text{otherwise} \end{cases}$$

Definition 2.2.16 (Rules of the game) Let $(\mathcal{B} = [0, clock], M, H(0), \mathcal{P}, \mathcal{C})$ be an initial configuration of the board, where $H(0)$ is a nondeterministically chosen initial state of the environment. We amend def. 2.2.10 as follows. The partial fluent *clock* starts beating time. Let the players take turns. The rules for the ego are not amended. The rules for the environment are amended as follows.

- At time $s = clock$ the ego communicates its decision to start a new action A by adding the element (s, A) to $\mathcal{C}$. Then the move of the environment consists in revising the board as follows: At time s, the environment nondeterministically selects a trajectory $h = T_i(A, H(s))$ from the trajectory set $Trajs(A, H(s))$. From s to $t = s + k$, where $k = duration(h)$, the state changes according to h: $(f_j, \varphi_{ij}(s, \rho, k)) \in H(\rho)$ for all $\rho \in [s, s + k]$. At time t the environment adds the element (s, t, A) to $\mathcal{P}$ and resets $\mathcal{C}$ to the empty set.

- At time $t = clock$ the ego communicates its decision to end the running action A by removing (s, A) from $\mathcal{C}$ and adding (s, t, A) to $\mathcal{P}$. Then the environment persists in its values during $[t, clock]$, until the ego starts a new action.
- The ego has not yet decided to start a new action, i.e. the current-action set $\mathcal{C}$ is empty at $n_{\mathcal{B}} = \nu$. Then the environment persists in its values during $[\nu, clock]$, until an initialisation message is received from the ego.

The agent does not perceive necessarily, although it may do so; in this case, we use $[\Box]A$ instead of $[s, t]A$ in SCD, meaning that the action A is executed at all times. All sensors are part of the agent's body, which is an object of the physical environment, and thus the «environment», as a player, is in charge of their management during the game.

The property of strict inertia (p. 48), including the frame axiom, is implemented by requiring $(f_j, \varphi_{ij}(\nu, \rho, l)) \in H(\rho)$ for all $\rho \in [\nu, \nu + l]$. As a corollary, sensing actions do not influence features other than those explicitly specified in the causal law.

The game is nondeterministic because of the nondeterministic initial configuration of the board and the nondeterministic selection of the trajectory h from *Trajs*. The environment, therefore, is a player with non-fixed strategy.

Definition 2.2.17 (Reasoning problem, scenario or narrative) We amend definition 2.2.11 as follows. Let $\mathcal{T}$ be the time point domain, $\mathcal{F}$ the set of all feature symbols, $\mathcal{V}$ the domain for features, $\mathcal{E}$ the set of all names for actions. Let $(\mathcal{H}, \sqsubseteq)$ be the lattice whose elements, called *observations*, are members of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \wp(\mathcal{O}) \times \mathcal{V}$ and the order relation $\sqsubseteq$ applies as follows: $(t_1, f_1, v_1) \sqsubseteq (t_2, f_2, v_2)$ iff $t_1 \sqsubseteq t_2$. The element $(t, f, unknown)$ is an abbreviation for $\bigvee_i (t, f, v_i)$, varying i over all possible elements (t, f, v_i) of $\mathcal{H}$. Let $(\mathcal{D}, \sqsubseteq)$ be the lattice whose elements, called *rigid occurrences of actions*, are members of $\mathcal{D} = \mathcal{T} \times \mathcal{T} \times \mathcal{E}$ and the order relation $\sqsubseteq$ applies as follows: $(s_1, t_1, A_1) \sqsubseteq (s_2, t_2, A_2)$ iff $s_1 \sqsubseteq s_2$. The order relation $\sqsubset$ is an abbreviation for $\sqsubseteq \wedge \neq$. The relation $(s_1, t_1, A_1) = (s_2, t_2, A_2)$ means that A_1 and A_2 start at the same time point, while $(s_1, t_1, A_1) \sqsubset (s_2, t_2, A_2)$ means that A_1 starts earlier than A_2. Let Y be a scenario.

- The part OBS of Y is a sub-lattice of $(\mathcal{H}_Y, \sqsubseteq)$, whose elements are members of $\mathcal{H}_Y = \mathcal{T} \times \mathcal{F}_Y \times \wp(\mathcal{O}) \times \mathcal{V} \subseteq \mathcal{H}$, where F_Y is the set of all features explicitly occurring in Y.
- The part SCD of Y is a sub-lattice of $(\mathcal{D}, \sqsubseteq)$. Each element of SCD specifies an action scheduled for execution, with the timepoint when the action starts, the timepoint when the action ends, and the action's own name.
- The function $\Rrightarrow : \mathcal{D} \rightarrow \wp(\mathcal{H}_Y)$ maps each element of SCD to a set of non-empty lattices of observations. The function $\Rrightarrow$ is parametric on the action type, and the part LAW of Y consists in the definition of $\Rrightarrow$ as a set of causal laws in Full Trajectory Normal Form, one causal law for each action type. The Full Trajectory Normal Form for the causal laws is a mapping $(s, t, A) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_{ij}$ for which the action occurrence (s, t, A) is *expanded* into a formula in Full Disjunctive Normal Form, that is into a disjunction of conjunctions of trajectory formulae S_{ij}, each of which corresponds to the feature f_j in the alternative i. For a given feature f_j in $\mathcal{F}$ and partial fluent φ_j defined over $D \subseteq [s, t] \subset \mathcal{T}$, $s \neq t$, a trajectory formula is the following first-order formula:

$$S_{ij} \equiv \forall\, \tau \in [s, t] \subset \mathcal{T}\,.\,[\tau] f_j = \varphi_j(s, \tau, t)$$

- RA is the family of characteristics regarding the ontological nature of the environment. The environment must enjoy the property of *strict inertia*, time points must be non-negative real numbers (R), and actions may have alternative results (A). There are no concurrent actions.

Definition 2.2.18 (Intended Models) We amend definition 2.2.12 as follows. The family of ontological characteristics is RA, instead of IA. The set $Mod(Y)$ is obtained, at point (2), using the rules of the game by definition 2.2.16.

Proposition 2.3 $\Sigma_{K\text{-}IA}(Y) \subseteq \Sigma_{K\text{-}RA}(Y) \subseteq [\![Y]\!]$.

Proof The model set $\Sigma_{K\text{-}RA}(Y)$ equals $\Sigma_{K\text{-}IA}(Y)$ in case of discrete environments with Aristotelian Clock, and $\Sigma_{K\text{-}IA}(Y)$ is empty in case of continuous or many-sorted environments, so that $\Sigma_{K\text{-}IA}(Y) \subseteq \Sigma_{K\text{-}RA}(Y)$. The set $[\![Y]\!]$ allows models which do not satisfy inertia and models

where there are additional actions besides those specified in SCD, so that $\Sigma_{K\text{-}RA}(\Upsilon) \subseteq [\![\Upsilon]\!]$. q.e.d.

Definition of the class *K-RACi*

In *K-RA* only one action can be executed at a time, including sensing actions; the environment is inhabited by a single agent, the agent can execute a single action at a time, and the action takes complete control of the influenced features. We shall now amend *K-RA* to the case of concurrency of independent actions. Following the binomial nomenclature [169, p. 26], the case is designated with the ontological characteristic *Ci*.

We make no structural extensions to *K-RA*. We first amend its rules to allow the case of n ego players, where each ego plays its game with the environment, concurrently and independently from each other. Each ego player is also allowed to start several concurrent independent actions. We then amend the scenario to allow more than a single element of the current-action set. The construction of the intended model set for *K-RA* is identical to *K-RACi*'s own. We finally show the relation between *K-RACi* models and *K-RA* models.

Definition 2.2.19 (Rules of the game) Using definition 2.2.16, if an ego starts several independent actions at the same time τ, then the environment selects one trajectory function nondeterministically. If the partial fluents $\varphi_j^1, \ldots, \varphi_j^n$ describe the same feature f_j during $[\tau, \mu]$, then only one partial fluent is considered for that feature, namely the composed partial fluent $\varphi = \varphi_j^1 \circ \cdots \circ \varphi_j^n$ of the given partial fluents, and $Domain(\varphi)$ is the time period $[\tau, \mu]$. When such composed function is a constant function c, then persistence holds for the influenced feature if and only if it exists a small positive $\varepsilon \in \mathcal{T}$ such that $H(\tau - \epsilon) = c(\xi)$, for all ξ in $[\tau, \mu]$ (inertia with influencing actions).

At the beginning of the game the ego may start all sensing actions for the sensors in its body, each running concurrently and independently with the others. If a new sensor becomes available, a new sensing action can be added. If a sensor dies, its sensing actions may end.

Definition 2.2.20 (Reasoning problem, scenario or narrative) We amend definition 2.2.17 as follows. The family of ontological characteristics is

RACi. The environment must enjoy the property of *strict inertia* and time points must be non-negative real numbers (*R*). Actions may have alternative results (*A*). Actions may be concurrent, i.e. $\#\mathcal{C} \geq 0$ at any time point (*C*). Concurrent actions must be independent from each other although their set of influenced features may overlap, i.e. the set $\bigcap_{i=1}^{k} Infl(A_i, \sigma_s)$ may not be empty (*Ci*).

Proposition 2.4 Let $\Upsilon_1, \ldots, \Upsilon_n$ be scenarios, each of which is of the form

$$\Upsilon_i = (K, \mathcal{O}, (RA, \textsc{law}_i), \textsc{scd}_i, \textsc{obs})$$

Let $\cup_{i=1}^{n} \textsc{scd}_i$ be a schedule appropriate for the *RACi* ontological family. The following relation holds.

$$\Sigma_{K\text{-}RACi}(\bigcup_{i=1}^{n} \Upsilon_i) = \bigcup_{i=1}^{n} \Sigma_{K\text{-}RA}(\Upsilon_i)$$

Proof Independent concurrent actions are compositional, because their joint effect is the sum of their individual effects. q.e.d.

Corollary 2.5 $\Sigma_{K\text{-}RA}(\Upsilon) \subseteq \Sigma_{K\text{-}RACi}(\Upsilon) \subseteq [\![\Upsilon]\!]$

2.3 Results (summary)

In summary, the method consists in collecting models as data, arrange the collection in a table, according to shared epistemological and ontological characteristics, and decide equivalence and subsumption relations with former classified models. The results consist in proofs by structural induction on the length of the causal chain, thus accepting the recommendation of Bacon and Frege about making progress using induction.

The table at page 63 summarises the results obtained using the described method, spanning seventy-two years of research in the field. The table includes the Handbook's former summary of results [174, p. 492]. The new results are identified with a star.

2.4 Discussion

By comparison with the original results [174, p. 492], the new results show the broader scope of the extended method. Formerly unclassifiable models can now be compared. Brandano's *Calculus of Fluents* has the broadest

range of correct applicability, and subsumes all classified models to date. The results also show that various models have essentially the same underlying structure, that is, they show how different authors discovered essentially the same reasoning using different languages.

Our experience of research shows that the classification of any given reasoning model generates more than just another theorem. The classifications are instructive, because they communicate the distilled essence of a model's thought processes with a detailed examination of its elements and structure. For any given model, the research exercise begins by collecting relevant literature, reading broadly and thoroughly to understand both the anatomy and physiology of the model, seeking for its characteristics. The research unfolds in a refinement process, where data and patterns build up, being guided by the classification method. From the first sketch of classification up to its final version, one understands the model deeply enough to purge its original literature from any appeal to intuition, and thus write a new concise description of the model in addition to its classification. We did not add confusion to an already chaotic field of research. We believe we attained order and clarity.

	MODEL OF CAUSAL REASONING	CLASS	COMPLEXITY CLASS
*	Brandano's Calculus of Fluents (NAS, or Non-simulative Algebraic Semantics)	*K-RACi* [26]	
*	**Classical Mechanics**	*Ksp-RAdCi*	
*	Calculus of Events (Circumscriptive, Continuous, redesigned)	*Ksp-RA*	
*	Calculus of Events (Circumscriptive, Continuous)	*Ksp-RA*	
	PMON (Pointwise Minimisation of Occlusion with No-change premisses)	*K-IA* [169, p. 243,247]	
	CAMC (Chronological Assignment and Minimisation of Change)	*K-IA* [169, p. 218,228]	
	GMONF (GMON with Filtering)	*K-IA* [169, p. 243,248]	
	CAMOC (Chronological Assignment and Minimisation of Occlusion and Change)	*K-IA* [169, p. 241,245]	
	CMOC (Chronological Minimisation of Occlusion and Change)	*K-IAe* [169, p. 240,246]	
	PGMF (PGM with Filtering)	*K-IAe* [169, p. 214,215]	
	PCMF (PCM with Filtering)	*K-IAex* [169, p. 213,215]	
*	Answer Set (Stable Model) Programming [81] [65, p. 251] [94]	*K-IbsAd*	
*	Logic Programming with Abduction and Integrity Constraints	*K-IbsAd*	coNP-complete
*	PDL (Dynamic Logic) [70, 154]	*K-IbsAd*	
	ADL by Pednault [147]	*K-IbsAd* [174, p. 488], [95]	
	Explanation-closure approach by Reiter [163]	*K-IbsAd* [174, p. 488], [95]	
	State-space approach by Baker [19]	*K-IbsAd* [174, p. 488], [95]	
*	Calculus of Events (Circumscriptive, Discrete)	*Ksp-IA*	
*	**Turing's Choice Machines**	*Ksp-IA*	
*	Logic Programming (SAS, or Simulative Algebraic Semantics)	*Ksp-IAd* [24]	
*	**Turing's Automatic Machines**	*Ksp-IAd*	
*	Calculus of Events (Circumscriptive, Boolean)	*Ksp-IbA* [28]	
*	**Turing's Computing Machines**	*Ksp-IbAd*	
	TMOC (Two-stage Minimisation of Occlusion and Change) [148]	$\subseteq$ *Kr-IA* [169, p. 244,250,276]	
	GMON (Global Minimisation of Occlusion with No-change premisses) [127, 128]	$\subseteq$ *Kr-IA* [169, p. 242,249,276]	
	PGM (Prototypical Global Minimisation) [127, 128]	$\subseteq$ *Kr-IA* [169, p. 200,208,276]	
	PCM (Prototypical Chronological Minimisation)	$\subseteq$ *Kr-IA* [169, p. 198,203,276]	
	OCM (Original Chronological Minimisation) [97]	$\subseteq$ *Kr-IA* [169, p. 199,206,276]	

Table 2.1: Results

3

MODELS (PART I)

3.1 Introduction

The search for a small number of laws that could explain the whole range of observed phenomena has been the aim of Physics for centuries, where Physics played the role of experimental science and Mathematics provided both the formal language and the computational theories for its models. The result of this endeavour is Classical Mechanics, a mathematical model of causal reasoning about the physical environment. This model, described in three laws, demolished the common sense picture of the world that the pre-Galileans were trying to develop within Mathematics. Although it was a common sense absurdity that objects could exert mutual influence if they were not in contact, this model proved that the physical environment is governed by forces that one could neither see with naked eye nor touch with bare hands. Four hundred years later, the aim of using Mathematical Logic for representing natural-language information about the effects of natural and artificial events in the physical environment led to the *frame problem* [134, p. 487]. Since then, a community has been trying to develop a common sense picture of the world within Mathematical Logic [180].

Despite all efforts, Classical Mechanics is a renown mathematical model of causal reasoning that no mathematical logic has yet succeeded to emulate. Classical Mechanics involves epistemological and ontological assumptions. Our aim is to classify this model, to measure the distance of classified logics with respect to the target class.

3.2 Methods

We used the systematic paradigm, as precisely described in chapter 2. The assessment required extensions to the paradigm. We described these extensions in chapter 2. In the following sections we therefore proceed with the development of the theory and assume that it is understood what is meant

by «correctness», «classification», «relations of equivalence and subsumption», «full abstraction», and the definition of the class *K-RACi*.

3.3 Results

Let us recall the fundamental laws of Classical Mechanics and a relevant theorem by Cauchy. Summarised by Newton [143], the first two laws were known to Galilei [80], the third law was known to work on fluids by Archimedes.

First Law: «Every body persists in its state of either rest or uniform motion along a straight line, unless compelled to change that state by forces impressed thereon.» **Second Law:** Whenever a net force F acts on a body it produces an acceleration a, along the direction of the force, which is proportional to the magnitude of the force and inversely proportional to the mass m of the body, that is $F = ma$. **Third Law:** «To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.» **Principle of independence of simultaneous actions:** The net force exerted on a body is the vectorial sum of the individual forces exerted on the body by the material systems interacting with it.

Theorem 3.1 (Cauchy [35]) Let a and b be continuous functions over a time period $I \subset \mathbb{R}$. Given $t_0 \in I$ and $x_0 \in \mathbb{R}$, it exists and is unique the solution to the following problem:

$$\begin{cases} x(t_0) = x_0 \\ x'(t) = a(t)\, x(t) + b(t) \end{cases}$$

The solution $x(t)$ is computable with the following formula:

$$x(t) = e^{A(s)} \left(x_0 + \int_{t_0}^{t} e^{-A(s)} b(s)\, ds \right), \text{ where } A(s) = \int_{t_0}^{t} a(s)\, ds.$$

We shall now prove the following.

Proposition 3.2 The epistemological and ontological characteristics of Classical Mechanics are *Ksp-RAdCi*.

Proof Let F be the component of the external force along the direction of the movement, let m be the mass (constant) of the object, let a be the acceleration of the object, and let $x(t)$ be the position in space of the object at time t. The equation $F = ma$ is the second order differential equation $F(t, x(t)) = mx''(t)$, which is valid for any point in time t. To mathematically describe the motion of the object exactly means to find a function $x(t)$ as the solution to such differential equation that satisfies the initial conditions of position and velocity of the object. The equation $x''(t) = \frac{1}{m}F(t, x(t))$ then reduces to the following problem:

$$\begin{cases} v_1'(t) = v_2(t) \\ v_2'(t) = \frac{1}{m}F(t, v_1(t)) \end{cases} \quad \text{where} \quad \begin{cases} v_1(t) = x(t) \\ v_2(t) = x'(t) \end{cases}$$

By Cauchy's theorem, given the initial conditions $v_1(0)$ and $v_2(0)$, the solution to the resulting system exists and is unique, that is the motion is deterministic.

Newton's laws and Cauchy's theorem call for the *Ksp-RAdCi* family of characteristics. The first law, or principle of inertia, jointly with the use of the continuum in both Newton's theory and Cauchy's theorem, finds an immediate correspondence with the property of strict inertia in continuous time, which is built explicitly in the underlying semantics for *K-RACi*. The epistemological sub-characteristic *s* restricts *K* to the case of complete knowledge about the initial state, which is the necessary condition to apply Cauchy's theorem. The epistemological sub-characteristic *p* restricts *K* to the case of no information at any later state than the initial one, because such information is not required to apply Cauchy's theorem. The ontological sub-characteristic *d* restricts *A* to the case of deterministic change, for which the effect of an action is completely determined by the state of the physical environment at the time when the action started, which is properly the result of applying Cauchy's theorem. The ontological sub-characteristic *i* restricts *C* to the case of concurrency of independent actions, whose definition, by construction, finds an immediate correspondence with the principle of independence of simultaneous actions, where the net force exerted on a body *B* and consisting of *n* forces, corresponds to *n* simultaneous actions influencing the feature *position in space* of the

object B. The second and third law involve information that can be represented explicitly in the scenario. q.e.d.

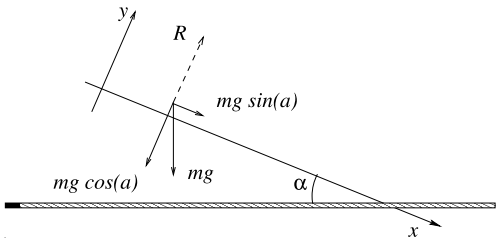
We note that although Kuipers did not acknowledge Cauchy in his work [106, p. 316], he did refer to a Cauchy problem as the definition of a «dynamical system».

3.4 Discussion

The following positive examples are merely demonstrative of the result.

Example 3.4.1 (Galilei's inclined plane [80]) Given both a horizontal and an inclined plane, let $\alpha \geq 0$ be the angle between the two surfaces. We place an iron ball of mass m on the inclined place, and we release it. Will this object move along the inclined plane? What are the involved actions at any given point in time τ and what is their effect?

Figure 3.1: Galilei's inclined plane



It easily verified that all theories of causal reasoning, as summarized by Shanahan [180], they all fail to reason correctly about this problem. Indeed, according to their temporal inertia, the object will persist in its initial state, because no action is directly exerted on the object by an agent.

We shall now translate the above reasoning problem from its natural-language description to a formal scenario. The knowledge about the initial state of the environment is complete. The mass of the object is m, the angle between the inclined plane and the horizontal plane is α, and thus we write both $[0]mass \mathrel{\hat{=}} m$ and $[0]angle \mathrel{\hat{=}} \alpha$ in the OBS part of the scenario. Two forces are always exerted on the object, the weight force $W = mg$ and the reaction R of the plane. As no other force is exerted, the SCD part of the scenario consists of the two formulae $[\Box]W(o)$ and $[\Box]R(o)$, where $o \in \mathcal{O}$

is an object constant. Finally, the LAW part of the scenario consists of the causal laws for W and R:

$$[s,t]W(obj) \Rrightarrow \forall \tau \in [s,t].[\tau]force(obj) = \varphi_1(s,\tau,t)$$

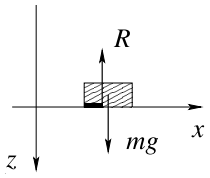
$$[s,t]R(obj) \Rrightarrow \forall \tau \in [s,t].[\tau]force(obj) = \varphi_2(s,\tau,t)$$

where s and t are temporal variables, obj is an object variable, and $force$ is the feature we are interested in. If $\circ$ is the symbol for the vectorial sum, then the partial fluent φ_1 is defined as the vectorial sum of $[0]mass * g * sin([0]angle)$ and $[0]mass * g * cos([0]angle)$, and φ_2 is defined as $-[0]mass * g * cos([0]angle)$. According to the underlying semantics, the force exerted on the object o at time τ is $H(\tau, force(obj)) = \varphi_1(s,\tau,t) \circ \varphi_2(s,\tau,t) = mg\, sin(\alpha)$. The following may occur:

$$mg\, sin(\alpha) = \begin{cases} > 0 & \text{if } 0 < \alpha < \frac{\pi}{2} \\ < 0 & \text{if } \frac{\pi}{2} < \alpha < \pi \\ = 0 & \text{if } \alpha = 0 \text{ or } \alpha = \pi \\ = mg & \text{if } \alpha = \frac{\pi}{2} \end{cases}$$

If $0 < \alpha < \frac{\pi}{2}$ or $\frac{\pi}{2} < \alpha < \pi$, then the object moves along the inclined plane with uniformly-accelerated motion. If $\alpha = 0$ then the object is lying on a horizontal plane; this is the case of inertia with influencing actions. If $\alpha = \frac{\pi}{2}$, then the inclined plane is a vertical plane and $\varphi_2(s,\tau,t) = 0$. (If we use glue or a magnet to exert adequate resistance on the object, then $\varphi_2(s,\tau,t) = -\varphi_1(s,\tau,t)$.) Passing time α may change because of a lifting action; the underlying semantics is *elaboration tolerant* in this respect, because the causal laws have α as parameter.

Figure 3.2: Galilei's inclined plane: case of inertia



Galilei's inclined plane involves the frame problem, as well as the prediction and causal-qualification problems. The **prediction problem** is described as follows [90]; «given an initial description of the physical environment [known values for a set of features at time point 0], the occurrence of some events, and some notion of causality (that an event occurring can cause a fact to become true), what facts are true once all the events have occurred?» The **qualification problem** is «the phenomenon that if an action normally causes an effect at a later time, the effect may be modified or inhibited because of other events in the intervening period» [167].

The frame problem has its solution with Newton's principle of inertia. The property of *strict inertia* (p. 48, 58) is a generalisation of Newton's principle to all features of objects in the environment, where the position in space is one of the possible features. The temporal projection problem here consists in predicting the motion of the object given the initial conditions of position and velocity, and is addressed in its general terms by proposition 3.2. The causal-qualification problem occurs as follows. The gravity force normally causes the objects to fall; the phenomenon may be fully inhibited because of the friction exerted by the horizontal plane, which causes the object to rest in its position according to the principle of inertia. The underlying semantics gives the same solution via the operator $\circ$; the free-fall situation corresponds to the case $\alpha = \frac{\pi}{2}$ without friction; the situation where the phenomenon is inhibited corresponds to any α with sufficient friction to exert a force equal and opposite to $mg\ sin(\alpha)$.

The inclined plane scenario is useful in practical applications. Robots are physical agents with limited resources, and a robot with wheels, for example, may have to climb an inclined platform with the available energy, taking into account its own weight and the minimum expense of energy that is required to achieve this goal with a certain desired speed. The above formulation shows how this problem can be represented in a $Ksp\text{-}RAdCi$ scenario.

The following variant of the above scenario leads to the dividing-instant problem. The **dividing-instant problem** is described as follows: determine the value of a feature at that instant in time dividing two periods where the feature has neat different values. One either violates the law of contradiction, by claiming both values, or the law of excluded third, by claiming

neither. The problem is commonly attributed to Plato [150].

Example 3.4.2 (Newton's apple, variant of) Given the above scenario 3.4.1, let be $\alpha = 0$ and let F be an initial force that throws the object. The object will raise along a straight vertical line. When the object reaches a certain hight h, it returns to the horizontal plane. Is the object raising or falling at the apex of its revolution?

In mathematics, if a continuous function $f(x)$ admits two different values for the same x, then the function admits a discontinuity in that point. If the function admits no value for a given x, then the function is simply not defined for that x. In classical logic, the former case corresponds to the law of contradiction; the latter corresponds to the law of excluded third. With no need to violate neither laws, continuity models the common experience that things change smoothly. An immediate consequence of continuity, which is respected by all systems of qualitative reasoning, is that a changing quantity must pass through all intermediate values[9], thus passing through the dividing-instant point. The solution to the above problem is, that at the apex of its revolution the object is neither raising nor falling.

The problem of establishing whether the object is at rest or at motion admits immediate solution. By common sense, in fact, an object is resting if it is lying motionless, which situation corresponds to the above case of inertia with influencing actions, with $h = 0$ and $R > 0$. In the present scenario, the object is not resting, because $h \neq 0$ and $R = 0$, even though it is neither raising nor falling at the apex of its revolution.

The dividing-instant problem seems to arise only if actions with instantaneous effects are allowed in the theory. In that case, the consequences of waiving classical logic, when violating either laws, must be accepted. The dividing-instant problem does *not* occur for *K-RACi* scenarios, because trajectories are nonempty sets, and thus instantaneous effects are *not* allowed.

The following variant of Galilei's inclined plane scenario leads to the struc-

[9] By Bolzano's theorem, if f is a continuous function on a closed interval $[a, b]$ and $f(a)$ and $f(b)$ have opposite signs, then there exists a number c in the open interval (a, b) such that $f(c) = 0$.

ture-based ramification problem. The **ramification problem** is «the phenomenon where the physical environment consists of a number of objects with relationships between them, and changes in one object (for example, the effects of actions on that object) are dependent on features of adjacent objects in the structure» [167].

Example 3.4.3 Given the above scenario 3.4.1, let be $\alpha = 0$ and let a wet ice-cube o_2 stand on a block o_1. We push gently on o_1. By common experience, we know that o_2 will move along with o_1.

The scenario is similar to Baker's ice-cream scenario [19], where an agent (o_1) holds the ice-cream (o_2) while moving slowly to a different location. We represent the problem using the operator $\circ$. The block o_2 rests on the block o_1 according to the known case of inertia with influencing actions; the force F is so tiny that block o_2 does not fall off o_1. At the same time, block o_1 moves along a straight line with constant velocity. The block o_2, which forms a single system with o_1, moves with o_1. If, however, the force F is sufficiently strong, we know by common experience that o_2 will fall off o_1. To address this case, we can introduce inertia in the equations, and $\circ$ still represents the problem successfully.

The following scenario involves concurrent actions with both cumulative and cancelling effects.

Example 3.4.4 Given the above scenario 3.4.1, we place an edged stand on the horizontal plane, the inclined plane on top of the stand, and a wet ice-cube on top of the inclined place. Two distinct lifting/lowering actions can be exerted at the edges of the inclined plane. The first task consists in moving the inclined plane so that the ice-cube will stand still on it. We then remove the stand, or lift the inclined plane, to achieve the same result.

A similar scenario was studied by Gelfond, Lifschitz and Rabinov [83], and Shanahan [181], where an agent moves a soup bowl sideways, trying not to spill the soup (move the ice-cube). The scenario is also a classical problem of control theory in various industrial applications. The problem admits immediate solution using the operator $\circ$. The ice-cube moves along the inclined plane, i.e, the bowl spills soup, iff the effect of the lifting/lowering actions are not cancelling, i.e., the net force exerted on the

object is not null.

Example 3.4.5 (Domino Effect) We align five Domino tokens so that if the token $i-1$ falls then the token i is pushed. If we push the first token and wait, we expect the last token to fall after a short while. We run the experiment, and observe that the last token falls as expected. If we realign the tokens, insert and hold a card between the fourth and the fifth token, we expect the last token to persist in its initial state. We run the experiment, and observe that the last token does not fall, as expected. We then generalise the first experiment to any number of tokens, using any arrangement of multiple paths, where any token falls only if at least two preceding tokens in its path fall on it.

A domino effect is a situation in which one event causes a series of similar events to happen one after the other (Oxford English Dictionary). The situation involves the following problems of causal reasoning: cause-and-effect chains, structure-based ramification, delayed effects of actions, concurrency of independent actions, and may also involve causal qualification. In the above example, and in particular in its last but one case, all these problems occur at once. Despite its apparent complexity, we readily solve the example via the operator $\circ$. Any token *obj* falls only if the net force exerted on *obj* is sufficiently strong, i.e., only if the following holds:

$$H(\tau, force(obj)) = \varphi_1(s,\tau,t) \circ \ldots \circ \varphi_i(s,\tau,t) \circ \ldots \circ \varphi_k(s,\tau,t) \geq C$$

where $k \geq 1$ is the number of independent forces that are exerted concurrently on *obj*, $C = 1$ in any case of the example except the last one where $C = 2$, and the pushing is described as follows:

$$\varphi_i(s,\tau,t) = \begin{cases} 0 & \text{if } \tau \in [s,t) \\ 1 & \text{if } \tau = t \end{cases}$$

We use a single causal law to model the action exerted on a generic token: $[s,t]Press(obj_i, obj_j) \Rrightarrow \forall\, \tau \in [s,t].[s]follows(obj_i, obj_j) \rightarrow [\tau]force(obj_j) = \varphi_i(s,\tau,t)$. We represent the occurrence of our initial action with the single formula $[0]Press(thumb, t_1)$ in the part SCD of the scenario. Finally, we rep-

resent any specific arrangement of possibly multiple paths via relations between tokens, with formulae of type $[0]follows(t_i, t_j)$ in the part OBS of the scenario. It is easily verified that the underlying semantics models the scenario as expected.

Example 3.4.6 (Achilles and the tortoise, by Zeno) The tortoise defied the fleet-footed Achilles to a single contest. They would have competed in running, which according to the tortoise would have made her victorious, on condition of a small initial advantage (one step) which Achilles would have allowed.

The given scenario is properly an example of rectilinear and uniform motion in Classical Mechanics. We describe the scenario as follows:

$$
\begin{array}{ll}
\text{OBS:} & [0.0]position(Achilles) \mathrel{\hat{=}} 0.0 \\
 & [0.0]position(Tortoise) \mathrel{\hat{=}} 1.0 \\
 & [\Box] \sqsubset (\mathbf{s}_2, \mathbf{s}_1) \mathrel{\hat{=}} true \\
\text{SCD:} & [\mathbf{s}_1, t_1]Run(Achilles) \\
 & [\mathbf{s}_2, t_2]Run(Tortoise) \\
\text{LAW:} & [s,t]Run(athlete) \Rrightarrow \forall \tau \in [s,t].[\tau]position(athlete) = \varphi(s, \tau, t)
\end{array}
$$

where $\mathbf{s}_1$ and $\mathbf{s}_2$ are temporal constants, and s, t, t_1, t_2 are variables. The scenario assumes rectilinear motion with null speed variation, therefore the partial fluent φ is $\varphi(s, \tau, t) = V(athlete) * \tau$ for all τ in $[s, t]$, and represents the covered space by a certain running athlete. The velocity $V(athlete)$ is a certain constant velocity V_A for Achilles, and a certain constant velocity V_T for the tortoise, where $V_A > V_T > 0$. The scenario clearly falls within *Ksp-RAdCi*, and the trajectory of the running action describes the continuous change during the scheduled time period. Let now solve Zeno's riddle. *Is there a point in time and space where Achilles reaches the tortoise?*

Let A and T be the points in space representing Achilles and the tortoise. If $d = 1.0$ is the initial distance between A and T, then the covered spaces S_T and S_A are described by the law for the uniform motion: $S_A(\tau) = V_A * \tau$ and $S_T(\tau) = d + V_T * \tau$. Let denote with τ_1 the point in time where A reaches the new position $S_T(\tau_1)$, that is $S_A(\tau_2) = S_T(\tau_1)$. As τ_2, τ_3, etc, are denoted in the similar way, we define the succession (τ_n) as $S_A(\tau_1) = d$ and $S_A(\tau_{n+1}) = S_T(\tau_n)$, and thus $S_A(\tau_{n+1}) = V_A * \tau_{n+1} = d + V_T * \tau_n = S_T(\tau_n)$. The following relations hold, because $V_A * \tau_1 = S_A(\tau_1) = d$:

$$\tau_1 = \frac{d}{V_A}, \tau_{n+1} = \frac{d + V_T * \tau_n}{V_A}, \tau_{n+1} - \tau_n = \left(\frac{V_T}{V_A}\right)^n * \tau_1$$

Zeno asserts that Achilles will never reach the tortoise, that is the sum of all time periods $(\tau_{n+1} - \tau_n)$ diverges. The famous argument goes that in order for Achilles to overtake the tortoise he must first reach the position where the tortoise is, which is impossible, because by the time he gets there the tortoise has crawled forward. The argument remained valid for about 2300 years until Leibniz, Newton, Cauchy, Weierstrass, Dedekind and Cantor devised a proper definition of length, distance and sum of an infinite series. According to the resulting account on series, we have:

$$\tau_1 + \sum_{k=1}^{+\infty}(\tau_{k+1} - \tau_k) = \tau_1 + \sum_{k=1}^{+\infty}\left(\frac{V_T}{V_A}\right)^k * \tau_1 = \tau_1 * \left(1 + \sum_{k=1}^{+\infty}\left(\frac{V_T}{V_A}\right)^k\right) =$$

$$= \tau_1 * \sum_{k=0}^{+\infty}\left(\frac{V_T}{V_A}\right)^k \overset{\star}{=} \tau_1 * \frac{1}{1 - \frac{V_T}{V_A}} = \frac{V_A * \tau_1}{V_A - V_T} = \frac{d}{V_A - V_T} = \tau^*$$

where $(\star)$ $\frac{V_T}{V_A} < 1$ by hypothesis. The fleet-footed Achilles will then reach the tortoise at τ^*, in fact is $S_A(\tau^*) = S_T(\tau^*)$. It is readily verified that the underlying semantics meets the mathematical solution, that is: $[\tau^*]position(Achilles) = [\tau^*]position(Tortoise)$.

In the following example, we use an external partial fluent of the first type as a function $\varphi_i : \mathbb{R}^{3+} \to \mathbb{R}^3$ to represent the i-th airway within a certain airspace, as generated by an air-traffic scheduler (which is "external" to the aircraft) so that for any $\tau \in [s, t]$ no other aircraft is using the i-th airway at τ (the row i of the matrix is null except at column j). Passing time, the agent receives a φ_i from the air-traffic scheduler, to control the aircraft accordingly. The aircraft is the agent and the co-pilot system is its ego. The airspace is represented by the matrix of partial fluents, where rows are airways, columns are aircrafts, and the resulting causal law is part of a Traffic Collision Avoidance System.

Example 3.4.7 (Traffic Collision Avoidance System) Given an air space with n air-ways in it, we assume the followings. Let φ_w be the partial fluent describing the air-way w, for any $w = 1 \dots n$. Let a_i^w be the i-th aircraft

using the air-way w, for any $i = 1 \dots f(w)$, where $f(w)$ is the number of aircrafts that can fly along the air-way w at the same time (w is a pipeline of length $f(w)$). Suppose each a_i^w is moving of rectilinear and uniform motion, on cruise control, and its position along φ_w is given by its GPS. Further, suppose that each a_i^w has a TCAS such that, if a_i^w has distance $\leq d_i^w$ from another aircraft, then they both start collecting information on their respective positions in space, via the GPS and the radar, and a safety action is performed when needed, so to keep the safety distance and avoid a collision. Given the initial position of all aircrafts at time t_0, will the aircrafts avoid the collisions?

Example 3.4.8 (static obstacles) The air space has no predefined airways, and a single aircraft has to fly safely through skyscrapers.

Example 3.4.9 (dynamic cooperating and negotiating obstacles) The air space has no predefined airways, and skyscrapers are replaced by other aircrafts. Each aircraft is moving freely and independently from the others. All aircrafts cooperate and negotiate to avoid collisions.

Example 3.4.10 (dynamic hostile obstacles) The air space has no predefined airways. The dynamic obstacles are non-cooperating and non-negotiating.

The scenario 3.4.7 falls within the *Ksp-RAdCi* family of characteristics; it is similar to the «Achilles and the tortoise» problem, where athletes run along various directions at different speeds. It is a simple exercise to write a causal law such that the property $alert(a_i)$ becomes 'true' if $H(\nu)position(a_i) - H(\tau)position(a_i) = d_i$, where ν is the present time point and τ is the point in time where the collision would otherwise take place. The difference between this scenario and its three variants rests entirely in the representation of trajectories: scenario 3.4.8 uses no trajectories at all, scenario 3.4.9 uses external partial fluents of the first type, and scenario 3.4.10 uses external partial fluents of the second type. Scenario 3.4.9 requires the extension of *K-RACi* to the case of concurrency of interacting actions.

The following examples show limitations of the result.

Scenario 3.4.11 (Ghost Forces) Suppose two agents are sitting one in front of the other. The table between them is perfectly horizontal and frictionless. One of the players gently pushes away the disk of dry ice, along the table, towards the other player. The table is long enough for both players to observe the trajectory of the disk. The player knows exactly how to push the disk, and knows also that the trajectory will ideally follow a straight line. The experiment is repeated a number of times, where the disk behaves as expected. The experiment is then repeated once more. At this time, the disk follows an unexpected trajectory, still being able to reach the other side of the table. The players, surprised, can not explain the phenomenon.

This is a reasoning scenario for a theory of relative truth. According to Classical Mechanics (Einstein's Special Theory of Relativity), an *observer* is needed. It can be located in an inertial or non-inertial system of co-ordinates, with consequences in his comprehension of phenomena. Two cameras recorded the experiment from above. One camera was fixed to the table, the other to the ceiling. The first camera recorded the experiment as we described it. According to the other camera, instead, during all the experiments the disk moved along a straight line. The first camera was located in the inertial system, the second camera in the non-inertial one. During the last experiment the table was rotating gently, so to affect the disk; the players, surrounded by wholes, could not be aware of it.

The class *K-RACi* is unable to describe (represent and reason correctly about) this type of problems, namely to observe the same situation from different viewpoints, which we regard as a special case of the more general ability to change either the context or the reasoning. Logic models of viewpoints do exist; see, e.g. [14, 188, 15, 16, 17]. The open problem then consists in defining a relevant extension of *K-RACi*, sufficient and good to assess this type of scenarios.

Also known as the *principle of action and reaction*, the third law deduces one of the fundamental and most popular principles of the whole Physics, the *principle of preservation of the quantity of motion*: «the total quantity of motion ($Q = mv$) of a system with two point-like objects which are subject exclusively to their mutual interaction, remains constant passing time». The latter principle deduces the former as well, and both are well-known classical

constructions. There are no known exceptions to the principle of preservation of the quantity of motion. In any isolated system, each time the principle seems violated during an experiment, a third interacting object is discovered which is external to the (known) system, and the validity of the principle is reestablished by including the object in the system. That was the case, for example, with the neutron and the photon, and the planet Neptune. To establish a similar principle within formal nonmonotonic reasoning, hence within the underlying semantics, it would mean to establish a reasoning rule for which when reasoning with incomplete knowledge leads to conclusions that are inconsistent with the known facts, what becomes crucial is the ability to postulate additional knowledge and to validate the consistency of the resulting theory. It was Turing who first introduced the idea that being intelligent implies also making errors. He then imagined a machine equipped with a method of drawing conclusions by induction. Methods of drawing conclusions and conjecturing hypothesis by induction have then been developed since the eighties, like the program BACON [108] and inductive learning in Logic Programming [140, 141]. Although the information provided by the third principle can be explicitly represented in the scenario and involves no evident additional characteristics, the principle itself is equivalent to the principle of preservation of the quantity of motion, which use involves the reasoning skill we have just mentioned.

An additional limitation arises from a letter of Einstein to Popper [153, p. 522]: «I would like to repeat that, in my opinion, you are not right in sustaining the thesis that it is impossible to derive statistical conclusions out of a deterministic theory. It is sufficient to think about classical statistical Mechanics (gas theory or the theory of the Brownian motion). Example: a material dot moves of constant velocity along a closed circle; I can calculate the probability of finding it, at a certain given time, at a certain position along the perimeter. What is essential is the following: I do not know the initial state, or I do not know it with precision!»—Although the full *K-RACi* class allows for environments with incomplete knowledge about the initial state, Einstein's example clearly does not belong to the class. The work by Pearl [145, 146] may be a starting point in this respect.

3.4.1 Conclusions

Using a theorem by Cauchy on the solution of first-order linear differential equations, the subclass *Ksp-RAdCi* of *K-RACi* is shown to include the epistemological and ontological assumptions that were implicitly involved in the three laws of Classical Mechanics. The boundaries of the result are outlined, suggesting two growth directions: viewpoints and learning from inconsistencies.

This report updates [25, 27]. For a relation between non-monotonic reasoning and Quantum Physics, see the 2002 report by Gabbay and Engesser [64], although their work does not include any assessment result.

4

MODELS (PART II)

4.1 Introduction

The Horn-clause fragment of Tarskian first-order logic, also known as the angelic fragment of positive logic programming [100, 213], is a renown model of human reasoning and general purpose model of computation. Despite being computationally complete [192, 176], this model has been extended with negation [41, 42, 81], abduction [65], temporal constraints, causal laws and axioms [103] tailored for solving test examples of the frame problem at a time where non-monotonic reasoning seemed to be a promising direction for research beyond the limits of both Tarskian logic and Turing machines.

Is it really true that the angelic fragment of positive logic programming is inadequate as a model of causal reasoning and, specifically, as a solution to the frame problem? What is the added value of its non-monotonic extensions?

4.2 Methods

The standard method of research for such models consisted in «appeals to intuition» through test examples of common sense reasoning [46, 180]. For a given theory, a progress report consisted in evidence of its failure through a test example, the description of an ad-hoc extension, and appeals to intuition to support the claim that the resulting theory passed the new test. According to both Dijkstra [56] and McCarthy [129, 132], this corresponds to the activity of pragmatic engineers, although Dijkstra despises it and McCarthy argues for it.

Our interest is the scientific modeling of human causal reasoning. Our aim is *not* to make a program to solve a test example. Our aim is to decide equivalence and subsumption relations between the classes of computations for which scientific models of human causal reasoning are provably

correct, and thus we depart from the above method of research. We use the systematic paradigm [chapter 2] as gold standard for the assessments.

We can solve different problems using essentially the same reasoning. We refer to this ability as **problem (domain) independence**. We can also express essentially the same reasoning using different languages. We refer to this ability as **language independence**. We expect these human abilities to have a correspondence in the formal models. For any given model of causal reasoning, we assess its problem independence by classifying its range of correct applicability; indeed we know that any classified model is correctly applicable to a whole class of reasoning problems, and not only to single examples, and that full-abstraction follows from the classification. We assess language independence by observing the effect of induced language shifts on its range of correct applicability. We conjecture that positive logic programming may not be language independent, namely that changing its language changes its problem-solving power. The following assessments shall answer to this conjecture.

In the following sections we therefore proceed with the development of the theory and assume that it is understood what is meant by «correctness», «classification», «relations of equivalence and subsumption», «full abstraction», and the definition of the class *K-RACi*. Working knowledge of lattice theory is also required, jointly with the original work by van Emden and Kowalski [100, 213].

4.3 Results

A scenario Y (def. 2.2.11 at p. 48) is a recursive definition where the formulae in OBS and SCD are regarded as *facts* and the formulae in LAW are regarded as *rules* which infer new facts from a number of known facts given as premisses. In the specific case of LAW, inferred facts are elements of $\mathcal{H}$. The semantics of Y can be defined as the least fixpoint of a mapping T associated with Y itself. An interpretation for Y is any subset of $\mathcal{H}_Y$. The set of interpretations for Y is the power set $\wp(\mathcal{H}_Y)$, which is a complete lattice under the partial order of set inclusion. The top element of this lattice is $\mathcal{H}_Y$, the bottom element is the empty set. For all $I \in \wp(\mathcal{H}_Y)$, we define the mapping $T_Y : \wp(\mathcal{H}_Y) \to \wp(\mathcal{H}_Y)$ as follows:

$$
\begin{aligned}
T_{\Upsilon}(I) = \{(\tau, f, v) \in \mathcal{H}_{\Upsilon} : \\
& (\tau, f, v) \in \textsc{obs} \text{ and } \tau = 0, \text{ or} \\
& \text{exists } (\mathbf{s}, \mathbf{t}, A) \in \textsc{scd}, \\
& \text{exists } (s, t, A) \Rrightarrow \bigwedge_{j=1}^{m} S_j \in \textsc{law} \text{ and} \\
& \text{exists a valuation } \theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\} \\
& \text{such that:} \\
& (\tau, f, v) \in \mathit{Consequents}(s\theta, \tau, t\theta, S_j(\theta)) \\
& \text{and } \mathit{Antecedents}(s\theta, \tau, t\theta, S_j(\theta)) \subseteq I \}
\end{aligned}
$$

$$
\begin{aligned}
\mathit{Antecedents}(s, \tau, t, S) &= \{(s, f, \varphi(s, s, t)) \in S \, : \, s \sqsubseteq \tau \sqsubseteq t \} \\
\mathit{Consequents}(s, \tau, t, S) &= \{(\tau, f, \varphi(s, \tau, t)) \in S \, : \, s \sqsubset \tau \sqsubseteq t \}
\end{aligned}
$$

By construction, the reasoning performed by T_{Υ} is identical to the reasoning of van Emden and Kowalski's T_P function:

$$
\begin{aligned}
T_P(I) = \{A \in B_P : \\
& A \leftarrow A_1 \wedge \ldots \wedge A_n \text{ is a ground instance} \\
& \text{of a clause in } P \text{ and } \{A_1, \ldots, A_n\} \subseteq I\}
\end{aligned}
$$

The order in which causal laws are retrieved in Υ is not relevant for its semantics, as all matching occurrences are taken into account and, by construction, there is exactly one occurrence of S_j for each feature in a given causal law. The selection rule of S_j in the body of the causal law is the same as the selection of the atom in the body of a definite Horn clause.

The above compositional model-theoretic fixpoint semantics is a formal model originally designed to speak the formal language of definite Horn clauses. Using the above construction, this model now speaks the language of *K-IA*. The reasoning model is still the original reasoning model by van Emden and Kowalski. We shall now study this model in detail. The following are formal properties of T_{Υ}.

Proposition 4.1 An interpretation I is a model for Υ iff $T_{\Upsilon}(I) \subseteq I$.

Proof I is a model for Υ *iff* for all $(\mathbf{s}, \mathbf{t}, A) \in \textsc{scd}$, $\mathit{Antecedents}(\tau, A) \subseteq I$ implies $\mathit{Consequents}(\tau, A) \subseteq I$ *iff* $T_{\Upsilon}(I) \subseteq I$. q.e.d.

Proposition 4.2 $T_{\Upsilon} : \wp(\mathcal{H}_{\Upsilon}) \rightarrow \wp(\mathcal{H}_{\Upsilon})$ is a monotonic mapping.

Proof We have to prove that $I_1 \subseteq I_2$ implies $T_\Upsilon(I_1) \subseteq T_\Upsilon(I_2)$, for each $I_1, I_2 \in \wp(\mathcal{H}_\Upsilon)$. If $I_1 \subseteq I_2$, then is $I_2 = I_1 \cup \{(t_1, f_1, v)_1, \ldots, (t_n, f_n, v_n)\}$. By definition of T_Υ, if $I_1 \subseteq I_2$ then is $T_\Upsilon(I_2) = T_\Upsilon(I_1) \cup I_3$, where $I_3 \in \wp(\mathcal{H}_\Upsilon)$ is the set of all consequences of those actions whose preconditions are satisfied in I_2 but not in I_1; if no such actions exist, I_3 is simply the empty set. As Υ meets the $\mathcal{K}$ epistemological characteristic, its actions are not allowed to have an empty set of postconditions, so that I_3 may never be an empty set for successfully executed actions. Then $T_\Upsilon(I_1) \subseteq T_\Upsilon(I_2)$ is true. q.e.d.

As $\wp(\mathcal{H}_\Upsilon)$ is a complete lattice and $T_\Upsilon : \wp(\mathcal{H}_\Upsilon) \rightarrow \wp(\mathcal{H}_\Upsilon)$ is a monotonic mapping, the ordinal powers of T_Υ can be defined as follows:

$$
\begin{aligned}
T_\Upsilon \uparrow 0 &= \{(0, f, v) \in \mathcal{H}_\Upsilon : (0, f, v) \in \textsc{obs}\} \\
T_\Upsilon \uparrow \tau &= \{(\tau, f, v) \in \mathcal{H}_\Upsilon : \\
&\qquad \text{exists } (\mathbf{s}, \mathbf{t}, A) \in \textsc{scd}, \\
&\qquad \text{exists } (s, t, A) \Rrightarrow \textstyle\bigwedge_{j=1}^{m} S_j \in \textsc{law} \text{ and} \\
&\qquad \text{exists a valuation } \theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\} \\
&\qquad \text{such that:} \\
&\qquad (\tau, f, v) \in \mathit{Consequents}(s\theta, \tau, t\theta, S_j(\theta)) \text{ and} \\
&\qquad \mathit{Antecedents}(s\theta, \tau, t\theta, S_j(\theta)) \subseteq T_\Upsilon \uparrow (\tau - 1)\}
\end{aligned}
$$

Theorem 4.3 (Tarski, 1955 [200]) Let $\mathcal{U} = (A, \leq)$ be a complete lattice and let $f : A \rightarrow A$ be a monotonic mapping. Then f has a least fixpoint, $lfp(f)$, and a greatest fixpoint, $gfp(f)$. Furthermore, $gfp(f) = lub\{x : f(x) = x\} = lub\{x : f(x) \geq x\}$ and $lfp(f) = glb\{x : f(x) = x\} = glb\{x : f(x) \leq x\}$.

Corollary 4.4 T_Υ admits a least fixpoint that is equal to the greatest lower bound of its pre-fixpoints, that is $lfp(T_\Upsilon) = glb\{I : T_\Upsilon(I) = I\} = glb\{I : T_\Upsilon(I) \subseteq I\}$.

The corollary guarantees the existence of $lfp(T_\Upsilon)$. The constructive characterisation is given in terms of the ordinal powers of T_Υ, i.e. $lfp(T_\Upsilon) \supseteq T_\Upsilon \uparrow \omega$.

Proposition 4.5 T_Υ is a continuous mapping.

Proof We have to prove that $T_\Upsilon(lub(X)) = lub(T_\Upsilon(X))$, for each directed subset X of $\wp(\mathcal{H}_\Upsilon)$. Let X be a directed subset of $\wp(\mathcal{H}_\Upsilon)$, and let

be $(\tau, f, v) \in \mathcal{H}_Y$. $(\tau, f, v) \in T_Y(lub(X))$ if and only if both $(\tau, f, v) \in Consequents(\tau, A)$ and $Antecedents(\tau, A) \subseteq lub(X)$, for some executed action A. This is true if and only if both $(\tau, f, v) \in Consequents(\tau, A)$ and $Antecedents(\tau, A) \subseteq I$, for some $I \in X$. This is true if and only if $(\tau, f, v) \in T_Y(I)$, for some $I \in X$. This is true if and only if $(\tau, f, v) \in lub(T_Y(X))$.

q.e.d.

Theorem 4.6 (Kleene, 1952 [99, p. 349]) Let $\mathcal{U} = (A, \leq)$ be a complete lattice and $f : A \to A$ be continuous. Then, $lfp(f) = f \uparrow \omega$, where ω is the first limit ordinal.

Corollary 4.7 $lfp(T_Y) = T_Y \uparrow \omega$.

The fixpoint semantics of Y is defined as the least fixpoint of T_Y. Given a scenario Y and a query $(t, f, v) \in \mathcal{H}$, the semantics of (t, f, v) is defined as the truth-value resulting from the application of the membership function for $lfp(T_Y)$ to (t, f, v) itself.

We have a finite lifetime, and thus any model of human causal reasoning must take this limitation into account. We then assume that the interaction between any agent and the environment leads to a finite game. This implies that any scenario has a finite description Y. Using this limitation, we can apply another well-known result of fixpoint theory: as Y is finite, then T_Y is a monotonic function over the finite and complete lattice $(\mathcal{H}_Y, \sqsubseteq)$, so that T_Y just needs a finite number of iterations to reach its least fixpoint. More precisely, there exists an ordinal τ such that $T_Y \uparrow (\tau + 1) = T_Y \uparrow \tau$. If Y has the same intended model set for $Y(\mu)$, where μ is the maximum time point occurring in Y, then $T_Y \uparrow \omega = T_{Y(\mu)} \uparrow \omega = T_{Y(\mu)} \uparrow \mu$. Thus, the fixpoint semantics may also be understood as an operational semantics for finite Ys, although its practical application would be inefficient.

We shall now assess the described model. Let the relation $(t, f, v) \in \Sigma(Y)$ be a shorthand for «(t, f, v) is true in $\Sigma(Y)$», i.e. «it exists an interpretation (M, H) such that the development $(\mathcal{B}, M, H, \mathcal{A}, \mathcal{C})$ is in $Mod(Y)$ and $H(t, f) = v$», according to the known definitions of intended model set.

Theorem 4.8 (Classification) For any $Y \in Ksp\text{-}IAd$ and for any $(t, f, v) \in \mathcal{H}_Y$, the following relation is true: $(t, f, v) \in T_Y \uparrow \omega \Leftrightarrow (t, f, v) \in$

$\Sigma_{Ksp\text{-}IAd}\mathrm{Y}$.

Proof The game between the ego and the environment starts at time $\tau = 0$. Because of the epistemological sub-characteristic s, the initial state of the environment is known and is represented by elements $(0, f, v)$ of OBS. This corresponds to $T_\mathrm{Y} \uparrow 0 = \{(t, f, v) \in \mathcal{H}_\mathrm{Y} : (t, f, v) \in \text{OBS and } t = 0\}$, and thus the thesis is true for $t = 0$. – The environment persists until the ego communicates its intention to perform an action, i.e. no element occurs in SCD whose starting time is $\tau - 1$ and $\sigma_0 = \sigma_1 = \ldots = \sigma_{\tau-1}$. This corresponds to $T_\mathrm{Y} \uparrow 0 = \ldots = T_\mathrm{Y} \uparrow (\tau - 1)$, and thus the thesis is true for $t = \tau - 1$. – Suddenly, the ego adds the element (τ, E) to the current-action set $\mathcal{C}$, where τ is the current time. Then, by definition, the environment executes the action and ends it at τ' by removing (τ, E) from $\mathcal{C}$ and adding (τ, τ', E) to the past-action set $\mathcal{A}$. The ego may also decide to end E earlier, let say at $\tau'' \in (\tau, \tau')$, i.e. it may autonomously remove (τ, E) from $\mathcal{C}$ and add (τ, τ'', E) to $\mathcal{A}$. The correspondence with T is as follows. By definition, $T_\mathrm{Y} \uparrow \tau$ finds the element (τ, τ', E) (or (τ, τ'', E)) in SCD, then evaluates whether the antecedents for E are satisfied. If the antecedents are satisfied, the knowledge is increased accordingly; otherwise the action is executed without any effect and the knowledge is not increased (we already proved T_Y is rising monotonic). Because of the epistemological sub-characteristic p, no observation occurs in OBS at later time points than the origo, and no alternative results of actions are allowed because of the ontological restriction d upon A. The game ranges to infinity, where the intended model set is defined. By Kleene's theorem, T_Y reaches the greatest lower bound of its pre-fixpoints with its least fixpoint, that is $lfp(T_\mathrm{Y}) = glb\{I : T_\mathrm{Y}(I) = I\} = glb\{I : T_\mathrm{Y}(I) \subseteq I\} = T_\mathrm{Y} \uparrow \omega$. q.e.d.

Corollary 4.9 For all $\mathrm{Y} \in Ksp\text{-}IAd$, $T_\mathrm{Y} \uparrow \omega \subseteq [\![\mathrm{Y}]\!]$.

Proof $T_\mathrm{Y} \uparrow \omega = \Sigma_{Ksp\text{-}IAd}\mathrm{Y} \subseteq \Sigma_{K\text{-}IA}\mathrm{Y} \subseteq [\![\mathrm{Y}]\!]$. q.e.d.

The full abstraction of any model with respect to the equivalence of causal reasoning scenarios is a default corollary of its classification chapter 4.

We can translate the model in theorem 4.8 into a PROLOG program. It is sufficient to rewrite any member (τ, f, v) of OBS into atoms $HoldsAt(\tau, f, v)$, any member (s, t, A) of SCD into atoms $Happens(s, t, A)$, any member

$(\tau, f, \varphi(s,\tau,t))$ of the S part of LAW into $HoldsAt(\tau, f, \varphi(s,\tau,t))$ atoms, and any member $(s,t,A) \Rrightarrow \bigwedge_{j=1}^{m} S_j$ of LAW into $Happens(s,t,A) \leftarrow \bigwedge_{j=1}^{m} S_j$ definite Horn clauses. We remark the absence of negation symbols. We also remark that this is not a program, to solve a reasoning problem; this is a class of computations that display a desired behaviour, to solve a class of reasoning problems.

Theorem 4.10 (Classification) Abductive Logic Programming with integrity constraints (ALP) belongs to the class *K-IbsAd* and the computational complexity of its entailment problem is coNP-complete.

Proof ALP is sound and complete with respect to the Action Language $\mathcal{A}$ [50, 49, 51] [82]; the result consists in a sound and complete transformation from $\mathcal{A}$ scenarios to *open logic programs* with integrity constraints, where the reasoning procedure adopted for the resulting programs is the SLDNFA Resolution Rule. The range of correct applicability of the Action Language $\mathcal{A}$ is *K-IbsAd* and the computational complexity of its entailment problem is coNP-complete [111]. The thesis is true by transitivity on these results.

q.e.d.

Theorem 4.11 (Classification) Answer Set Programming belongs to the class *K-IbsAdCi*.

Proof Answer Set Programming is sound with respect to the Action Language $\mathcal{A}$ [82].The range of correct applicability of the Action Language $\mathcal{A}$ is *K-IbsAd*. The thesis is true by transitivity on these results. q.e.d.

4.4 Discussion

As a model of human causal reasoning, the range of correct applicability of positive logic programming is null *if we use its original language*. As we know well, the model fails with the Hanks-McDermott problem [89, 90], and *Ksp-IAd* is the smallest class to include its correct solution [169]. This means that there is at least one problem in *Ksp-IAd* for which the model fails, and thus, the model is correctly applicable to either a sub-class of *Ksp-IAd* or no class at all. The available evidence suggests that *Ksp-IAd* is the elementary class, and thus the above cases overlap.

However, theorem 4.8 shows that positive logic programming is adequate as a model of causal reasoning. The model is not language inde-

pendent, because its range of correct applicability changes with the language. The correctness result was obtained inducing a language shift: we divided the reasoning from the original language of definite Horn-clauses, and used *K-IA*'s own language for the classification. The language shift changed the language but not the logic. Using ordinals as time points and the upward inductive process as master-clock, the proof shows the model's ability to simulate the game semantics of *K-IA* when the environment has a fixed strategy, leading precisely to the class *Ksp-IAd*. The proof also shows the model's ability to simulate a many-valued non-monotonic temporal reasoning with a two-valued monotonic non-temporal semantics. We can demonstrate this ability by solving the renown Hanks-McDermott problem. We represent this problem as follows.

OBS is the set $\{(0, alive, true), (0, loaded, false)\}$
SCD is the set $\{(2, 4, Load), (6, 8, Shoot)\}$
LAW is the set containing the following elements:

$$[s,t]Load \Rrightarrow \forall\, \tau \in [s,t]\,.\,[\tau]loaded = to\_load(s,\tau,t)$$

$$[s,t]Shoot \Rrightarrow \forall\, \tau \in [s,t]\,.\,[\tau]alive = to\_die(s,\tau,t) \wedge$$
$$[\tau]loaded = to\_shoot(s,\tau,t)$$

The actions are described by the following partial fluents:

$$to\_load(s,\tau,t) = \begin{cases} unknown & \text{if } \tau \in (s,t) \\ true & \text{if } \tau = t \end{cases}$$

$$to\_shoot(s,\tau,t) = \begin{cases} true & \text{if } \tau = s \\ unknown & \text{if } \tau \in (s,t) \\ false & \text{if } \tau = t \end{cases}$$

$$to\_die(s,\tau,t) = \begin{cases} unknown & \text{if } \tau \in (s,t) \\ false & \text{if } \tau = t \end{cases}$$

The following is true by theorem 4.8.

$$(5, alive, true) \in T_Y \uparrow \omega \quad (5, loaded, true) \in T_Y \uparrow \omega$$
$$(9, alive, false) \in T_Y \uparrow \omega \quad (9, loaded, false) \in T_Y \uparrow \omega$$

The proof of theorem 4.8 gives step-by-step instructions into how this is done, and ensures that the reasoning model gives the correct answers to this specific reasoning problem as well as to any other problem in the class *Ksp-IAd* of epistemological and ontological characteristics.

The thesis of theorem 4.8 cannot be improved. It is readily verifiable by counter-examples that any problem Y in the class $K\text{-}IA \setminus Ksp\text{-}IAd$ falls beyond the scope of the given model. Indeed, for any Y using $K \setminus Ksp$, T_Y fails because it is unable to perform backward reasoning. Further, for any Y using $IA \setminus IAd$, T_Y fails because it is unable to perform nondeterministic reasoning.

By measuring the progress with respect to both ALP and ASP, the classifications show that part of the original problem-solving power of positive logic programming was lost in these non-monotonic extensions: the epistemological characteristics improved from *Ksp* to full *K*, but an ontological characteristics regressed from *I* to *Ibs*. This suggests a growth direction for both theories ALP and ASP.

5

MODELS (PART III)

5.1 Introduction

The Calculus of Events [103] is a useful application of logic programming with negation [100, 213, 42], where programs are designed to solve problems in legal reasoning, medical informatics and cognitive robotics. As chiefly reviewed by Shanahan [180], many designs aimed at improving the problem-solving power of such programs, where the proposed correctness consisted in appeals to intuition through test examples. This approach to program correctness is unsatisfactory from the standpoint of disciplined programming. The test examples can be an effective way to show the presence of errors and limitations on a design, as shown by Hanks and McDermott [90], but are hopelessly inadequate for proving the absence of errors [56] and the range of correct applicability [56, 174]. Further, the appeals to intuition do not generalise to relevant computational properties, such as the full abstraction of the calculus (the program function) with respect to the equivalence of reasoning scenarios (the input), and give no formal instrument to decide equivalence and subsumption relations of the calculus with respect to other models of causal reasoning.

The aim of this work is to answer to the above problems. We thus present assessments for the Calculus of Events in its latest descriptions, namely Shanahan's circumscriptive axiomatisations defined using Sandewall's filtering technique [179, 180, 181, 182], to gain more insight into this model of causal reasoning and raise the confidence level of its programs significantly.

5.2 Methods

We used the systematic paradigm, as precisely described in chapter 2. The assessments required extensions to the paradigm. We described these extensions in chapter 2. In the following sections we therefore proceed with

the development of the theory and assume that it is understood what is meant by «correctness», «classification», «relations of equivalence and subsumption», «full abstraction», and the definition of the class *K-RACi*.

5.3 Results

5.3.1 Boolean CCE (Full Event Calculus)

The following description consists in the original 1986 [103] Calculus of Events based on predicate completion [41, 42, 162], simplified in 1992 [102], extended in 1995 [179] to use filtered predicate circumscription [166] [180, ch. 16 and p. 81] [115, 128, 126], further extended in 1997 [180, ch. 16] to model actions over time periods, further simplified in 1999 [181, §1,3] and 2000 [182, p. 209], and thus referred to as Full Event Calculus. The description preserves the original decisions to use time points and events as a primary notion. The reasoning model truly is Kowalski's Calculus of Events, extended with Sandewall's filtering technique and McCarthy's predicate circumscription, then reduced to Clark's predicate completion via the Lifschitz reduction theorems to compute circumscription [115, 112].

Definition 5.3.1 (Boolean CCE) The model uses classical first-order logic as base logic, augmented with the formulae in fig. 5.1 (p. 105) and axioms in fig. 5.2 for representing the scenario of interest and for controlling deduction, it then uses predicate circumscription [128] with forced separation as model-preference criterion. The formal language is defined in fig. 5.1. Let Γ_1 be a finite conjunction of *Initiates*, *Terminates* and *Releases* formulae (the scenario). Let Γ_2 be a finite conjunction of $Initially_P$ and $Initially_N$ formulae (the initial situation) and of *Happens* and temporal ordering formulae (the narrative). Let Γ_3 be a finite conjunction of Uniqueness of Names Axioms for the actions and the features mentioned in Γ_1. Following Tarski's definition of logical consequence, the set of logical consequences is $\{\alpha : \Delta \wedge \Gamma \vDash \alpha\}$, where α is a finite conjunction of $(\neg)HoldsAt$ formulae, Δ is the conjunction of axioms $A1 \ldots A7$ in fig. 5.2 and Γ is the following formula, where CIRC is the circumscription of the given predicates:

$$\text{CIRC}(\Gamma_1; Initiates, Terminates, Releases) \wedge \text{CIRC}(\Gamma_2; Happens) \wedge \Gamma_3$$

The minimisation of *Happens* corresponds to the default assumption

that no unexpected events occur. The minimisation of *Initiates*, *Terminates* and *Releases* corresponds to the default assumption that actions have no unexpected effects. We interpret time points as natural numbers, with 0 as initial element. The structure of time is the classical structure $\mathbb{N}$ of natural numbers with a total order relation, i.e. linear (non branching) time.

Although the structure of time in Shanahan's own text is $\mathbb{R}^+$ [178] [181, p. 411], the ontological characteristic of continuous time is only useful to model continuous change, and thus the above restriction to $\mathbb{N}$ does not affect the classification.

The given description shows that the Calculus of Events is not merely a logic program: it is a full-blown preferential logic. Preferential logics [187, p. 74] [121] are the result of associating a preference relation on interpretations to any base logic with compositional model-theoretic semantics. In this case, the base logic consists in the Horn clause fragment of classical, Tarskian first-order logic. The conceptual basis of the adopted preference relation on interpretations is the partitioning of the set of premisses and the application of local preference relations to their interpretations; the set of preferred interpretations is chosen by filter preferential entailment, using predicate circumscription as preference relation. The definition combines filtering with *occlusion* [166], a technique to block temporal inertia for specified formulae at specified times.

We shall now classify the described model. Let the relation $(t, f, v) \in \Sigma(\mathrm{Y})$ be a shorthand for «(t, f, v) is true in $\Sigma(\mathrm{Y})$», that is, «it exists an intended interpretation (M, H) such that $(\mathcal{B}, M, H, \mathcal{P}, \mathcal{C}) \in \mathit{Mod}(\mathrm{Y})$ and $(f, v) \in H(t)$» according to the known definition of intended model set. Let the relation $(t, f, \mathit{true}) \in S(T(\mathrm{Y}))$ be a shorthand for $\Delta \wedge \Gamma \vDash \mathit{HoldsAt}(f, t)$. Let the relation $(t, f, \mathit{false}) \in S(T(\mathrm{Y}))$ be a shorthand for $\Delta \wedge \Gamma \vDash \neg \mathit{HoldsAt}(f, t)$, where Δ is the conjunction of axioms $A1 \ldots A7$, Γ is the formula

$$\textsc{circ}(\Gamma_1; \mathit{Initiates}, \mathit{Terminates}, \mathit{Releases}) \wedge \textsc{circ}(\Gamma_2; \mathit{Happens}) \wedge \Gamma_3$$

and, by definition 5.3.2, all formulae in Γ_1 and Γ_2 are in $T(\mathrm{Y})$.

Definition 5.3.2 Let L_M be the set of legal sentences in *K-IA*, and let L_O be the set of legal sentences of Boolean CCE. We define $T\colon L_M \to L_O$ as

follows.

- $T((0, f, true)) = \{Initially_P(f)\}$ and $T((0, f, false)) = \{Initially_N(f)\}$;
- $T((s, t, a)) = \{Happens(a, s, t-1)\}$ if s and t are temporal constants, and $T((s, t, a)) = \{\forall s, t.Happens(a, s, t-1)\}$ otherwise;
- $T(s < t) = \{s < t\}$ and $T(s \leq t) = \{s \leq t\}$;
- $T((s, t, a) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_{ij})$, where $n \in \{1, 2\}$, is translated into the following set of formulae. A single $Initiates(a, f, s)$ formula for each feature f becoming true as the effect of a deterministic action a. A single $Terminates(a, f, s)$ formula for each feature f becoming false as the effect of a deterministic action a. A single $Releases(a, f, s)$ formula for each feature f becoming either true or false as the effect of a nondeterministic action a. A single $HoldsAt(f, s)$ formula for each antecedent $(s, f, true)$ to the successful execution of the action a, and a single $\neg HoldsAt(f, s)$ formula for each antecedent $(s, f, false)$ to the successful execution of the action a. The antecedents are explicit conditions for the truth of *Initiates*, *Terminates* and *Releases* formulae.

The first rule maps the part OBS of Y into a Boolean CCE initial situation. The second and third rule map SCD and TC into a Boolean CCE narrative. The fourth rule maps LAW into a Boolean CCE scenario.

The reason for mapping t into $t - 1$ in the second rule is the following. For any successfully executed action a during $[t1, t2]$, the occlusion predicate prevents the value of influenced features from being seen during $(t1, t2)$ [169, p. 234,238], while the *Releases* predicate prevents the value of influenced features from being seen during $(t1, t2]$. In fact, the constraint $t2 < t$ in the CCE axioms $A2$ and $A5$ causes the action a to be neither true nor false at $t2$. The semantics by the metalanguage and the object language are then identical during $(t1, t2)$ but differ at $t2$; in the object language, the effect of a is exerted at any time $t > t2$, while in the metalanguage it is exerted at any time $t \geq t2$.—The alternative mapping would be the linear mapping $T((s, t, A)) = Happens(A, s, t)$, and the modification of axioms $A2$ and $A5$ by replacing the constraint $t2 < t$ with $t2 \leq t$. This would trigger the need for an additional modification, namely the substitution of $t1 \leq t2$

with $t1 < t2$ in axiom $A7$, to allow actions with non-null duration only, and avoid the dividing-instant problem. Please note that this would be a modification, not a correction. It is our aim to classify the original reasonin model, as given by its authors, and thus no modification is welcome.

Theorem 5.1 (Classification) For any reasoning scenario Y in $Ksp\text{-}IbA \subset K\text{-}RACi$ and for any element (t, f, v) of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \mathcal{V}$ the following relation holds true: $(t, f, v) \in S(T(\mathrm{Y}))$ iff $(t, f, v) \in \Sigma_{Ksp\text{-}IbA}(\mathrm{Y})$.

To prove this thesis we need two propositions by Lifschitz; we reproduce them as in Shanahan [180, p. 280].

Proposition 5.2 [115, prop. 3.1.1] $\textsc{circ}(\Gamma \wedge \forall \overline{x}.\rho(\overline{x}) \leftarrow \phi(\overline{x}); \rho)$ is equivalent to $\Gamma \wedge \forall \overline{x}.\rho(\overline{x}) \leftrightarrow \phi(\overline{x})$ if Γ and $\phi(\overline{x})$ do not mention the predicate ρ.

Proposition 5.3 [115, prop. 7.1.1] An occurrence of a predicate symbol in a formula ϕ is *positive* if it is in the scope of an even number of negations in the equivalent formula ψ that is obtained by eliminating the connectives $\rightarrow$ and $\leftrightarrow$ from ϕ. Let $\overline{\rho}$ be the n-tuple of predicate symbols $\rho_1, \ldots, \rho_n$. If all occurrences in Γ of the predicate symbols in $\overline{\rho}$ are positive, then $\textsc{circ}(\Gamma; \overline{\rho}) = \textsc{circ}(\Gamma; \rho_1) \wedge \ldots \wedge \textsc{circ}(\Gamma; \rho_n)$.

Proof *of theorem 5.1* The following standard reduction applies to Γ. By prop. 5.3, the second-order formula

$$\textsc{circ}(\Gamma_1; \mathit{Initiates}, \mathit{Terminates}, \mathit{Releases})$$

reduces to the following second-order formula:

$$\textsc{circ}(\Gamma_1; \mathit{Initiates}) \wedge \textsc{circ}(\Gamma_1; \mathit{Terminates}) \wedge \textsc{circ}(\Gamma_1; \mathit{Releases})$$

By prop. 5.2 each CIRC minimisation in both the above formula and in $\textsc{circ}(\Gamma_2; \mathit{Happens})$ reduces to first-order predicate completion. In what follows, this reduction is used at each evaluation of $S(T(\mathrm{Y}))$, and the reference to any CCE axiom involves the application of the Uniqueness of Names Axioms in Γ_3. The proof is by induction.

(i) The simulative game starts at time $\tau = 0$. The initial state of the environment is represented by elements $(0, f, \mathit{true})$ or $(0, f, \mathit{false})$ of OBS

in $\Sigma(\mathrm{Y})$. This results either in $HoldsAt(f,0) \in S(T(\mathrm{Y}))$ by axiom $A1$, or in $\neg HoldsAt(f,0) \in S(T(\mathrm{Y}))$ by axiom $A4$ respectively.

(ii) The environment, as a player, persists until the ego player communicates its intention to perform an action, so that no element of SCD starts at the present time τ. This results in temporal inertia, by either axiom $A1$ or $A4$ depending on how f was initialised, or by axiom $A2$ or $A5$ depending on how f was last modified.

(iii) At time point τ, the ego player suddenly adds (τ, E) to the current-action set $\mathcal{C}$. Then the environment executes the action, and ends it at τ' by removing (τ, E) from $\mathcal{C}$ and adding (τ, τ', E) to the past-action set $\mathcal{P}$. The ego may also decide to end E earlier than τ', let say at $\tau'' \in (\tau, \tau')$, so that it may autonomously remove (τ, E) from $\mathcal{C}$ and add (τ, τ'', E) to $\mathcal{P}$. Let show the corresponding logical consequences of Boolean CCE, pointwise. By definition 5.3.2, it exists either a single $Happens(E, \tau, \tau' - 1)$ or a single $Happens(E, \tau, \tau'' - 1)$ formula to refer to. **If** the feature f does not belong to the set of those features which would be modified by a successful execution of E, i.e. $f \notin Infl(E, \sigma_t)$, **then** the feature is neither *Clipped* nor *Declipped*, and the situation described at (ii) then holds up to τ' (τ''). Otherwise is $f \in Infl(E, \sigma_t)$, and the following holds. If at least one antecedent for the successful execution of the action E is not met, the action E is executed without any effect and the situation described at (ii) then holds up to τ' (τ''). Otherwise, **if** all antecedents for the action E are successfully met (i.e. all *HoldsAt* and $\neg$*HoldsAt* test conditions for *Initiates*, *Terminates* and *Releases* clauses are met by axioms $A3$ and $A6$), or no antecedent exists at all (in which case the above tests are immediately met), **then** E is successfully executed and the following holds.

- If $t = \tau$, then either of the following holds by temporal inertia:
  - $(t, f, true) \in \Sigma(\mathrm{Y})$,
    $Initially_P(f)$ by definition of $T(\mathrm{Y})$,
    $\neg Clipped(0, f, t)$ by axiom $A3$ and
    $HoldsAt(f, t) \in S(T(\mathrm{Y}))$ by axiom $A1$, or
  - $(t, f, false) \in \Sigma(\mathrm{Y})$,
    $Initially_N(f)$ by definition of $T(\mathrm{Y})$,
    $\neg Declipped(0, f, t)$ by axiom $A6$ and

$\neg HoldsAt(f,t) \in S(T(\mathrm{Y}))$ by axiom $A4$.

- If $\tau < t < \tau'$, it is $(f,v) \in Trajs(E,\sigma_t)$, $(t,f,v) \in \Sigma(\mathrm{Y})$ and either of the following holds:

  - $Initiates(a,f,\tau) \vee Releases(E,f,\tau)$ by definition of $T(\mathrm{Y})$ and $Declipped(\tau,f,\tau')$ by axiom $A6$, or
  - $Terminates(a,f,\tau) \vee Releases(E,f,\tau)$ by definition of $T(\mathrm{Y})$ and $Clipped(\tau,f,\tau')$ by axiom $A3$,

  so that it is neither $HoldsAt(f,t) \in S(T(\mathrm{Y}))$ by axiom $A2$ nor $\neg HoldsAt(f,t) \in S(T(\mathrm{Y}))$ by axiom $A5$ and $v = true \vee false$ (occlusion).
- If $t = \tau'$, it is $(f,v) \in Trajs(E,\sigma_t)$, $(t,f,v) \in \Sigma(\mathrm{Y})$ and one of the following holds:

  - $v = true$, then
    $Initiates(a,f,\tau)$ by definition of $T(\mathrm{Y})$ and
    $HoldsAt(f,\tau') \in S(T(\mathrm{Y}))$ by axiom $A2$, or
  - $v = false$, then
    $Terminates(a,f,\tau)$ by definition of $T(\mathrm{Y})$ and
    $\neg HoldsAt(f,\tau') \in S(T(\mathrm{Y}))$ by axiom $A5$, or
  - $v = true \vee false$, then
    $Releases(a,f,\tau)$ by definition of $T(\mathrm{Y})$, then it is both
    $Declipped(\tau,f,\tau')$ and $Clipped(\tau,f,\tau')$, so that it is neither
    $HoldsAt(f,t) \in S(T(\mathrm{Y}))$ by axiom $A2$, nor $\neg HoldsAt(f,t) \in S(T(\mathrm{Y}))$ by axiom $A5$ (nondeterminism).

The case for τ'' in place of τ' is identical to the above case.

(iv) The simulative game ranges to infinity, where the intended-model set is completely defined. The situations described at (ii) and (iii) repeat themselves to the infinity, for both semantics, the semantics mirroring the underlying semantics.

q.e.d.

The use of definition 5.3.1 for solving a few celebrated test examples was demonstrated by Shanahan [181] [180, p. 322-323]. The above result gives general insight into how this is done, and guarantees that the reasoning

model gives the correct answers for those specific scenarios, as well as for all other problems in the class *Ksp-IbA*.

Corollary 5.4 Boolean CCE is fully abstract with respect to the equivalence of reasoning scenarios in *Ksp-IbA*.

Proof True by proposition 2.1 at p. 40. q.e.d.

5.3.2 Continuous CCE (Extended Event Calculus)

The following description is a fragment of the Extended Event Calculus [181, p. 424] [180, §16.4]; it extends Boolean CCE to the case of unoccluded change for non-Boolean features.

Definition 5.3.3 (Continuous CCE) The language of Boolean CCE is extended with formulae of type $Initially_P(f2(v))$ to express the initial value v of a non-Boolean feature $f2$, and formulae of type $Trajectory(f1, t, f2(v), d)$ to express the unoccluded change of $f2$. The intended meaning of the latter formula is as follows: if the feature $f1$ is initiated at time t then the feature $f2$ has value v at time $t + d$. The logical machinery of Boolean CCE is extended accordingly, with the following axiom.

$$\begin{aligned} HoldsAt(f2, t3) \longleftarrow\ & Happens(a, t1, t2) \wedge Initiates(a, f1, t1) \wedge \\ & t2 < t3 \wedge t3 = t2 + d \wedge Trajectory(f1, t1, f2, d) \wedge \\ & \neg Clipped(t1, f1, t3) \end{aligned} \tag{A8}$$

Definition 5.3.4 Let L_M be the set of legal sentences of *K-RA*, and let L_O be the set of legal sentences of discrete CCE. The mapping $T\colon L_M \rightarrow L_O$ is identical to def. 5.3.2 for Boolean features and their respective actions. The following extension only applies to non-Boolean features and their respective actions.

- $T((0, f, v)) = \{Initially_P(f(v))\}$
- $T((s, t, a)) = \{Happens(a_{start}, s, s), Happens(a_{end}, t - 1, t - 1)\}$ if s and t are temporal constants, and
  $T((s, t, a)) = \{\forall t.Happens(a_{start}, t, t), \forall t.Happens(a_{end}, t - 1, t - 1)\}$ otherwise.

- $T(s < t) = \{s < t\}$ and $T(s \leq t) = \{s \leq t\}$
- $(s, t, a) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_{ij}$ is translated into the following set of formulae. For each pair (i, j), where $i = 1 \ldots n$, $j = 1 \ldots m$,
  $\forall t.Initiates(a_{start}, f1, t)$
  $\forall t.Terminates(a_{end}, f1, t)$
  $\forall t, v.Releases(a_{start}, f2(v), t)$
  $\forall t, v.Initiates(a_{end}, f2(v), t) \leftarrow HoldsAt(f2(v), t)$
  $\forall t, d.Trajectory(f1, t, f2(v), d) \leftarrow v = ...$

The first rule maps the part OBS of Y into a CCE initial situation, the second and third rule map SCD and TC into a narrative, and the fourth rule maps LAW into a CCE scenario. The actions a_{start} and a_{end} are introduced to initiate and terminate the Boolean feature $f1$, as required by definition 5.3.3.

We shall now assess the described model. Let the relation $(t, f, v) \in \Sigma(\mathrm{Y})$ be a shorthand for «(t, f, v) is true in $\Sigma(\mathrm{Y})$», that is, «it exists an intended interpretation (M, H) such that $(\mathcal{B}, M, H, \mathcal{P}, \mathcal{C}) \in Mod(\mathrm{Y})$ and $(f, v) \in H(t)$», according to the definition of intended model set for *K-RACi*. Let the relation $(t, f, v) \in S(T(\mathrm{Y}))$ be a shorthand for $\Delta \wedge \Gamma \vDash HoldsAt(f(v), t)$, where Δ is the conjunction of axioms $A1 \ldots A8$, Γ is the known conjunctive formula

$$\Gamma \equiv \textsc{circ}(\Gamma_1; Initiates, Terminates, Releases) \wedge \textsc{circ}(\Gamma_2; Happens) \wedge \Gamma_3$$

and, by definition 5.3.4, all formulae in Γ_1, Γ_2 and Γ_2 are in $T(\mathrm{Y})$.

Theorem 5.5 (Classification) For any reasoning scenario Y in the class *Ksp-RA* and for any element (t, f, v) of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \mathcal{V}$ the following relation holds true: $(t, f, v) \in S(T(\mathrm{Y}))$ iff $(t, f, v) \in \Sigma_{Ksp\text{-}RA}(\mathrm{Y})$.

Proof By theorem 5.1 the thesis is true for any Y in $Ksp\text{-}IbA \subset Ksp\text{-}IA \subset Ksp\text{-}RA$. We must prove the thesis for any Y in $Ksp\text{-}RA \setminus Ksp\text{-}IbA$. The proof is identical to the given one for Boolean CCE, except for the following. When reasoning about non-Boolean features, the axioms A4 and A5 always fail. By absurd, the meaning of a successful axiom A4 (A5) would be that the non-Boolean feature f does not have the value v at time 0 (t). As v is not a truth-value, however, this would cause an infinite computation (if v

ranges on the naturals, for example) and the computed answers would not be the intended ones. Concerning the unoccluded change, the following holds. If $\tau < t < \tau'$ (point 3b of the proof), then both $Happens(a_{start}, \tau, \tau)$ and $Initiates(a_{start}, f1, \tau)$ hold by definition 5.3.4; thus $Clipped(\tau, f1, t)$ does not hold, and $Trajectory(f1, \tau, f(v), d)$ computes the value v; thus $HoldsAt(f(v), t) \in S(T(\Upsilon))$ by axiom $A8$. If $t = \tau'$ (point 3c of the proof), then both $Happens(a_{start}, \tau, \tau)$ and $Initiates(a_{start}, f(), \tau)$ hold by definition of $T(\Upsilon)$, and the axiom $A3$ always fails; thus $HoldsAt(f(v), t) \in S(T(\Upsilon))$ by axiom $A2$. q.e.d.

Corollary 5.6 Continuous CCE is fully abstract with respect to the equivalence of reasoning scenarios in *Ksp-RA*.

Corollary 5.7 (Classification of Discrete CCE) When Continuous CCE is applied to discrete scenarios, we refer to it as Discrete CCE. For any reasoning scenario Υ in the class *Ksp-IA* and for any element (t, f, v) of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \mathcal{V}$ the following relation holds true: $(t, f, v) \in S(T(\Upsilon))$ iff $(t, f, v) \in \Sigma_{Ksp\text{-}IA}(\Upsilon)$.

5.3.3 Continuous CCE (redesigned)

The Continuous CCE does not include a standard representation of trajectory descriptors. This is due to the original mathematical error of specifying the value of a function as its parameter in the formula $Trajectory(f1, t, f2(v), d)$. We solve this problem using *partial fluents* as new trajectory descriptors. The partial fluents can be implemented by functions in any PROLOG system where such objects are formally allowed, and by the usual PROLOG gymnastics otherwise.

Definition 5.3.5 (Continuous CCE) We extend the language of Boolean CCE with formulae of type $Initially_P(f, v)$ to express the initial value of non-Boolean features, and formulae of type

$$\forall t, d. Trajectory(a, t, d, f, v) \leftarrow v = \varphi(t, t + d, t + d)$$

to express the unoccluded change, where φ is a partial fluent. The intended meaning of the latter formula is as follows. Let a be an action such that $Happens(a, t1, t2)$ holds. The feature f has value v at time $t + d$ only if the

Boolean feature f' was initiated at time $t1$ and it was not terminated or released between times $t1$ and t. The logical machinery of Boolean CCE is extended accordingly, with the following axiom.

$$\begin{aligned} HoldsAt(t,f,v) \longleftarrow\ & Happens(a,t1,t2) \wedge t1 < t2 \wedge \\ & t1 < t \wedge t =< t2 \wedge \\ & d = t - t1 \wedge Trajectory(a,t1,d,f,v). \end{aligned}$$

We described the resulting axiomatisation in figures 5.5–5.6 (p. 107). It is evident from it that axioms A4, A5, A6, and the formulae $Initially_N$, *Initiates*, *Terminates* and *Releases* are now redundant. We then redesigned the model, as described in figures 5.7–5.8 (p. 108).

Definition 5.3.6 Let L_M be the set of legal sentences of *K-RA*, and let L_O be the set of legal sentences of the redesigned Continuous CCE (p. 108). We define $T\colon L_M \to L_O$ as follows.

- $T((0,f,v)) = \{Initially(f,v)\}$;
- $T((s,t,a)) = \{Happens(a,s,t)\}$ if s and t are temporal constants, and $T((s,t,a)) = \{\forall s,t.Happens(a,s,t)\}$ otherwise;
- $(s,t,a) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_{ij}$, for each pair (i,j), where $i = 1 \ldots n$, $j = 1 \ldots m$, it is translated as $\forall t,d.Trajectory(a,s,d,f_{ij},v) \longleftarrow Precondition \wedge v = \varphi_{ij}(s,s+d,s+d)$, where *Precondition* consists in a single $HoldsAt(s,f,v)$ formula for each antecedent (s,f,v) to the successful execution of the action a.

The first rule maps the part OBS of Y into a CCE initial situation. The second and third rule map the SCD and LAW part of Y into a CCE scenario.

We shall now classify the described model. Let the relation $(t,f,v) \in \Sigma(\mathrm{Y})$ be the usual shorthand for «(t,f,v) is true in $\Sigma(\mathrm{Y})$», that is, «it exists an intended interpretation (M,H) such that $(\mathcal{B},M,H,\mathcal{P},\mathcal{C}) \in Mod(\mathrm{Y})$ and $(f,v) \in H(t)$», according to the definition of intended model set for *K-RACi*. Let the relation $(t,f,v) \in S(T(\mathrm{Y}))$ be a shorthand for $\Delta \wedge \Gamma \vDash HoldsAt(t,f,v)$, where Δ is the conjunction of axioms $A1 \ldots A4$, Γ is the conjunctive formula $\Gamma \equiv \mathrm{CIRC}(\Gamma_2; Happens) \wedge \Gamma_3$ and, by definitions

5.3.5, 5.3.6 , all formulae in Γ_2 and Γ_3 are in $T(\Upsilon)$.

Theorem 5.8 (Classification) For any reasoning scenario Υ in the class *Ksp-RA* and for any element (t, f, v) of $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times \mathcal{V}$ the following relation holds true: $(t, f, v) \in S(T(\Upsilon))$ iff $(t, f, v) \in \Sigma_{Ksp\text{-}RA}(\Upsilon)$.

Proof By proposition 5.2, CIRC$(\Gamma_2; Happens)$ reduces to first-order predicate completion. In what follows, this standard reduction is used at each evaluation of $S(T(\Upsilon))$, and the reference to any CCE axiom involves the application of the Uniqueness of Names Axioms in Γ_3. The proof is by induction.

(i) The simulative game starts at time $\tau = 0$. The initial state of the environment is represented by elements $(0, f, v)$ of OBS in Υ. This results in $HoldsAt(0, f, v) \in S(T(\Upsilon))$ by axiom $A1$.

(ii) The environment, as a player, persists until the ego player communicates its intention to perform an action, so that no element of SCD starts at the present time τ. This results in temporal inertia, by either axiom $A1$, depending on how f was initialised, or by axiom $A2$, depending on how f was last modified.

(iii) At time point τ, the ego player suddenly adds (τ, E) to the current-action set $\mathcal{C}$. Then the environment executes the action, and ends it at τ' by removing (τ, E) from $\mathcal{C}$ and adding (τ, τ', E) to the past-action set $\mathcal{P}$. The ego may also decide to end E earlier than t', let say at $\tau'' \in (\tau, \tau')$, so that it may autonomously remove (τ, E) from $\mathcal{C}$ and add (τ, τ'', E) to $\mathcal{P}$. Let show the corresponding logical consequences of the redesigned Continuous CCE, pointwise. By definition 5.3.6, it exists either a single $Happens(E, \tau, \tau')$ or a single $Happens(E, \tau, \tau'')$ formula to refer to. **If** the feature f does not belong to the set of those features which would be modified by a successful execution of E, i.e. $f \notin Infl(E, \sigma_t)$, **then** the feature is not *Clipped*, and the situation described at (ii) then holds up to τ' (τ''). Otherwise is $f \in Infl(E, \sigma_t)$, and the following holds. If *Precondition* fails, then the action E is executed without any effect and the situation described at (ii) then holds up to τ' (τ''). If *Precondition* succeeds, then E is successfully executed and the following holds.

- If $t = \tau$, then is $Initially(f, v)$ by definition of $T(\Upsilon)$, $\neg Clipped(0, f, t)$ by axiom $A4$, and $HoldsAt(t, f, v) \in S(T(\Upsilon))$ by axiom $A1$;

- If $\tau < t \leq \tau'$, then is $Clipped(\tau, f, \tau')$ by axiom $A4$, and $HoldsAt(t, f, v) \in S(T(\Upsilon))$ by axiom $A3$.

The case for τ'' in place of τ' is identical to the above case.

(iv) The simulative game ranges to infinity, where the intended-model set is completely defined. The situations described at (ii) and (iii) repeat themselves to the infinity, for both semantics, the semantics mirroring the underlying semantics.

q.e.d.

Corollary 5.9 The redesigned Continuous CCE is fully abstract with respect to the equivalence of reasoning scenarios in *Ksp-RA*.

5.3.4 Abductive CCE

Ksp-IA is the subclass of *K-IA* with the following characteristics: accurate and complete information about actions (*K*), complete knowledge about the initial state of the environment (*Ks*) and no information at any later state than the initial one (*Kp*), together with strict inertia in integer time (*I*) of actions with alternative results (*A*). In *Ksp-IA* are the problems of reasoning forwards in time, from causes to effects. In *K-IA* $\setminus$ *Ksp-IA* are the problems of reasoning backwards in time, from effects to causes, and problems of reasoning both forwards and backwards in time for a single query. The part OBS of any reasoning scenario in *K-IA* $\setminus$ *Ksp-IA* explicitly includes observations about features at strictly later states than the initial one. However, by definition, CCE can only represent information about the initial state of the environment, to reason forwards in time, and thus any scenario in *K-IA* $\setminus$ *Ksp-IA* falls beyond CCE's expressiveness and reasoning ability. In CCE's literature, reasoning backwards in time is understood as abductive reasoning. Abductive CCE was first studied in [177] [180, p. 330, 347-361], then defined as follows [182].

Definition 5.3.7 (Abductive CCE) Following definition 5.3.1, let α be the *goal* (ground). A *plan* for α is a narrative Γ' such that $\Delta \wedge \Gamma \vDash \alpha$, where $\Delta \wedge \Gamma$ is a ground and consistent set of premisses, and Γ is the following conjunctive formula:

$$\text{CIRC}(\Gamma_1; Initiates, Terminates, Releases) \wedge \text{CIRC}(\Gamma_2 \wedge \Gamma'; Happens) \wedge \Gamma_3.$$

As mentioned above, $K\text{-}IA \setminus Ksp\text{-}IA$ is partitioned in the set of problems requiring pure backward reasoning and the set of problems requiring both backward and forward reasoning for a single query. The following scenarios are not necessarily the most representative members of this latter set, yet they show that this latter set is not empty, and Abductive CCE does *not* represent and reason correctly about them.

Scenario 5.3.1 $\text{OBS} = \{(8, f, true)\}$, $\text{SCD} = \{(4, 6, a)\}$,
$\text{LAW} = \{(s, t, a) \Rrightarrow (s, f, true) \Rightarrow (t, g, true)\}$.

Scenario 5.3.2 $\text{OBS} = \{(15, g, false)\}$, $\text{SCD} = \{(4, 6, a), (8, 10, a2)\}$,
$\text{LAW} = \{(s, t, a) \Rrightarrow (s, f, true) \Rightarrow (t, g, true),$
$(s, t, a2) \Rrightarrow (s, f, true) \Rightarrow (t, g, false)\}$.

Scenario 5.3.3 Any variant of the above scenario where arbitrarily many actions are executed between a and $a2$ which are either independent from f or use f as precondition.

In these scenarios, the reasoning problem consists in deciding whether the action a was executed successfully. This is not a planning problem; we know that a has been executed, because it appears explicitly in the schedule. In the first scenario, the action a was successfully executed during $[4, 6]$; in fact, f holds both at 8 and at any previous time point, including the starting point 4. In the second scenario, a was executed successfully, for exactly the same reason. The only difference between the two scenarios is, that g is known to be true in the former, and is true as precondition of the action $a2$ in the latter ($a2$ is successfully executed, because g is false at $\tau = 15$). Abductive CCE fails with the scenarios due to both its object language and its reasoning.

Concerning the object language, Abductive CCE has the same expressiveness of ground Boolean CCE. In fact, definition 5.3.7 does not extend the expressiveness of definition 5.3.1; on the contrary, it constrains its expressiveness to the case of ground scenarios. If we translate the first scenario into the language of Abductive CCE, we obtain the following.

$$Happens(a, 4, 6)$$
$$Initiates(a, g, 4) \leftarrow HoldsAt(4, f)$$

The element in OBS. The element represents information about the environment at later states than the initial one. Abductive CCE has the same expressiveness of ground Boolean CCE, where such information can not be represented. This type of information could only be represented as part of α (the goal), but this representation of knowledge is not desirable, because it confuses the data with the query. We recall that this is not a planning problem; we know that a has been executed, because it occurs in the schedule.

Concerning the reasoning, we observe that Boolean CCE and Abductive CCE work independently from each other, thus requiring two distinct human-triggered runs for the same query. By definition of Boolean CCE, the action a is executed successfully only if the antecedent f is true by past information only, which turns out to be false by axioms $A1$ and $A4$. By Abductive CCE, the set $\Gamma' = \{Happens(4, 6, A)\}$ is a plan for α, but this does not help Boolean CCE in solving the problem. For the similar reasons, Boolean CCE also fails with the second and third scenarios.

5.3.5 Concurrent CCE

The class *Ksp-RACi* is the extension of *Ksp-RA* to concurrency of possibly independent actions. The part SCD of any reasoning scenario in *Ksp-RACi* $\setminus$ *Ksp-RA* explicitly includes scheduled actions at strictly overlapping time periods.

Concurrent CCE, or Extended Event Calculus is built upon Continuous CCE. The language of Continuous CCE is extended with formulae of type $Happens(a1\&a2, t1, t2)$, meaning that $a1\&a2$ is the compound action comprising the two actions $a1$ and $a2$ occurring during the time period $[t1, t2]$, with formulae of type $Cancels(a1, a2, b)$, meaning that «the occurrence of $a1$ cancels the effect of a simultaneous occurrence of $a2$ on feature b», and formulae of type $Cancelled(a, b, t1, t2)$, meaning that «some event occurs from time $t1$ to time $t2$ which cancels the effect of acion a on feature b». The logical machinery of Continuous CCE is extended accordingly, including the following:

$$Happens(a1\&a2, t1, t2) \longleftarrow Happens(a1, t1, t1) \wedge Happens(a2, t1, t2).$$

The following simple scenarios are not necessarily the most representative members of *Ksp-RACi* $\setminus$ *Ksp-RA*, yet they show that this latter set is

not empty. It is easily verified that Concurrent CCE does not reason correctly about them.

Scenario 5.3.4 OBS $= \{()\}$, SCD $= \{(4, 8, a1), (2, 6, a2)\}$, LAW $= \{...\}$.

Scenario 5.3.5 OBS $= \{()\}$, SCD $= \{(4, 8, a1), (6, 10, a2)\}$, LAW $= \{...\}$.

Let $a1$ compute, say, $mg * sin(30)$ and $a2$ compute $mg * cos(30)$. In both scenarios, $a1$ and $a2$ do not cancel each other, but blend into the vectorial sum mg, respectively during $[4, 6]$ and $[6, 8]$. The logical $\wedge$ is not sufficient to model this case correctly.

5.4 Discussion

In summary, we presented assessments for the *Calculus of Events* in its various Circumscriptive axiomatisations defined using the filtering technique (CCE). The available axiomatisations reduced to five models. Boolean CCE is correct with respect to *Ksp-IbA*. Continuous CCE, both original and redesigned with partial fluents, belongs to the class *Ksp-RA*. Discrete CCE belongs to *Ksp-IA*. Abductive CCE is not correctly applicable to *K-IA* $\setminus$ *Ksp-IA*. Concurrent CCE is not correctly applicable to *Ksp-RACi*.

The results show a general limitation of the CCE models to the family *Ksp* of epistemological characteristics, and no CCE model is correct for *K-RACi* $\setminus$ *Ksp-RA*. This suggests work towards a better design of Concurrent and Abductive CCE, and their integration into a single formal model.

Shanahan raises the following open question [183]: «knowing that a logic has sufficient expressive power to represent a given problem domain is no help when it comes to actually constructing such a representation». Our answer to Shanahan consists in a three step process: (1) to learn about the expressive power of a reasoning model, we classify the model according to the systematic paradigm; (2) to establish whether the given reasoning model solves a given reasoning problem, we establish whether the reasoning problem belongs to the class for which the reasoning model is provably correct; (3) to construct the representation of the problem, we apply the described synthesis technique, which consists in a corollary of the formal classification. By construction, if we supply the representation of the problem to the reasoning model, the reasoning model accepts this input as syntactically correct, and its output is formally correct. This holds

because the formal classification is also a formal correctness result of the reasoning model with respect to *any* reasoning problem in the class, and the synthesis technique is a corollary of the formal classification. This occurs within the systematic paradigm. By comparison, the method based upon «appeals to intuition» offers opinions on the expressive power of the logic, no synthesis technique for the reasoning problems, and no general evidence of correctness.

SPECIFICATION AND SYNTHESIS. If we read the metalanguage as a specification language, then Y is a formal specification for the corresponding CCE formalisation, for any scenario Y in the correctness class. For any such Y, the corresponding formalisation is $\Delta \wedge T(\mathrm{Y})$, where Δ is the conjunction of axioms $A1, A2, A3, \ldots$ and $T(\mathrm{Y})$ results from the mechanical application of the translator T to Y. Plain knowledge of the metalanguage, and the use of a compiler for T, allows anyone to use the calculus correctly. Preliminary knowledge of CCE is no longer a requirement for its correct use.

VERIFICATION. For any CCE formalism, not necessarily written using the above synthesis technique, the inverse mechanical application of T is an immediate technique to verify whether the formalisation corresponds to any Y in the target class, that is, to *decide* whether the given formalism fulfills the set of specified requirements.

The above technique to synthesise and verify the scenarios answers to the open question posed by Shanahan in [183, p. 142]. The whole work answers to the open problem posed by Sandewall in [172, p. 272].

Figure 5.1: The language of Boolean CCE

Formula	Meaning
t	time point (natural number)
f	feature
a	action
	What is true when (OBS):
$Initially_P(f)$	f holds true from time 0
$Initially_N(f)$	f does not hold from time 0
	What happens when (SCD):
$Happens(a, t1, t2)$	a starts at time $t1$ and ceases at time $t2$
	Temporal Constraints (TC):
$t1 < t2$, $t1 \leq t2$	standard order relations between natural numbers
	What actions do (LAW):
$Initiates(a, f, t)$	f starts to hold after action a at time t
$Terminates(a, f, t)$	f ceases to hold after action a at time t
$Releases(a, f, t)$	f is not subject to inertia after action a at time t
	Logical Machinery (table 5.2):
$HoldsAt(f, t)$	f holds at time t
$Clipped(t1, f, t2)$	f is terminated/released between times $t1$ and $t2$
$Declipped(t1, f, t2)$	f is initiated/released between times $t1$ and $t2$

Figure 5.2: The axioms of Boolean CCE

$$HoldsAt(f,t) \longleftarrow Initially_P(f) \wedge \neg Clipped(0,f,t) \quad \text{(A1)}$$

$$HoldsAt(f,t) \longleftarrow t2 < t \wedge Happens(a,t1,t2) \wedge Initiates(a,f,t1) \wedge \neg Clipped(t1,f,t) \quad \text{(A2)}$$

$$Clipped(t1,f,t4) \longleftrightarrow \exists\, a,t2,t3\,[\, t1 < t3 \wedge t2 < t4 \wedge Happens(a,t2,t3) \wedge [Terminates(a,f,t2) \vee Releases(a,f,t2)]] \quad \text{(A3)}$$

$$\neg HoldsAt(f,t) \longleftarrow Initially_N(f) \wedge \neg Declipped(0,f,t) \quad \text{(A4)}$$

$$\neg HoldsAt(f,t) \longleftarrow t2 < t \wedge Happens(a,t1,t2) \wedge Terminates(a,f,t1) \wedge \neg Declipped(t1,f,t) \quad \text{(A5)}$$

$$Declipped(t1,f,t4) \longleftrightarrow \exists\, a,t2,t3\,[\, t1 < t3 \wedge t2 < t4 \wedge Happens(a,t2,t3) \wedge [Initiates(a,f,t2) \vee Releases(a,f,t2)]] \quad \text{(A6)}$$

$$Happens(a,t1,t2) \longrightarrow t1 \leq t2 \quad \text{(A7)}$$

Figure 5.3: The language of Continuous CCE

Formula	Meaning
t	time point (natural number)
f	feature
a	action
	What is true when (OBS):
$Initially_P(f)$	f holds from time 0
$Initially_N(f)$	f does not hold from time 0
	What happens when (SCD):
$Happens(a, t1, t2)$	a starts at time $t1$ and ends at time $t2$
	What actions do (LAW):
$Initiates(a, f, t)$	f starts to hold after action a at time t
$Terminates(a, f, t)$	f ceases to hold after action a at time t
$Releases(a, f, t)$	f is not subject to inertia after action a at time t
$Trajectory(f', t, f, d)$	f starts to hold at time $t + d$ if f' is initiated at time t
	Temporal Constraints:
$t1 < t2$, $t1 \leq t2$	standard order relations between natural numbers
	Logical Machinery (table 5.4):
$HoldsAt(f, t)$	f holds at time t
$Clipped(t1, f, t2)$	f is terminated/released between times $t1$ and $t2$
$Declipped(t1, f, t2)$	f is initiated/released between times $t1$ and $t2$

Figure 5.4: The axioms of Continuous CCE

$$HoldsAt(f,t) \longleftarrow Initially_P(f) \wedge \neg Clipped(0,f,t) \quad \text{(A1)}$$

$$HoldsAt(f,t) \longleftarrow Happens(a,t1,t2) \wedge Initiates(a,f,t1) \wedge t2 < t \wedge \neg Clipped(t1,f,t) \quad \text{(A2)}$$

$$HoldsAt(f2,t3) \longleftarrow Happens(a,t1,t2) \wedge Initiates(a,f1,t1) \wedge t2 < t3 \wedge t3 = t2 + d \wedge Trajectory(f1,t1,f2,d) \wedge \neg Clipped(t1,f1,t3) \quad \text{(A8*)}$$

$$Clipped(t1,f,t4) \longleftrightarrow \exists\, a,t2,t3\, [\, Happens(a,t2,t3) \wedge t1 < t3 \wedge t2 < t4 \wedge [Terminates(a,f,t2) \vee Releases(a,f,t2)]] \quad \text{(A3)}$$

$$\neg HoldsAt(f,t) \longleftarrow Initially_N(f) \wedge \neg Declipped(0,f,t) \quad \text{(A4)}$$

$$\neg HoldsAt(f,t) \longleftarrow Happens(a,t1,t2) \wedge Terminates(a,f,t1) \wedge t2 < t \wedge \neg Declipped(t1,f,t) \quad \text{(A5)}$$

$$Declipped(t1,f,t4) \longleftrightarrow \exists\, a,t2,t3\, [\, Happens(a,t2,t3) \wedge t1 < t3 \wedge t2 < t4 \wedge [Initiates(a,f,t2) \vee Releases(a,f,t2)]] \quad \text{(A6)}$$

$$Happens(a,t1,t2) \longrightarrow t1 \leq t2 \quad \text{(A7)}$$

Figure 5.5: The language of Continuous CCE (first redesign)

Formula	Meaning
t	time point (natural number)
f	feature
a	action
	What is true when (OBS):
$Initially_P(f, v)$	f has value v from time 0
$Initially_N(f, v)$	f does not have value v from time 0
	What happens when (SCD):
$Happens(a, t1, t2)$	a starts at time $t1$ and ends at time $t2$
	What actions do (LAW):
$Initiates(a, (f, v), t)$	f starts to have value v after action a at time t
$Terminates(a, (f, v), t)$	f ceases to have value v after action a at time t
$Releases(a, (f, v), t)$	f is not subject to inertia after action a at time t
$Trajectory(f', t, (f, v), d)$	f starts to have value v at time $t + d$ if f' is initiated at time t
	Temporal Constraints:
$t1 < t2$, $t1 \leq t2$	standard order relations between natural numbers
	Logical Machinery (table 5.6):
$HoldsAt(t, f, v)$	f has value v at time t
$Clipped(t1, f, t2)$	f is terminated/released between times $t1$ and $t2$
$Declipped(t1, f, t2)$	f is initiated/released between times $t1$ and $t2$

Figure 5.6: The axioms of Continuous CCE (first redesign)

$$HoldsAt(t, f, v) \longleftarrow Initially_P(f, v) \wedge \neg Clipped(0, f, t) \quad \text{(A1)}$$

$$HoldsAt(t, f, v) \longleftarrow Happens(a, t1, t2) \wedge Initiates(a, (f, v), t1) \wedge t2 < t \wedge \neg Clipped(t1, f, t) \quad \text{(A2)}$$

$$HoldsAt(t, f, v) \longleftarrow Happens(a, t1, t2) \wedge t1 < t2 \wedge t1 < t \wedge t =< t2 \wedge d = t - t1 \wedge Trajectory(a, t1, d, f, v). \quad \text{(A8)}$$

$$Clipped(t1, f, t4) \longleftrightarrow \exists\, a, t2, t3\, [\, Happens(a, t2, t3) \wedge t1 < t3 \wedge t2 < t4 \wedge [Terminates(a, (f, v), t2) \vee Releases(a, (f, v), t2)]] \quad \text{(A3)}$$

$$\neg HoldsAt(t, f, v) \longleftarrow Initially_N(f, v) \wedge \neg Declipped(0, f, t) \quad \text{(A4)}$$

$$\neg HoldsAt(t, f, v) \longleftarrow Happens(a, t1, t2) \wedge Terminates(a, (f, v), t1) \wedge t2 < t \wedge \neg Declipped(t1, f, t) \quad \text{(A5)}$$

$$Declipped(t1, f, t4) \longleftrightarrow \exists\, a, t2, t3\, [\, Happens(a, t2, t3) \wedge t1 < t3 \wedge t2 < t4 \wedge [Initiates(a, (f, v), t2) \vee Releases(a, (f, v), t2)]] \quad \text{(A6)}$$

$$Happens(a, t1, t2) \longrightarrow t1 \leq t2 \quad \text{(A7)}$$

Figure 5.7: The language of Continuous CCE (redesigned)

Formula	**Meaning**
t	time point (natural number)
f	feature
a	action
	What is true when (OBS):
$Initially(f, v)$	f has value v at time 0
	What happens when (SCD):
$Happens(a, t1, t2)$	a starts at time $t1$ and ends at time $t2$
	What actions do (LAW):
$Trajectory(a, t, d, f, v)$	f has value v at time $t + d$
	Temporal Constraints:
$t1 < t2, t1 \leq t2$	standard order relations between natural numbers
	Logical Machinery (table 5.8):
$HoldsAt(t, f, v)$	f has value v at time t
$Clipped(s, f, t)$	f is influenced between times s and t

Figure 5.8: The axioms of Continuous CCE (redesigned)

$$HoldsAt(t, f, v) \longleftarrow Initially(f, v) \wedge \neg Clipped(0, f, t) \quad \text{(A1)}$$

$$HoldsAt(t, f, v) \longleftarrow Happens(a, t1, t2) \wedge t1 < t2 \wedge t2 < t \wedge \neg Clipped(t2, f, t) \wedge d = t2 - t1 \wedge Trajectory(a, t1, d, f, v). \quad \text{(A2)}$$

$$HoldsAt(t, f, v) \longleftarrow Happens(a, t1, t2) \wedge t1 < t2 \wedge t1 < t \wedge t =< t2 \wedge d = t - t1 \wedge Trajectory(a, t1, d, f, v). \quad \text{(A3)}$$

$$Clipped(s, f, t) \longleftarrow Happens(a, t1, t2) \wedge t1 < t2 \wedge s < t2 \wedge t1 < t \wedge d = t2 - t1 \wedge Trajectory(a, t1, d, f, \_). \quad \text{(A4)}$$

6

MODELS (PART IV)

6.1 Introduction

We present the *Calculus of Fluents* in its up-to-date description and assessment [26]. The aim of the assessment is threefold: to give the distilled essence of the model's though processes, with a detailed explanation of its elements and structure, to classify its epistemological and ontological characteristics, and compare it with former classified models.

The name *Calculus of Fluents* originates from our interest in computing partial fluents, being our extension of Sandewall's notion of fluent to continuous change. The name distinguishes our model from both the Calculus of Events [180, 103] and the Calculus of Situations [164, 134]. The difference is both methodological and conceptual. Our model follows the systematic paradigm, while both rival models are «constrained by appeals to intuition». Concerning the underlying conception, the Calculus of Events deals with histories and global states, using Clark's negation as failure, the Calculus of Situations deals with histories and local events, still using negation as failure, and the Calculus of Fluents deals with the punctual value of features traced over time, using a new algebraic semantics of our design. The rival models proved inadequate to represent and reason correctly about the class *K-RACi* of epistemological and ontological characteristics. By comparison, the *Calculus of Fluents* provably encompasses the target class.

6.2 Methods

We used the systematic paradigm, as precisely described in chapter 2. The assessment required extensions to the paradigm. We described these extensions in chapter 2. In the following sections we therefore proceed with the development of the theory and assume that it is understood what is meant by «correctness», «classification», «relations of equivalence and subsumption», «full abstraction», and the definition of the class *K-RACi*. Work-

ing knowledge of lattice theory is also required, jointly with the original work of classification for positive logic programming, here referred to as SAS (chapter 4).

6.3 Results

6.3.1 Description of the model

Let Y be a scenario and let (τ, f, v) be an element of $\mathcal{H}$. The decision problem we are about to address consists in assigning the intended truth-value to (τ, f, v) using the set of premisses Y. The problem is equivalent to asking whether Y, augmented with (τ, f, v) in its OBS part, is a consistent set. As more than one intended model is possible, we also want to identify that single set V of values in $\mathcal{V}$ such that, for any $v \in V$, (τ, f, v) is *true* in at least one intended model. The ultimate task is to conclude everything about the values of features at different points in time.

According to SAS, an interpretation for Y is any subset of $\mathcal{H}_Y$ and the model for Y is the least fixpoint of a continuous mapping $T_Y : \wp(\mathcal{H}_Y) \rightarrow \wp(\mathcal{H}_Y)$, where the power set $\wp(\mathcal{H}_Y)$ is a complete lattice under the partial order of set inclusion. The reasoning is model-theoretic; it consists of a step-by-step approach to temporal inertia, where the ordinals are used as time points and the upward inductive process is used as master-clock. The model is built by successive approximations, iterating from the origo of time points up to the limit ordinal. The decision problem consists in deciding whether the element (τ, f, v) belongs to the set $T_Y \uparrow \omega$. As shown by its classification, this model builds the preferred history of game developments for *K-IA* when the environment has a fixed strategy, leading to the proven restricted range of applicability.

We shall now design the converse non-simulative model. The decision problem is addressed immediately, by focusing the formalised reasoning on the set of all action alternatives that are strictly relevant to the feature of interest. The model relies heavily on the order relation of the temporal structure; time is assumed as *given*, and thus no master-clock is necessary.

The following is the bottom-up scheme underlying the overall top-down construction: (1) for every (t, f, v) in OBS, (t, f, v) is true in all models; (2) for every (t, f, v) that may be generated as the effect of at least one alternative of a causal law without antecedents, (t, f, v) is true in at least one model;

(3) for every (t, f, v) that may be generated as the effect of at least one alternative of a causal law for which every antecedent is satisfied, (t, f, v) is true in at least one model; (4) for every (t, f, v) that may be generated as the abductive effect of at least one alternative of a causal law for which every consequent is satisfied, (t, f, v) is true in at least one model; (5) for every (t, f, v) for which none of the above hold, (t, f, v) is true in all models. Points 3 and 4 are then weakened by allowing not every antecedent (consequent) satisfied, but a non empty subset of them, while others must correspond to unknown feature values. The scheme is an extension of the fixpoint scheme to backward and forward reasoning about actions with alternative results.

The converse top-down construction consists in (1) analysing the contribution of feature values at τ and at past time points with respect to τ, gathered under the name of *greatest lower observations* (glo) for f at τ; (2) analysing the contribution of feature values at future time points with respect to τ, gathered under the name of *least upper observations* (luo) for f at τ; (3) determining the solution to the decision problem (τ, f, v) as the *consistent union* of $glo(\tau, f)$ and $luo(\tau, f)$, which is, by definition, the non-simulative *History*. A number of complete lattices are associated to Y, being sub-lattices of $(\wp(\mathcal{I}); \sqsubseteq)$ and hence referred to as *causal chains*, so that temporal priorities determine the candidate answers $glo(\tau, f)$ and $luo(\tau, f)$.

Definition 6.3.1 (consistent union) We define the function $\overset{*}{\cup} : \wp(\mathcal{H}) \times \wp(\mathcal{H}) \to \wp(\mathcal{H})$ as follows. For any O_1 and O_2 non-empty sets, $O_1 \overset{*}{\cup} O_2$ is the set of all $\sigma_1 \cup \sigma_2 \in \wp(\mathcal{H})$ such that $\sigma_1 \in O_1$, $\sigma_2 \in O_2$ and $\sigma_1 \cup \sigma_2$ is consistent. If both O_1 and O_2 are empty sets, we impose $O_1 \overset{*}{\cup} O_2 = \mathcal{H}$.

For example, let the following be candidate partial-states:

$$\sigma_{11} = \{(\_, f, true), (\_, g, true)\} \quad \sigma_{12} = \{(\_, f, true), (\_, g, false)\}$$
$$\sigma_{21} = \{(\_, f, true)\} \quad \sigma_{22} = \{(\_, f, false)\}$$

If $O_1 = \{\sigma_{11}, \sigma_{12}\}$, $O_2 = \{\sigma_{21}, \sigma_{22}\}$, then $O_1 \overset{*}{\cup} O_2 = \{\sigma_{11} \cup \sigma_{21}, \sigma_{12} \cup \sigma_{21}\}$. In the last case of the definition, the reason for having $O_1 \overset{*}{\cup} O_2 = \mathcal{H}$ rather than the empty set, as one would otherwise expect, it is due to the underlying semantics itself, as shown by lemma 6.3 of the classification.

Definition 6.3.2 (non-simulative history) We define the function $History : \mathcal{T} \times \mathcal{F} \rightarrow \wp(\mathcal{V})$ as follows:

$$History(\tau, f) = \{v \in \mathcal{V} : (t, f, v) \in glo(\tau, f) \overset{*}{\cup} luo(\tau, f)\}.$$

For any scenario Y and element $(\tau, f, v) \in \mathcal{H}$, we say that (τ, f, v) is true if and only if $v \in History(\tau, f)$.

Note that *History* has its values in $\wp(\mathcal{V})$, while the function *history* of the underlying semantics has its values in $\mathcal{V}$. For a given feature f, it is our aim to collect all the values that are intended for the feature f at the given time point τ. For example, in the tossing-coin scenario, the underlying semantics generates two possible developments as the effect of tossing, namely 1 and 0 for head and cross respectively, so that if the tossing ends at τ, the function $history(\tau, face)$ generates the feature value 1 for a game simulation, and 0 for the other, and thus $History(\tau, face) = \{1, 0\}$. The following three important functions are strictly defined in terms of *History*, the third of them formally defining the model itself.

$$Fluent(f) = \{(t, v) \in \mathcal{T} \times \mathcal{V} : v \in History(t, f)\}$$

Given a scenario Y and a feature $f \in \mathcal{F}$, $Fluent(f)$ is the set of all values per f on flowing time, according to Y.

$$State(\tau) = \{(f, v) \in \mathcal{F} \times \mathcal{V} : v \in History(\tau, f)\}$$

Given a scenario Y and a time point $\tau \in \mathcal{T}$, $State(\tau)$ is the state of the environment at time point τ, according to Y.

$$Comp(\mathrm{Y}) = \{(t, f, v) \in \mathcal{H} : (f, v) \in State(t)\}$$

Given a scenario Y, the *Completion Set* of Y, written $Comp(\mathrm{Y})$, is the set of all relevant states of the environment on flowing time. This set defines the non-simulative algebraic semantics for Y (NAS).

To complete the description of the model, we shall now define the concept of relevant action alternative, when it is successfully executed, what is meant by causal chains, and thus define the greatest-lower and

least-upper sets of formal observations.

Definition 6.3.3 (relevant action alternative) Let A be an action scheduled for execution, that is, it exists $(\mathbf{s}, \mathbf{t}, A)$ in SCD for some temporal expressions $\mathbf{s}$ and $\mathbf{t}$. Let $(s, t, A) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_{ij}$ be the corresponding causal law in LAW, where $n \geq 1$ is the number of action alternatives, and let θ be a valuation such that $\theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\}$. An *instantiated alternative of the action A* (iaa) is the element

$$(s\theta, t\theta, \bigcup_{j=1}^{m} S_{ij}(\theta)) \in \mathcal{T} \times \mathcal{T} \times \wp(\mathcal{H})$$

For a given feature f we say that the iaa $(s\theta, t\theta, S_{ij}(\theta))$ is relevant to f if and only if exist τ and v such that $(\tau, f, v) \in (s\theta, t\theta, S_{ij}(\theta))$.

For a given feature and a relevant action alternative, we must decide whether its execution was successful, namely, whether the feature changes by its execution. Let first recall the approach adopted in SAS.

$$\begin{aligned} T_{\mathrm{Y}}(I) = \{(\tau, f, v) \in \mathcal{H}_{\mathrm{Y}} : \\ & (\tau, f, v) \in \text{OBS and } \tau = 0, \text{ or} \\ & \text{exists } (\mathbf{s}, \mathbf{t}, A) \in \text{SCD}, \\ & \text{exists } (s, t, A) \Rrightarrow \textstyle\bigwedge_{j=1}^{m} S_j \in \text{LAW and} \\ & \text{exists a valuation } \theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\} \\ & \text{such that:} \\ & (\tau, f, v) \in \mathit{Consequents}(s\theta, \tau, t\theta, S_j(\theta)) \\ & \text{and } \mathit{Antecedents}(s\theta, \tau, t\theta, S_j(\theta)) \subseteq I \} \end{aligned}$$

where

$$\begin{aligned} \mathit{Antecedents}(s, \tau, t, S) &= \{(s, f, \varphi(s, s, t)) \in S \ : \ s \sqsubseteq \tau \sqsubseteq t \} \\ \mathit{Consequents}(s, \tau, t, S) &= \{(\tau, f, \varphi(s, \tau, t)) \in S \ : \ s \sqsubset \tau \sqsubseteq t \} \end{aligned}$$

According to the definition of T_{Y}, the action A is successfully executed if and only if $\mathit{Antecedents}(s, \tau, t, S) \subseteq I$, reasoning from premisses to consequences. Differently from T_{Y}, our approach consists in the following three consistency checks.

$$\mathit{Sat\_Obs} : \mathcal{I} \to \{\mathit{true}, \mathit{false}\}$$

$Sat\_Obs((s,t,S))$ succeeds if and only if, for any $(\tau,f,v) \in$ OBS such that $s \sqsubseteq \tau \sqsubseteq t$, it is $(\tau,f,v) \in S$. The contribution of *Sat_Obs* is analogous to the intersection of $[\![\text{OBS}]\!]$ with $Min(\ll, [\![\text{LAW}[\text{SCD}]]\!])$ in filter preferential entailment.

$$Sat\_Pre : \mathcal{I} \rightarrow \{true, false\}$$

$Sat\_Pre((s,t,S))$ succeeds if and only if $Sat\_Obs((s,t,S))$ succeeds and the preconditions $(s,f,v) \in S$ are either members of $glo(s,f)$ or correspond to unknown feature values. An empty set of preconditions is allowed. We say that it *succeeds explicitly* if and only if at least one of its preconditions is a member of $glo(s,f)$.

$$Sat\_Post : \mathcal{I} \rightarrow \{true, false\}$$

$Sat\_Post((s,t,S))$ succeeds if and only if $Sat\_Obs((s,t,S))$ succeeds and the postconditions $(t,f,v) \in S$ are either members of $luo(t,f)$ or correspond to unknown feature values. An empty set of postconditions is not allowed. We say that it *succeeds explicitly* if and only if at least one of its postconditions is a member of $luo(t,f)$.

Causal Chains

We mentioned that SAS beats time with the master clock and builds a semantic model from the origo of time points up to τ. By comparison with SAS, we focus on the feature f of the query, using the order relation of the given temporal structure. We do so by collecting all relevant action alternatives that have been successfully executed before τ, because they have influenced f. This collection is the set of all *lower causal chain* (lcc) for f at τ. We also collect all relevant action alternatives that have been successfully executed after τ, because they have influenced f. This collection is the set of all *upper causal chain* (ucc) for f at τ.

Let $\mathcal{I} = \mathcal{T} \times \mathcal{T} \times \wp(\mathcal{H})$ be the domain of all relevant action alternatives that have been successfully executed. The order relation $\sqsubseteq$ associated with the basic time structure applies as follows on members of $\mathcal{I}$: $(s_1,t_1,S_1) \sqsubseteq (s_2,t_2,S_2)$ iff $s_1 \sqsubseteq s_2$, for any (s_1,t_1,S_1) and (s_2,t_2,S_2) in $\mathcal{I}$. The order relation $\sqsubseteq$ is a partial order on $\wp(\mathcal{I})$, because for any I_1 and I_2 in $\wp(\mathcal{I})$ it

is $I_1 \sqsubseteq I_2$ if and only if $i_1 \sqsubseteq i_2$ for any $i_1 \in I_1$ and $i_2 \in I_2$. Furthermore, $\wp(\mathcal{I})$ is a complete lattice under $\sqsubseteq$, because the least upper bound ($\sqcup$) of a collection of subsets of $\mathcal{I}$ is their minimum element and the greatest lower bound ($\sqcap$) is their maximum element ($(\mathcal{T}, \sqsubseteq)$ is a complete lattice). The function $Min\colon (\wp(\mathcal{I}); \sqsubseteq) \rightarrow (\wp(\mathcal{I}); \sqsubseteq)$ is a monotone (order-preserving) function, because $I_1 \sqsubseteq I_2 \Rightarrow Min(I_1) \sqsubseteq Min(I_2)$ for any $I_1, I_2 \in \wp(\mathcal{I})$. The function Min is also complete, since $Min(\sqcup I) = \sqcup(Min(I)) = Min(I) = \sqcup(I)$ for every directed subset I of $\wp(\mathcal{I})$. The similar property holds for the function $Max\colon (\wp(\mathcal{I}); \sqsubseteq) \rightarrow (\wp(\mathcal{I}); \sqsubseteq)$.

Definition 6.3.4 (lower causal chain)

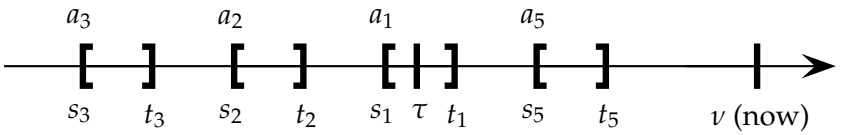
$$lcc : \mathcal{T} \times \mathcal{F} \rightarrow \wp(\mathcal{I})$$

Let $\mathcal{P}(\tau, f)$ be the set of all relevant action alternatives that have been successfully executed before τ, or that are being currently executed at τ. We compute this set as follows. For each relevant action alternative (s, t, S), occurring as mentioned, we check wether both $Sat\_Pre((s, t, S))$ and $Sat\_Post((s, t, S))$ succeed, ensuring that at least one of them succeeds explicitly. We thus define $lcc(\tau, f)$ as the set $Max(\mathcal{P}(\tau, f); \sqsubseteq)$ union the set of all members of $\mathcal{P}(\tau, f)$ that have not been overruled. We say that a member $(s_1, t_1, S_1) \in \mathcal{P}(\tau, f)$ is overruled if and only if it exists another member $(s_2, t_2, S_2) \in \mathcal{P}(\tau, f)$ such that: $t_1 \sqsubseteq s_2$, there is no other member between them, and either of the followings hold: **(a)** (s_2, t_2, S_2) has an empty set of preconditions, **(b)** both $(t_1, f, v) \in S_1$ and $(s_2, f, v) \in S_2$ hold, **(c)** f is in the postconditions of S_2.

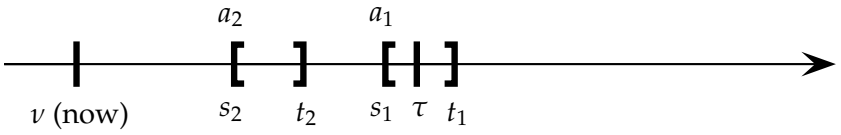
The definition of *upper causal chain* (ucc) is symmetrical to lcc. The only difference worth mentioning is that the upper chain does not include current activities, and thus the upper causal chain for f at τ consists of all and only those relevant action alternatives (s, t, S) that have successfully occurred after τ ($\tau \sqsubseteq s\theta$) and thus exert backward abductive influence on the feature of interest. To avoid infinite recursion, its consistency check must be limited to the postconditions.

Event Horizon

When describing scenarios in natural language, we use a temporal reference and verbs that acknowledge this temporal reference. The temporal reference may be an expression like «now» (often implicit), «yesterday», «next week», or «in 1620». For a given temporal reference to a past time (see τ in the diagram), we use the *past progressive tense* to denote an ongoing action a_1, the *past perfect tense* to denote an action a_2 completed before a_1, and the *pluperfect tense* to denote an action a_3 completed before a_2. When t_1 overlaps ν, we may denote a_1 using the *perfect progressive tense* or the rare *past perfect progressive tense*. The descriptive grammar of the English language [157] does not record any tense that can acknowledge the given τ to denote an action a_4 completed before a_3; to denote one such action, English speakers use a new temporal reference. Further, there is no record of a «future in the past» tense [109, p. 52]; to denote a future action a_5 with respect to τ, English speakers either use the uncommon *would* plus infinitive or '*was/were to*' plus infinitive, or switch temporal reference to the time of speaking ν and denote a_5 using the ordinary *past tense*. The used verbs define a time period (s_3, t_5), where s_3 is the time where a_3 starts and t_5 is the time where a_5 ends. We refer to this time period as the *past event horizon* for a given temporal reference τ.



The similar time period occurs when denoting actions with respect to a temporal reference $\tau > \nu$. For a given temporal reference to a future time (see τ in the second diagram), which may be implicit, we use the *future tense* to denote a 'predicted' action a_1, and the form *will* plus *perfect infinitive* to denote a 'past in the future' action a_2. Other future tenses have unprecise temporal references, and are therefore omitted. The used verbs define a time period (s_2, t_1), where s_2 is the time where a_2 starts and t_1 is the time where a_1 ends. We refer to this time period as the *future event horizon* for a given temporal reference τ.



For a given temporal reference, its event horizon limits the scope of reasoning, ignoring any action beyond the horizon. Although the act of changing temporal reference allows visibility of actions beyond the present event horizon, each temporal reference limits one's reasoning. This evidence suggests that human causal reasoning does not consist in broad inference through one large database.

Further evidence of an event horizon is given by the «positional analysis» in the game of chess. Positional analysis is the understanding of what is happening at a given spatial reference. The event horizon is still finite, but is much broader than the above, and each person may broaden their own through ad-hoc training and exercise. The most successful players do not waste time with linguistic descriptions, but envision whole strategies and position. The complete development of a game can always be reduced to a description in a natural or formal language.

In general, we use the *perfect aspect* to denote an action viewed as complete, and the *progressive aspect* to denote an action viewed as incomplete. According to Quirk et al. [157, p. 190], «approximately ten per cent of finite verb phrases are perfective». When using natural language, about 90% of Human causal reasoning is concerned about progressive reasoning, that is, ongoing actions. Therefore, albeit important, the event horizon concerns only the 10% of Human causal reasoning.

We model the event horizon as follows: we limit the definition of *lower causal chain* to its upper four members, and the definition of *upper causal chain* to its lower two members. Without this restriction, the horizon consists in the full lower and upper causal chains.

6.4 Greatest Lower and Least Upper Observations

Definition 6.4.1 (Greatest Lower Observations) We define the function $glo : \mathcal{T} \times \mathcal{F} \rightarrow \wp(\mathcal{H})$ as follows. Given a scenario Y, a time point τ and a feature f, $glo(\tau, f)$ is the union of the following three sets: the contribution by explicit observations $glo_{obs}(\tau, f)$ from the OBS part of the

scenario, the contribution by the current actions set $glo_{cas}(\tau, f)$ from the schedule, and the contribution by the past actions set $glo_{pas}(\tau, f)$ from the schedule. Formally, $glo_{cas}(\tau, f)$ is the set of all $(\tau, f, v) \in \mathcal{H}$ such that $(s, t, S) \in lcc(\tau, f)$, $s \sqsubset \tau \sqsubseteq t$ and $(\tau, f, v) \in S$. Similarly, $glo_{pas}(\tau, f)$ is the set of all $(t, f, v) \in \mathcal{H}$ such that $(s, t, S) \in lcc(\tau, f)$, $t \sqsubset \tau$ and $(t, f, v) \in S$. Finally, $glo_{obs}(\tau, f)$ is the set of all $(t, f, v) \in \mathcal{H}$ such that $(t, f, v) \in$ OBS, $t \sqsubseteq \tau$ and either of the following holds. **(a)** If $lcc(\tau, f) = \varnothing$ then no relevant action alternative is successfully executed during $[t, \tau]$, and thus $(t, f, v) \in glo_{obs}(\tau, f)$ by strict inertia. **(b)** If (a) failed, and all members $(b, e, S) \in lcc(\tau, f)$ occur before t ($e \leq t$), then $(t, f, v) \in glo_{obs}(\tau, f)$. **(c)** If (b) failed, and exists at least one member $(b, e, S) \in lcc(\tau, f)$ occurring after t ($t \leq b$) with an empty set of preconditions, then (b, e, S) is successfully executed in all semantic models, and thus $(t, f, v) \notin glo_{obs}(\tau, f)$. **(d)** If (c) failed, all members $(b, e, S) \in lcc(\tau, f)$ occurring after t ($t \leq b$) have preconditions. Let $\mathcal{D}$ be their set. If it exists $(b, e, S) \in \mathcal{D}$ such that $glo(b, g) \setminus \{(\_, g, w)\} = \varnothing$ for at least one of its preconditions $(b, g, w) \in S$, then (b, e, S) is successfully executed in all semantic models, and thus $(t, f, v) \notin glo_{obs}(\tau, f)$. **(e)** If (d) failed and it exists $(b, e, S) \in \mathcal{D}$ such that $\{(f, v) : (b, f, v) \in (S\} = \{(f, v) : (t, f, v) \in glo(b, f)\}$, then $(t, f, v) \notin glo_{obs}(\tau, f)$. It is $(t, f, v) \in glo_{obs}(\tau, f)$ otherwise.

Definition 6.4.2 (Least Upper Observations) Given a scenario Y, a time point τ and a feature f, $luo(\tau, f)$ is the union of the following two sets: the contribution by explicit observations $luo_{obs}(\tau, f)$ from the OBS part of the scenario, and the contribution by the future actions set $luo_{fas}(\tau, f)$ from the schedule. The definitions of luo_{obs} and luo_{fas} are strictly symmetrical to glo_{obs} and glo_{pas} respectively.

6.4.1 Classification of the model

Let the relation $(t, f, v) \in \Sigma(\mathrm{Y})$ be a shorthand for «(t, f, v) is true in $\Sigma(\mathrm{Y})$», that is, «it exists an interpretation (M, H) such that the development $(\mathcal{B}, M, H, \mathcal{A}, \mathcal{C})$ is in $Mod(\mathrm{Y})$ and $H(t, f) = v$». Further, let $\mathcal{F}_{\mathrm{Y}}$ be the set of all features occurring in Y, and let $\mathcal{F} \setminus \mathcal{F}_{\mathrm{Y}}$ be the set of all features f such that Y is empty of f.

Lemma 6.1 For any $\mathrm{Y} \in Ksp\text{-}IAd$ and $(t, f, v) \in \mathcal{T} \times \mathcal{F}_{\mathrm{Y}} \times \mathcal{V}$ the following

relation holds: $(t,f,v) \in \Sigma_{Ksp\text{-}IAd}\mathrm{Y} \Leftrightarrow (t,f,v) \in Comp(\mathrm{Y})$.

Proof The relation $(t,f,v) \in \Sigma_{Ksp\text{-}IAd}\mathrm{Y} \Leftrightarrow (t,f,v) \in T_\mathrm{Y} \uparrow \omega$ holds by proposition 4.8 at page 83. We must prove that $(t,f,v) \in T_\mathrm{Y} \uparrow \omega \Leftrightarrow (t,f,v) \in Comp(\mathrm{Y})$. Since Y is in *Ksp-IAd*, we assumed complete knowledge about the initial state of the environment. This knowledge is represented by observations $(0,f,v)$ in the OBS part of Y. By definition, it is $T_\mathrm{Y} \uparrow 0 = \{(t,f,v) \in \mathcal{B}_\mathrm{Y} : (t,f,v) \in \text{OBS} \wedge t = 0\}$ and $glo(0,f) = glo_{obs}(0,f) \cup glo_{pas}(0,f) \cup glo_{cas}(0,f) = \{(t,f,v) \in \text{OBS} : t = 0\} \cup \varnothing \cup \varnothing$. Then $(t,f,v) \in T_\mathrm{Y} \uparrow 0$ iff $(t,f,v) \in glo(0,f)$. Suppose $(t,f,v) \in T_\mathrm{Y} \uparrow (\tau-1)$ iff $(t,f,v) \in glo(\tau-1,f)$. We must prove that $(t,f,v) \in T_\mathrm{Y} \uparrow \tau$ iff $(t,f,v) \in glo(\tau,f)$. One of the following holds:

- SCD $= \varnothing$. Then is $T_\mathrm{Y} \uparrow \tau = T_\mathrm{Y} \uparrow (\tau-1) = \ldots = T_\mathrm{Y} \uparrow 0$. As $glo_{cas}(\tau,f) = \varnothing$, then is $glo(\tau,f) = glo(\tau-1,f)$.
- SCD $\neq \varnothing$. Let $(\mathbf{s},\mathbf{t},A)$ be one of its elements. If $Antecedents(\tau,A) \subseteq T_\mathrm{Y} \uparrow (\tau-1)$ then $T_\mathrm{Y} \uparrow \tau = T_\mathrm{Y} \uparrow (\tau-1) \cup Consequents(\tau,A)$; otherwise is $T_\mathrm{Y} \uparrow \tau = T_\mathrm{Y} \uparrow (\tau-1)$. Concerning NAS, it exists $(s,t,A) \Rrightarrow \bigwedge_{j=1}^{m} S_j$ in LAW and exists a valuation $\theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\}$ such that $f \in Influence(\tau,(s\theta,t\theta,S_j(\theta)))$ and $glo_{cas}(\tau,f_j) = \{(\tau,f,v) \in \mathcal{H} : (s,t,S) \in lcc(\tau,f), s \sqsubset \tau \sqsubseteq t \text{ and } (\tau,f,v) \in S\} = Consequents(\tau,A)$; otherwise is $glo(\tau,f_j) = glo(\tau-1,f_j)$ because $glo_{cas}(\tau,f_j) = \varnothing$.

In each of the above situations is $glo(\tau,f) \overset{*}{\cup} luo(\tau,f) = glo(\tau,f)$, in fact for any $(t',f,v') \in luo(\tau,f)$ such that $(t',f,v') \notin glo(\tau,f)$, then also $(t',f,v') \notin T_\mathrm{Y} \uparrow \tau$. q.e.d.

Lemma 6.2 For any $\mathrm{Y} \in Ksp\text{-}RA$ and $(\tau,f,v) \in \mathcal{T} \times \mathcal{F}_\mathrm{Y} \times \mathcal{V}$ the following relation holds: $(\tau,f,v) \in \Sigma_{Ksp\text{-}RA}\mathrm{Y} \Leftrightarrow (\tau,f,v) \in Comp(\mathrm{Y})$.

Proof By lemma 6.1, the relation holds for the sub-class *Ksp-IAd*. However, the SCD part of the scenario is a discrete sub-lattice of $(\mathcal{D},\sqsubseteq)$ in fact, by definition, every action has a non-empty length. As such, there are at most as many time points in SCD as the natural numbers. Then the relation holds also for that subset of the intended model set for scenarios in *Ksp-RAd* where strict inertia in continuous time and discrete deterministic change are allowed. To reach the full class *Ksp-RA* we must show that NAS accepts continuous change and alternative results of actions. Let (M,H) be

an intended model in $\Sigma(\mathrm{Y})$ such that exists $(\mathbf{s}, \mathbf{t}, A)$ in SCD, exists $(s, t, A) \Rrightarrow \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} S_j$ in LAW and exists a valuation $\theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\}$ such that $s\theta \sqsubset \tau \sqsubseteq t\theta$, $f = f_j$, $f_j \in Infl(A, H(s\theta))$, and $H(\tau, f_j) = \varphi_{ij}(s\theta, \tau, t\theta) = v$, that is, $(\tau, f_j, v) \in H \rhd S_{ij}(\theta)$. The following holds:

$$\begin{aligned} f_j \in Infl(A, H(s\theta)) \quad &\Leftrightarrow \quad f_j \in Influence(\tau, (s\theta, t\theta, S_{ij}(\theta))) \\ &\Rightarrow \quad (\tau, f_j, v) \in glo_{cas}(\tau, f_j) \\ &\Rightarrow \quad (\tau, f_j, v) \in glo(\tau, f_j) \end{aligned}$$

Therefore, by definition of consistent union, $v \in History(\tau, f_j)$, $(f_j, v) \in State(\tau)$ and $(t, f_j, v) \in Comp(\mathrm{Y})$. q.e.d.

Lemma 6.3 For any $\mathrm{Y} \in$ *K-RACi*, $\mathcal{F}_\mathrm{Y} = \varnothing$ implies $Comp(\mathrm{Y}) = \mathcal{H}$.

Proof The thesis is a straightforward corollary of the following statement: $(\tau, f, v) \in \Sigma_{K\text{-}RACi}\mathrm{Y} \Leftrightarrow (\tau, f, v) \in Comp(\mathrm{Y})$ for any $(\tau, f, v) \in \mathcal{T} \times \mathcal{F} \setminus \mathcal{F}_\mathrm{Y} \times \mathcal{V}$. Let first prove that for any $(\tau, f, v) \in \mathcal{T} \times \mathcal{F} \setminus \mathcal{F}_\mathrm{Y} \times \mathcal{V}$ is $(\tau, f, v) \in \Sigma_{K\text{-}RACi}\mathrm{Y}$. By definition, the game starts with the board in an initial configuration, where the initial state $H(0)$ is a certain nondeterministically given σ_0. Because of the nondeterminism, any value v could be assigned to f. During the game, the environment will persist in that value per f until an action is invoked by the ego that influences the feature. No such action will be invoked, because Y is empty of f by hypothesis. At the end of the game, the value per f will be the nondeterministically assigned initial value v. As no specific choice is made on this initial value, no element $(0, f, v)$ occurs in the OBS part of Y that may restrict the number of models, so that for any $(\tau, f, v) \in \mathcal{T} \times \mathcal{F} \setminus \mathcal{F}_\mathrm{Y} \times \mathcal{V}$ it is $(0, f, v) \in \Sigma_{K\text{-}RACi}\mathrm{Y}$ by point 4 of the definition of complete development set. On the other hand, for any $(\tau, f, v) \in \mathcal{T} \times \mathcal{F} \setminus \mathcal{F}_\mathrm{Y} \times \mathcal{V}$ is also $(\tau, f, v) \in Comp(\mathrm{Y})$. In fact, as $f \in \mathcal{F} \setminus \mathcal{F}_\mathrm{Y}$, both $glo(\tau, f)$ and $luo(\tau, f)$ are trivially empty, so that $(\tau, f, v) \in Comp(\mathrm{Y})$ via the definition of *History* and, in particular, via the function of consistent union. q.e.d.

Theorem 6.4 (Classification) The following relation holds true for any $\mathrm{Y} \in$ *K-RACi* and for any $(\tau, f, v) \in \mathcal{H}$: $(\tau, f, v) \in \Sigma_{K\text{-}RACi}\mathrm{Y} \Leftrightarrow (\tau, f, v) \in Comp(\mathrm{Y})$.

Proof The relation holds true for any (t, f, v) in $\mathcal{T} \times \mathcal{F} \setminus \mathcal{F}_{\mathrm{Y}} \times \mathcal{V}$ by lemma 6.3. We must prove the case for $f \in \mathcal{F}_{\mathrm{Y}} \neq \varnothing$. By lemma 6.2, via the *glo* function, the relation holds for any Y in *Ksp-RA*, that is, for pure prediction problems, and thus we must prove the case for any Y in *K-RACi* \ *Ksp-RA*, that is, for pure postdiction and prepostdiction problems. The former case is straightforward; the definition of *luo* is symmetrical to that of *glo*, so that lemma 6.2 also proves the relation for pure postdiction problems, with similar converse technique. Concerning prepostdiction problems, the following holds according to the definition of intended model set: it exists $(\mathbf{s}, \mathbf{t}, A) \in$ SCD and an elements (M, H) such that $M(\mathbf{s}) \sqsubset \tau \sqsubseteq M(\mathbf{t})$, $f \in Infl(A, H(M(\mathbf{s})))$, $H(\tau, f) = v$, and exists $(M(\mathbf{s}), g, w)$ precondition of A such that $H(0, g) = H(M(\mathbf{s}), g) = w$ and, since $\mathrm{Y} \cup \{(\tau, f, v)\}$ is a prepostdiction problem, Y is empty of g for any $t \in [0, M(\mathbf{s}))$. Thus, the problem of deciding whether (τ, f, v) belongs to $Comp(\mathrm{Y})$ reduces to the problem of deciding whether $(M(\mathbf{s}), g, w)$ belongs to $Comp(\mathrm{Y})$, which is a pure postdiction problem. q.e.d.

The full abstraction of this model with respect to the equivalence of causal reasoning scenarios is a default corollary of its classification chapter 4.

6.4.2 Computability

We consider the problem of computing the completion set for Y.

Proposition 6.5 For any $\mathrm{Y} \in$ *K-IA*, $Comp(\mathrm{Y})$ is recursively enumerable.

Proof By definition, $Comp(\mathrm{Y}) = \{(t, f, v) \in \mathcal{H} : (f, v) \in State(t)\}$, where $State(\tau)$ is the set of all elements $(f, v) \in \mathcal{F} \times \mathcal{V}$ such that $v \in History(\tau, f)$, for any $\tau \in \mathbb{N}$. The set $History(\tau, f)$ is finite; in fact, if $a(i)$ is the number of action alternatives of the action i, and m is the number of action occurrences in the part SCD of Y, then the number of relevant and successfully executed action alternatives for f at τ is at most $\sum_{i=1}^{m} a(i)$. Time points are recursively enumerable, being natural numbers, ans thus $Comp(\mathrm{Y})$ is a recursively enumerable set. q.e.d.

The above fact does not hold for scenarios in *K-RACi*, because their completion set is continuous. However, the problem of computing the *whole* completion set is unrealistic. Would we really ask for more information

that we could ever hope to use? The purpose of scenarios is to describe the environment as perceived by a human being. In such situations, resources are limited. Any scenario Y is therefore a finite scenario, that is, the OBS, SCD and LAW components of Y are finite sets, and thus, both the set $\mathcal{T}_Y$ of all time points occurring in Y and the set $\mathcal{F}_Y$ of all features occurring in Y are finite sets too. In any model for Y, all changes due to an action are limited to a finite period in time, because each action occurring in SCD has a finite duration. Since no other event is allowed in *K-RACi* besides those specified in SCD, the set of all time periods where persistence arise is a finite set too. The $History(\tau, f)$ consists of a finite set of feature values and $State_Y(\tau)$, namely $State(\tau)$ for those features occurring in Y, it is a finite set too. The problem of computing $Comp(Y)$ then reduces to computing a finite number of finite states:

$$Ker(Y) = \{(\tau, f, v) \in \mathcal{T}_Y \cup \{0\} \times \mathcal{F}_Y \times \mathcal{V} : (f, v) \in State_Y(\tau)\}$$

We shall refer to $Ker(Y)$ as the *kernel* of the completion set for Y, the most representative of all decidable subsets of $Comp(Y)$. The value of any feature at any given time point can always be computed via the *History* function, where *computable partial fluents* are used for characterising features during time periods where their values change. Computing the set of all time points for which a given feature f has a given value v requires standard numerical-analysis techniques over those periods where actions influencing f are performed:

$$Holds\_at(f, v) = \{\tau \in \mathcal{T} : v \in History(\tau, f)\}$$

Adopting the non-simulative algebraic semantics as proof procedure, we defined a meta-theoretic extension of the Horn Clause Logic, using the non-ground representation of the object-level variables. The meta-interpreter consists in the Horn clause representation of the proof procedure, it is built on top of the EPSILON system (ESPRIT project P-530) [45, 110, 48]), and is executable as a conventional logic program by the SLD-resolution rule. The resulting calculus is domain independent and, due to theorem 6.4, it has a known range of correct applicability.

6.5 Discussion

We presented the *Calculus of Fluents* and its classification. The model is correct with respect to the full class *K-RACi* of epistemological and ontological characteristics.

The model subsumes both Lifschitz's and Thielscher's variants of McCarthy's Calculus of Situations, and the Calculus of Events in its various circumscriptive axiomatisations. The Calculus of Fluents subsumes all classified models to date, spanning seventy-two years of research in the field (see page 63)

7

MODELS (PART V)

7.1 Introduction

In 1936 Turing modelled «a human calculator, provided with pencil and paper and explicit instructions» [209, 210, 39]. A man had to perceive and change the local environment by strict instructions. His observational behaviour [209, §9a] was «compared to a machine» with a finite number of internal states [209, §1]. The abstract machine had input and output actions, it could read and write, perceive and change its local environment. The class of computations by this machine was independent of the details of its definition: the same machine could be programmed to compute different functions. Turing named his model of a human calculator «the universal computing machine», he referred to it as a useful invention [209, p. 241] [211, p. 7], he claimed that it «can be used to compute any computable sequence» and commented that «all [supporting] arguments which can be given are bound to be, fundamentally, appeals to intuition, and for this reason rather unsatisfactory mathematically» [209, p. 249]. Turing's claim, or «Church-Turing thesis» [98, p. 232], it is still an unproven conjecture. Turing's abstract machine is still the core reference in the theory of computation. Turing's own interest in thinking machines [212], studies on what Turing machines cannot do [122, 123], and the challenge to explore farther than the computational limits of Turing machines, led to various computational models of causal reasoning [174, 180, 116]. An objective measure of the added value of those models with respect to Turing's original model is still unknown.

Our aim is to classify the epistemological and ontological characteristics of Turing's model, to measure the distance of classified models with respect to the target class.

7.2 Methods

We used the systematic paradigm, as precisely described in chapter 2. In the following sections we therefore proceed with the development of the theory and assume that it is understood what is meant by «correctness», «classification», «relations of equivalence and subsumption», «full abstraction», and the definition of the class $K\text{-}IA$.

7.3 Results

Definition 7.3.1 (machine) «We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions $q_1, q_2, \ldots, q_R$ which will be called **m-configurations**. The machine is supplied with a 'tape' (the analogue of paper) running through it, and divided into sections (called 'squares') each capable of bearing a 'symbol'. At any moment there is just one square, say the r-th, bearing the symbol $s(r)$ which is 'in the machine'. We may call this square the **scanned square**. The symbol on the scanned square may be called the **scanned symbol**. The scanned symbol is the only one of which the machine is, so to speak, 'directly aware'. However, by altering its m-configuration the machine can effectively remember some of the symbols which it has 'seen' (scanned) previously. The possible behaviour of the machine at any moment is determined by the m-configuration q_m and the scanned symbol $s(r)$. This pair $q_m, s(r)$ will be called the **configuration**: thus the configuration determines the possible behaviour of the machine.» [209, p. 231]—«The [operations] of the machine [are] described [as follows:] 'R' means 'the machine moves so that it scans the square immediately on the right of the one it was scanning previously', 'L' means 'the machine moves so that it scans the square immediately on the left of the one it was scanning previously', 'E' means 'the scanned symbol is erased', 'P[x]' means 'prints' [the symbol x].» [209, p. 233]—«In addition to any of these operations the m-configuration may be changed. Some of the symbols written down will form the sequence [which] is being computed. The others are just rough notes to 'assist the memory'. It will only be these rough notes which will be liable to erasure.» [209, p. 231]—«If the machine is supplied with a blank tape and set in motion, starting from the correct initial m-configuration, the subsequence of the symbols printed by it which are of the

first kind will be called the *sequence computed by the machine*. At any stage of the motion of the machine, the number of the scanned square, the complete sequence of all symbols on the tape, and the m-configuration will be said to describe the *complete configuration* at that stage. The changes of the machine and tape between successive complete configurations will be called the **moves of the machine**.» [209, p. 232-3]—«A computable sequence γ is determined by a description of a machine which computes γ. [...] The initial m-configuration is always to be called q_1. [...] The lines of the table are now of form [**(configuration, behaviour)**, where the configuration is the known pair and the **behaviour** is the pair (operations, final m-configuration)]. [...] Let us write down all expressions so formed from the table for the machine and separate them by semi-colons. In this way we obtain a **complete description of the machine**.» [209, p. 239-240]

Definition 7.3.2 (automatic machine) «If at each stage the motion of a machine (def. 7.3.1) is *completely* determined by the configuration, we shall call the machine an automatic machine.» [209, p. 232]

Definition 7.3.3 (choice machine) We shall call "choice machine" any machine (def. 7.3.1) «whose motion is only partially determined by the configuration». «When such a machine reaches one of these ambiguous configurations, it cannot go on until some arbitrary choice has been made by an external operator.» [209, p. 232]

Definition 7.3.4 (computing machine) «If an automatic machine prints only two kinds of symbols, of which the first kind consists entirely of 0 and 1, then the machine will be called a computing machine.» [209, p. 232]

We shall now describe the Turing machine in the meta-language.

Definition 7.3.5 Let L_O be the set of instructions in a complete description of Turing's machine, and let L_M be the set of legal sentences in K-IA. We define $T\colon L_O \to L_M$ as follows. The machine is only able to read one square at the time, and thus the tape belongs to the environment and the observable features are the scanned square, the symbol in it, and the machine's own internal state. This leads to one fluent for each feature, whose definition shall correspond to Turing's *complete description of the machine*.

Let us call q_0 the initial m-configuration, instead of q_1. The first m-configuration maps to $(0, state, q_0) \in OBS$. The first scanned square maps to $(0, square, r) \in OBS$. The first scanned symbol maps to $(0, symbol, s_0) \in OBS$. The complete description of the machine, namely its set of rows $(configuration, behaviour) = (q_m, s(r); o_n, q_n)$, it maps to $(\square, move) \in SCD$ and

$$(s, t, move) \Rrightarrow \bigwedge_{j=1}^{3} \forall\, \tau \in [s, t] \subset \mathcal{T}\,.\,[\tau] f_j = \varphi_j(s, \tau, t) \in LAW,$$

with features $f_1 = state$, $f_2 = square$ and $f_3 = symbol$. The corresponding partial fluents are defined as follows: $o_n = R$ maps to $\varphi_2(s, \tau, t) = [s]square + 1$; $o_n = L$ maps to $\varphi_2(s, \tau, t) = [s]square - 1$; $o_n = E$ maps to $\varphi_3(s, \tau, t) = blank$; $o_n = Px$ maps to $\varphi_3(s, \tau, t) = x$; q_n maps to $\varphi_1(s, \tau, t) = q_n$.

Theorem 7.1 (Classification) For any complete description Γ of automatic machine, $(r_i, tape_i, q_i)$ describes its complete configuration at stage i if and only if $(i, square, r_i)$, $(i, symbol, s(r_i))$, $(i, state, q_i) \in \Sigma_{Ksp\text{-}IAd}(T(\Gamma))$.

Proof We refer to the initial complete configuration $S_0 = (r_0, tape_o, q_0)$ as *input*, and to the final complete configuration S_k as *output*. The generic move of the machine is the transition $S_i \to S_{i+1}$ determined by the configuration $q_i, s(r_i)$.

We show the thesis true for S_0. The initial configuration of the automatic machine is completely specified by the pair $q_0, s(r_0)$. The input is S_0, and thus the machine is only reading the scanned symbol from the scanned square. By def. 7.3.5, it is $(0, square, r_0) \in OBS$, $(0, symbol, s(r_0)) \in OBS$, $(0, state, q_0) \in OBS$, and thus, by definition of $K\text{-}IA$, it is $(0, square, r_0) \in \Sigma_{Ksp\text{-}IAd}(\mathrm{Y})$, $(0, symbol, s(r_0)) \in \Sigma_{Ksp\text{-}IAd}(\mathrm{Y})$ and $(0, state, q_0) \in \Sigma_{Ksp\text{-}IAd}(\mathrm{Y})$.

We show the thesis true for S_1. The configuration of the automatic machine is completely specified by the pair $q_0, s(r_0)$, and the resulting behaviour is completely specified by the pair o_1, q_1. The operation o_1 can be either of the followings: R, L, E, P. By definition of $K\text{-}IA$ and def. 7.3.5, the following holds: if $o_1 = R$, then $[1]square = [0]square + 1$ and thus $(1, square, s_0 + 1) \in \Sigma_{Ksp\text{-}IAd}(\mathrm{Y})$; if $o_1 = L$, then $[1]square =$

$[0]square - 1$ and thus $(1, square, s_0 - 1) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$; if $o_1 = E$, then $[1]symbol = blank$ and thus $(1, symbol, blank) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$; if $o_1 = Px$, then $[1]symbol = x$ and thus $(1, symbol, x) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$. It is also true that $(1, state, q_1) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$.

We assume the thesis true for S_i. We shall now prove the thesis for S_{i+1}. The $i+1$-th configuration of the automatic machine is completely specified by the pair $q_i, s(r_i)$, and the resulting behaviour is completely specified by the pair o_{i+1}, q_{i+1}. The operation o_{i+1} can be either of the followings: R, L, E, P. By definition of $K\text{-}IA$ and def. 7.3.5, the following holds: if $o_{i+1} = R$, then $[i+1]square := [i]square + 1$ and thus $(i+1, square, s_i + 1) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$; if $o_{i+1} = L$, then $[i+1]square := [i]square - 1$ and thus $(i+1, square, s_i - 1) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$; if $o_{i+1} = E$, then $[i+1]symbol = blank$ and thus $(i+1, symbol, blank) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$; if $o_{i+1} = Px$, then $[i+1]symbol = x$ and thus $(i+1, symbol, x) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$. It is also true that $(i+1, state, q_{i+1}) \in \Sigma_{Ksp\text{-}IAd}(\Upsilon)$.

At any stage of the motion of the machine, the game univocally responds with the correct observables. q.e.d.

As corollary of the above, Turing's computing machines belong to the class $Ksp\text{-}IbAd$, and Turing's choice machines belong to the class $Ksp\text{-}IA$.

7.4 Discussion

In summary, we presented assessments for Turing's models of causal reasoning. Turing's computing machines belong to the class $Ksp\text{-}IbAd$, Turing's automatic machines belong to the class $Ksp\text{-}IAd$ and Turing's choice machines belong to the class $Ksp\text{-}IA$.

By comparison with former results (see the table of results at page 63), we learn that Turing's computing machines are subsumed by the Circumscriptive Boolean Calculus of Events, Turing's automatic machines are equivalent to Positive Logic Programming (SAS), and Turing's choice machines are equivalent to the Circumscriptive Discrete Calculus of Events.

Following the systematic paradigm, the scientific study of human causal reasoning shows a hierarchy of equivalent or subsuming models. In this framework, subsumption has no implications for Computability Theory. We then ask the following questions: *Is there a subsumed model which can simulate its subsuming model? Is there a model which can simulate any other model in*

the present taxonomy? The answer to these questions is positive. The model that subsumes any other model in the taxonomy, namely Brandano's Calculus of Fluents, admits implementation in Positive Logic Programming without red cuts, whose fixpoint semantics is equivalent to Turing's Automatic Machines. Therefore, the class $Ksp\text{-}IAd$ is sufficiently broad to include models of any other class in the present taxonomy.

A

NOTES AND COMMENTS

a.1 On McCarthy's theory and its variants

As a general policy, we call theories and their variants after their original authors, and then seek for their binomial nomenclature after the formal classification of their range of correct applicability. Therefore, we refer to the «situation calculus» [124, 134, 133, 131] as *McCarthy's theory*, and we refer to the collaborative work in Toronto and Roma [164] as *Reiter's variant of McCarthy's theory*, to acknowledge its chief promoter and researcher. Other variants of McCarthy's theory are available, such as Thielscher's [205, 206]. According to the *Handbook* [174, p. 444], «the situation calculus is basically just a notation. Different authors introduce and use different axiomatizations for it, reflecting not only technical differences in their logics but also different notions of what a situation is». A direct classification of McCarthy's theory and its variants is still unknown.

The words «situation» and «fluent» are often used by different authors with different meanings or synonyms. According to McCarthy [130], «a situation is in principle a snapshot of the environment at an instant». *The theory of computation commonly refers to such objects as states.* «One never knows a situation; one only knows facts about a situation. Events occur in situations and give rise to new situations. There are many variants of situation calculus, and none of them has come to dominate. [...] In situation calculus, the formula $s' = result(e, s)$ gives the new situation s' that results when the event e occurs in situation s. [...] Continuous processes can be treated in the situation calculus, but the theory is so far less successful than in discrete cases.» Statements in McCarthy's theory are LISP-like statements of type $result(e_t, \ldots, result(e_2, result(e_1, result(e_0, s_0))))$. For a given current state s_i and event e_j, the state $s_{i+1} = result(e_j, s_i)$ is the effect of the single event e_j in the state s_i. The overall effect is defined as the effect of a single initial event in the initial state followed by the effect of a new com-

putation starting in the resulting new state. *One writes these statements from right to left, from the initial state to the final state, to model discrete instantaneous change using explicit states and implicit discrete time.* According to McCarthy, fluents are «functions of situations in situation calculus» [130]. The above function $result\colon E \times S \to S$ is a «situational fluent» [134]. *Situations are states, and thus McCarthy's situational fluents are state transitions.* Albeit originally motivated by the limitations of Turing machines, and described in terms of Tarskian logic and LISP-like statements, McCarthy's theory looks very much like a **deterministic one-way finite state machine**.

In McCarthy's own words, «the situation calculus is the most studied formalism for *doing* causal reasoning» [130]. We observe that McCarthy does not refer to the situation calculus as a model of human reasoning, either supported by scientific evidence or pending scientific validation; he refers to it as a formalism for doing something, that is to say a useful invention. However, there is no consent to its usefulness. According to Davis, for example, «Reiter extended and popularized the situation calculus, which prior to his work had been widely considered to be too inexpressive to be useful. [Reiter's] recent book detailing this exploration, *Knowledge in Action* [164], is an exemplar of the best work of foundationalist researchers: rigorous yet accessible; formal yet grounded in application» [47, p. 10].

As far as we can see, Reiter's variant of McCarthy's formalism consists in using Peano's arithmetics as **basic execution mechanism** of the above finite state machine, and Kowalski's notation for the Calculus of Events as formal notation [164]. Reiter does not refer to the resulting theory as a model of human reasoning; he refers to it as the specification for GOLOG, which is a PROLOG program for controlling academic robots in Toronto. Therefore, when reading Reiter's book we expected as main result the proof of soundness and completeness of GOLOG with respect to its specification. However, this proof is missing [164, p. 98-100, §6.3.3]. We also observe the absence of supporting evidence of the following claim: «GOLOG appears to offer significant advantages over current tools for applications in dynamic domains» [164, p. 119]. In what respects the logic program GOLOG is better than other similar programs? How did Reiter measure the claimed advantage? Did he describe the method, and the results, so that we can verify his findings? The answers are all negative, and thus we fear that the use of the

verb «appears» was truly meant to appeal to intuition, necessarily limited in scope to the direct experience of local researchers, and implicitly reject any demand for a scientific methodology by external observers.

The original aim of inventing computing machineries that display human-level intelligence is necessarily bound to the scientific study of human thought processes. However, McCarthy's theory is not presented as a scientific model of human causal reasoning. If the intended aim of both Reiter and McCarthy was to model human causal reasoning, to use the scientific results in electro-mechanical engineering (Cognitive Robotics), in what respects the situation calculus can be considered a good scientific theory? The works of both McCarthy and Reiter fail to answer to this question; the authors appeal to their own intuition, but make no effort to structure their work, dividing scientific from engineering research, and using an objective measure of progress when presenting and comparing their work with the work of other researchers in the field. Davis is, therefore, in plain error in his quoted review; in the past four hundred years, the best work of foundationalist researchers has always been based on raw facts about the natural world, and never on appeals to intuition about (unsupported claims of) useful inventions.

One could approach the above problems as follows: (1) sustain the claim that McCarthy's theory is a scientific model of human causal reasoning, (2) classify this theory, (3) compare it with other previously classified models of causal reasoning, and then (4) prove the soundness and completeness of the logic program GOLOG with respect to this theory as formal specification.

On the first of the above mentioned points, we observe that we have neither natural nor experimental evidence of people who reason about their environment using states as primary objects, and instantaneous state transitions. We have evidence of the contrary, however. As reported by Quirk et al. [157], about 90% of natural language denotations of causal reasoning are about non-perfect tenses. People like to describe life while it happens, using pointwise denotations and progressive tenses. Human reasoning, as represented in natural language, is chiefly concerned about continuous change over time periods. When modelling human reasoning, we cannot possibly ignore these facts. Further, states as primitive objects are a needless computational burden, because their database must stay up-

-to-date with the reasoning, whose core duty, however, is not to waste time and space in managing states, but to answer quickly to queries and be aware of the physical environment. Both McCarthy's formalism and Reiter's variant fail to address this problem.

a.2 On Allen's theory

Allen's theory [5, 6] does not correspond to statistical evidence of human causal reasoning by Quirk et al. As reported by Quirk et al. [157], time points do occur in natural language denotations of causal reasoning. Furthermore, about 90% of such denotations use non-perfect tenses, that is, humans like to describe life while it happens, using pointwise denotations and progressive tenses. Allen's theory is strictly based on time periods as primitive objects, that is, the theory is unable to represent time points. Allen's theory is also unable to model progressive tenses, and continuous change in particular.

a.3 On Kuipers and Shults, «Reasoning in Logic about Continuous Systems», in J. Doyle, E. Sandewall and P. Torasso eds., Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, 1994

The report describes a theory of common sense reasoning about continuous behaviour in the physical environment, proves its soundness with respect to the QSIM simulator of qualitative physics, and supports the resulting theory by discussing its applications to control theory. The described theory consists in the adaptation of Emerson's logic [61] «to work with» QSIM [106]. We raise the following questions.

On QSIM. Do we have evidence of correctness and termination of the QSIM algorithm? If it runs continuously, does it generate correct answers along the way, so that we can use them? What is the advantage of QSIM with respect to, say, computing a set of differential equations to solve a Cauchy problem? For any given logic, can we use QSIM as semantic reference (virtual environment) with no need to adapt the logic to QSIM's formal environment? The question is especially relevant when dealing with a society of heterogeneous agents. Further, are QSIM's answers generated with the exact timing that we get from the physical environment? Finally, how

broad is its range of correct simulation?

On the theory. What are the epistemological and ontological characteristics of Emerson's adapted logic? How it compares with previously classified models of causal reasoning?

BIBLIOGRAPHY

The entries are sorted by author. The abbreviations of type v(i):p identify the volume, issue, and page number in a journal series. The number in round brackets at the end of each entry identifies the page in this monograph where the work in question is mentioned.

[1] *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Washington, D.C., 4th edition, 1994. (33)

[2] *Peer Review: An assessment of recent developments*. The Royal Society, London, 1995. (17)

[3] *Declaring Independence*. S.P.A.R.C., 21 Dupont Circle, Suite 800, Washington, DC 20036 USA, 2001. (6)

[4] *Peer Review*, POSTNOTE 182, The Parliamentary Office of Science and Technology, 7 Millbank, London SW1P 3JA, Sept. 2002. (17, 27)

[5] J. F. Allen, *Maintaining Knowledge about Temporal Intervals*, Communications of the ACM, 26(11):832–843, 1983. (52, 133)

[6] J. F. Allen and G. Ferguson, *Actions and Events in Interval Temporal Logic*, Journal of Logic and Computation, 4(5):531–579, 1994. (133)

[7] Aristotle, *Physics*, Oxford University Press, 1936. (51)

[8] Aristotle, *De Generatione Animalium*, Oxford University Press, 1949. (29, 30, 57)

[9] Aristotle, *Prior Analytics*, Oxford University Press, 1949. (30)

[10] Aristotle, *The Nicomachean Ethics*, Oxford University Press, 1951. (30)

[11] Aristotle, *Metaphysics*, Oxford University Press, 1958. (31)

[12] Aristotle, *Topica et Sophistici Elenchi*, Oxford University Press, 1958. (30)

[13] Aristotle, *Analytica Posteriora*, Oxford University Press, 1964. (31)

[14] G. Attardi and M. Simi, *Metalanguage and Reasoning across Viewpoints*, in Artificial Intelligence, Proceedings of the European Conference, 1986. (76)

[15] G. Attardi and M. Simi, *Proofs in Context*, in Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, 1994. (76)

[16] G. Attardi and M. Simi, *A Formalisation of Viewpoints*, Fundamenta Informaticae, 23(3), 1995. (76)

[17] G. Attardi and M. Simi, *Communication across Viewpoints*, Journal of Logic, Language and Information, 7:53–75, 1998. (76)

[18] F. Bacon, *Novum Organum*, London, 1620. (16, 28, 30, 35, 40)

[19] A. B. Baker, *Nonmonotonic Reasoning in the Framework of the Situation Calculus*, Artificial Intelligence, 49(1-3):5–23, 1991. (63, 71)

[20] J. Bell, *Remarks on the Evaluation of Formal Theories of Causal Reasoning*, in N. Foo, P. Peppas, V. Lifschitz, E. Sandewall and M. Williams eds., International Joint Conference on Artificial Intelligence, Workshop on Non-monotonic Reasoning, Action, and Change, p. 14–21, 1995. (12, 24)

[21] J. S. Birman, *Scientific Publishing: A Mathematician's Viewpoint*, Notices of the AMS, 47(7):770–774, 2000. (6)

[22] N. J. Blackwood, R. J. Howard, R. P. Bentall and R. M. Murray, *Cognitive Neuropsychiatric Models of Persecutory Delusions*, American Journal of Psychiatry, 158:527–539, 2001. (33)

[23] G. Bossu and P. Siegel, *Saturation, Non-monotonic Reasoning and the Closed-World Assumption*, Artificial Intelligence, 25:13–63, 1985. (11)

[24] S. Brandano, *Features and Fluents for Logic Programming: Simulative Algebraic Semantics*, Technical report, University of London, Department of Computer Science at Queen Mary College, 1999. (63)

[25] S. Brandano, *K-RACi*, in International Joint Conference on Artificial Intelligence, Workshop on Non-monotonic Reasoning, Action, and Change, p. 9–16, 1999. ISBN 1-55860-613-0. (42, 78)

[26] S. Brandano, *Features and Fluents for Logic Programming: Non-simulative Algebraic Semantics*, in Formal and Applied Practical Reasoning, Proceedings of the International Conference, p. 131–143, 2000. ISSN 1469-4166. (38, 42, 63, 109)

[27] S. Brandano, *On the metatheoretic approach to nonmonotonic reasoning, its extension to the continuum and relation with Classical Mechanics*, Linköping Electronic Articles in Computer and Information Science, 5(42), 2000. Updated in 2009. (78)

[28] S. Brandano, *The Event Calculus Assessed*, in Temporal Representation and Reasoning, Proceedings of the IEEE International Symposium, p. 7–11, 2001. Indexed by ACM and IEEE. (38, 42, 63)

[29] S. Brandano, *Assessment Results for the Calculus of Events*, Electronic Transactions on Artificial Intelligence, 2002. To be submitted for open peer-review. (42)

[30] S. Brandano, P. Hayes, J. Ma et al., *Ontologies of Time (Public Debate)*, Electronic Newsletter on Reasoning about Actions and Change, 1998. `http://www.etai.org`. (54)

[31] R. A. Brooks, *Intelligence without Reason*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 569–595, 1991. Computers and Thought Award Lecture, Reprinted in [33, p. 133-86]. (33)

[32] R. A. Brooks, *Intelligence without Representation*, Artificial Intelligence, 47:139–159, 1991. Reprinted in [33]. (22, 33)

[33] R. A. Brooks, *Cambrian Intelligence: The early history of the new AI*, MIT Press, 1999. (33, 56, 136)

[34] A. L. Brown and Y. Shoham, *New results on semantical non-monotonic reasoning*, in Non-monotonic Reasoning, Proceedings of the International Workshop, vol. 346 of Lecture Notes in Artificial Intelligence, p. 19–26, 1989. (11)

[35] A. L. Cauchy, *Cours d'Analyse. Leçons de seconde année pour l'École Royale Polytechnique*. Paris, 1820. Reprinted as "Ordinary Differential Equations – Unpublished Course (Fragment)" by Johnson Reprint Corporation, Paris, 1981. (65)

[36] D. Chapman, *Planning for Conjunctive Goals*, Artificial Intelligence, 32:333–377, 1987. (50)

[37] A. Church, *A Note on the* Entscheidungsproblem, Journal of Symbolic Logic, 1(1):40–41, 1936. (10)

[38] A. Church, *Correction to a Note on the* Entscheidungsproblem, Journal of Symbolic Logic, 1(3):101–102, 1936. (10)

[39] A. Church, *Review of A. M. Turing, «On computable numbers, with an application to the Entscheidungsproblem», Proceedings of the London Mathematical Society, 2 s. vol. 42 (1936-7), p. 230–265*, Journal of Symbolic Logic, 2(1):42–43, 1937. (124)

[40] M. T. Cicero, *Le Orazioni*, vol. 4 of Classici Latini, UTET, Torino, 1983. (32)

[41] K. L. Clark, *Negation as Failure*, in H. Gallaire and J. Minker eds., Logic and Data Bases, Proceedings of the Workshop, p. 293–322, 1978. (11, 20, 79, 89, 137)

[42] K. L. Clark, *Predicate Logic as a Computational Formalism*, Ph.D. Thesis, University of London, Department of Computer Science at Queen Mary College, 1980. Partially published as [41]. (11, 20, 79, 88, 89)

[43] P. J. Cohen, *The Indipendence of the Continuum Hypothesis*, in Proceedings of the National Academy of Sciences, p. 1143–1148, 1963. *Fields Medal* by the International Congress of Mathematicians (Moscow, 1966). (28)

[44] B. Comrie, *Tense*, Cambridge University Press, 1985. (52)

[45] P. Coscia, P. Franceschi, G. Levi, G. Sardu and L. Torre, *Meta-level Definition and Compilation of Inference Engines in the EPSILON Logic Programming Environment*, in R. A. Kowalski and K. Bowen eds., Logic Programming, Proceedings of the International Conference, p. 359–373. MIT Press, 1988. (122)

[46] E. Davis, *Representation of Commonsense Knowledge*, Morgan Kauffman, 1990. (12, 22, 79)

[47] E. Davis and L. Morgenstern, *Progress in Formal Commonsense Reasoning*, Artificial Intelligence, 153:1–12, 2004. (12, 131)

[48] M. Degl'Innocenti, G. Levi and G. Sardu, *A methodology for the design and the automatic composition of structured Prolog meta-interpreters*, Technical report, May 1988. (122)

[49] M. Denecker, *Knowledge Representation and Reasoning in Incomplete Logic Programming*, Ph.D. Thesis, Katholieke Universiteteit Leuven, Sept. 1993. Partially published in [51]. (85)

[50] M. Denecker and D. D. Schreye, SLDNFA*: an abductive procedure for normal abductive programs*, in Logic Programming, Proceedings of the International Joint Conference and Symposium, p. 686–700, 1992. (85)

[51] M. Denecker and D. D. Schreye, *Representing Incomplete Knowledge in Abductive Logic Programming*, Journal of Logic and Computation, 5(5):553–578, 1995. (85, 137)

[52] R. Descartes, *Regulae ad directionem ingenii*, 1628. (29)

[53] R. Descartes, *Principia Philosophiae*, L. Elzevir, Amsterdam, 1644. (29)

[54] R. Descartes, *Discours de la méthode pour bien conduire sa raison et chercher la verité dans les sciences (1637)*, in C. Adam and P. Tannery eds., Oeuvres des Descartes, vol. VI, p. 1–78, L. Cerf, Paris, 1902. (29)

[55] E. W. Dijkstra, *Programming considered as a Human Activity*, in Proceedings of the IFIP Congress, p. 213–217. North-Holland, 1965. (20)

[56] E. W. Dijkstra, *The Humble Programmer*, Communications of the ACM, 15(10):859–866, 1972. Turing Award Lecture. (20, 79, 88)

[57] E. W. Dijkstra, *Society's role in mathematics*, circulated privately, Sept. 1998. (25)

[58] E. W. Dijkstra, *Under the spell of Leibniz's dream*, circulated privately, Apr. 2000. (26)

[59] E. W. Dijkstra. *Denken als Discipline (Discipline in Thought)*. VPRO, 2001. (26)

[60] J. Doyle, *A Truth Maintenance System*, Artificial Intelligence, 12:231–272, 1979. Reprinted in [84]. (11)

[61] E. Emerson, *Temporal and Modal Logic*, vol. B, ch. 16, p. 995–1072, 1990. (133)

[62] H. B. Enderton, *A Mathematical Introduction to Logic*, Academic Press, New York, 1972. (53)

[63] H. T. Engelhardt, Jr. and A. L. Caplan, *Patterns of controversy and closure: the interplay of knowledge, values, and political forces*, in H. T. Engelhardt, Jr. and A. L. Caplan eds., Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology, p. 1–23, Cambridge University Press, 1987. (16)

[64] K. Engesser and D. M. Gabbay, *Quantum logic, Hilbert space, revision theory*, Artificial Intelligence, 136:61–100, 2002. (78)

[65] K. Eshghi and R. A. Kowalski, *Abduction compared with Negation by Failure*, in Logic Programming, Proceedings of the 6th International Conference and 5th Symposium, p. 234–254. MIT Press, 1989. (11, 63, 79)

[66] A. Favaro ed., *Opere di Galileo Galilei*, Barbèra, Firenze, 1890–1909. Venti Volumi. (28)

[67] D. L. Feygin, J. E. Swain and J. F. Leckman, *The normalcy of neurosis: Evolutionary origins of Obsessive-Compulsive Disorder and related behaviors*, Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30:854–864, 2006. (33)

[68] R. E. Fikes and N. J. Nilsson, STRIPS*: A new approach to the application of theorem proving to problem solving*, Artificial Intelligence, 2(3-4):189–208, 1971. (50)

[69] A. Flanagin, L. A. Carey, P. B. Fontanarosa, S. G. Phillips, B. P. Pace, G. D. Lundberg and D. Rennie, *Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals*, Journal of the American Medical Association, 280:222–224, 1998. (7)

[70] N. Foo, D. Zhang, Y. Zhang, S. Chopra and B. Q. Vo, *Encoding Solutions of the Frame Problem in Dynamic Logic*, in T. Eiter, W. Faber and M. Truszczyński eds., Logic Programming and Non-monotonic Reasoning, Proceedings of the International Conference, vol. 2173 of Lecture Notes in Artificial Intelligence, p. 240–253, 2001. (63)

[71] G. Frege, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens (Ideography, a formula language, modeled upon that of arithmetic, for pure thought)*, Verlag von Louis Nebert, Halle an der Saale, Germany, 1879. (18, 40)

[72] G. Frege, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Blackwell, 1968. English translation by J. L. Austin of the original 1884 work. (18, 40)

[73] S. Freud, *The Dissection of the Psychical Personality (1933)*, in J. Strachey ed., Standard Edition of the Complete Psychological Works of Sigmund Freud, vol. 22, Vintage, The Hogarth Press and the Institute of Psycho-Analysis, 2001. (42)

[74] D. M. Gabbay, C. J. Hogger and J. A. Robinson eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*, Oxford University Press, 1993–1998. five volumes. (11)

[75] D. M. Gabbay, C. J. Hogger and J. A. Robinson eds., *Handbook of Logic in Artificial Intelligence and Logic Programming — Nonmonotonic Reasoning and Uncertain Reasoning*, vol. 3, Oxford University Press, 1994. (23, 40)

[76] M. Gabrielli, G. Levi and M. C. Meo, *Observable Behaviours and Equivalences of Logic Programs*, Information and Computation, 122(1):1–29, Oct. 1995. (39)

[77] G. Galilei, *Sidereus Nuncius*, Venezia, 1610. (30)

[78] G. Galilei, *Il Saggiatore (1622)*, Roma, 1623. (30)

[79] G. Galilei, *Dialogo sopra i due massimi sistemi del mondo (1630)*, Firenze, 1632. Declared heretical, it will appear in the *Index Librorum Prohibitorum* from 1633 to 1824. (29, 30)

[80] G. Galilei, *Discorsi e dimostrazioni matematiche intorno a due nuove scienze*, Louis Elsevier, Leiden, 1638. Collected in A. Favaro ed., Opere di Galileo, volume 8 di 20, Firenze 1898. In Italian. (65, 67)

[81] M. Gelfond and V. Lifschitz, *The Stable Model Semantics for Logic Programming*, in R. Kowalski and K. Bowen eds., Logic Programming, Proceedings of the International Conference, p. 1070–1080, 1988. (11, 63, 79)

[82] M. Gelfond and V. Lifschitz, *Representing Action and Change by Logic Programs*, Journal of Logic Programming, 17(2-4):301–321, 1993. (85)

[83] M. Gelfond, V. Lifschitz and A. Rabinov, *What are the limitations of the Situation Calculus?*, in R. S. Boyer ed., Automated Reasoning, p. 167–181, Kluwer Academic Publishers, 1991. (71)

[84] M. L. Ginsberg ed., *Readings in Nonmonotonic Reasoning*, Morgan Kauffman, 1987. (138, 143)

[85] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I (On Formally Undecidable Propositions of Principia Mathematica and Related Systems)*, Monatshefte für Mathematik und Physik, 38:173–198, 1931. English translation in [214, p. 596–616]. (10, 28)

[86] K. Gödel, *The consistency of the axiom of choice and of the generalised continuum hypothesis with the axioms of set theory*, Annals of Mathematics Studies, 3, 1940. (28)

[87] F. Godlee and T. Jefferson, *Peer review in health sciences*, BMJ Publishing Group, London, 2nd edition, 2003. (17)

[88] J. Y. Halpern and Y. Moses, *Towards a Theory of Knowledge and Ignorance: Preliminary Report*, in K. R. Apt ed., Logics and Models of Concurrent Systems, p. 459–476, Springer-Verlag, 1985. (11)

[89] S. Hanks and D. V. McDermott, *Default Reasoning, Non-monotonic Logics, and the Frame Problem*, in Artificial Intelligence, Proceedings of the AAAI Conference, p. 328–333, 1986. AAAI Best Paper Award (1986), AAAI Classic Paper Award (2005). (21, 85)

[90] S. Hanks and D. V. McDermott, *Nonmonotonic Logic and Temporal Projection*, Artificial Intelligence, 33(3):379–412, 1987. (21, 48, 69, 85, 88)

[91] J. E. Harmon, *The Structure of Scientific and Engineering Papers: A Historical Perspective*, Transactions on Professional Communication, 32(3):132–138, 1989. (27)

[92] D. Hilbert and W. Ackermann, *Grundzüge der Theoretischen Logik*, Springer-Verlag, 1928. Reprinted in 1972; English translation by Lewis Hammond et al., "Principles of Mathematical Logic," Chelsea, New York, 1950. (28)

[93] H. F. Judson, *Authorship, Ownership: Problems of Credit, Plagiarism, and Intellectual Property*, in The Great Betrayal, Harcourt, 2004. (7)

[94] A. C. Kakas, R. A. Kowalski and F. Toni, *The Role of Abduction in Logic Programming*, Handbook of Logic in Artificial Intelligence and Logic Programming, 5:235–324, 1998. (11, 63)

[95] G. N. Kartha, *Soundness and completeness theorems for three formalisations of action*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 724–729, 1993. (23, 63)

[96] G. N. Kartha and V. Lifschitz, *Actions with Indirect Effects (preliminary report)*, in Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, p. 341–350, 1994. (23)

[97] H. Kautz, *The Logic of Persistence*, in Artificial Intelligence, Proceedings of the AAAI Conference, p. 401–405, 1986. (37, 63)

[98] S. Kleene, *Mathematical Logic*, Wiley, 1967. (124)

[99] S. C. Kleene, *Introduction to Metamathematics*, Elsevier Science, 1952. (11th reprint, 1996). (83)

[100] R. A. Kowalski, *Predicate Logic as Programming Language*, in Proceedings of the IFIP Congress, p. 569–574, Stockholm, 1974. (79, 80, 88)

[101] R. A. Kowalski, *Logic for Problem Solving*, North-Holland, 1979. (11)

[102] R. A. Kowalski, *Database Updates in the Event Calculus*, Journal of Logic Programming, 12:121–146, 1992. (89)

[103] R. A. Kowalski and M. Sergot, *A Logic-based Calculus of Events*, New Generation Computing, 4(1):67–95, 1986. (11, 20, 52, 79, 88, 89, 109)

[104] T. S. Kuhn, *The Structure of Scientific Revolutions*, The University of Chicago Press, third edition, 1962. Third edition, 1996. (12, 17, 19)

[105] T. S. Kuhn, *The Function of Dogma in Scientific Research*, in A. C. Crombie ed., Scientific Change: historical studies in the intellectual, social, and technical conditions for scientific discovery and technical invention, from antiquity to the present. Proceedings of the Symposium held at the University of Oxford, July 9-15, 1961, p. 347–69, London, 1963, Heinemann Press. (12)

[106] B. J. Kuipers, *Qualitative Reasoning: Modelling and Simulation with Incomplete Knowledge*, MIT Press, 1994. (67, 133)

[107] I. Lakatos, *Essays in the Logic of Mathematical Discovery*, Ph.D. Thesis, University of Cambridge, England, 1961. (19)

[108] P. Langley, H. A. Simon, G. Bradshaw and J. Zytkow, *Scientific Discovery: An Account on the Creative Process*, MIT Press, 1986. (77)

[109] G. Leech, *Meaning and the English verb*, Longman, 2004. (116)

[110] G. Levi and G. Sardu, *Partial Evaluation of Metaprograms in a "Multiple Worlds" Logic Language*, New Generation Computing, 6(2,3):227–247, 1988. (122)

[111] P. Liberatore, *The Complexity of the Language $\mathcal{A}$*, Linköping Electronic Articles in Computer and Information Science, 2(6), 1997. `http://www.ep.liu.se/ea/cis/1997/006/`. (85)

[112] V. Lifschitz, *Computing Circumscription*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 121–127. Morgan Kauffman, 1985. (89)

[113] V. Lifschitz, *Benchmark Problems for Formal Nonmonotonic Reasoning – Version 2.00*, in Non-monotonic Reasoning, Proceedings of the International Workshop, vol. 346 of Lecture Notes in Artificial Intelligence, p. 202–219, 1988. (21)

[114] V. Lifschitz, *Toward a Metatheory of Action*, in J. Allen, R. Fikes and E. Sandewall eds., Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, p. 376–386, 1991. (23)

[115] V. Lifschitz, *Circumscription*, in D. M. Gabbay, C. J. Hogger and J. A. Robinson eds., Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3, p. 298–352, Oxford University Press, 1994. (11, 20, 89, 92)

[116] V. Lifschitz ed., *Formalizing Common Sense: Papers by John McCarthy*, Intellect, 1998. (30, 124)

[117] F. Lin and Y. Shoham, *Provably Correct Theories of Action (preliminary report)*, in Artificial Intelligence, Proceedings of the AAAI Conference, p. 349–354, 1991. (23)

[118] C. Linnaeus, *Philosophia Botanica*, 1751. (54)

[119] J. Locke, *An Essay Concerning Human Understanding*, London, 1690. (20)

[120] D. C. Makinson, *General Patterns in Nonmonotonic Reasoning*, in D. M. Gabbay, C. J. Hogger and J. A. Robinson eds., Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3, p. 35–110, Oxford University Press, 1994. (11, 20, 90)

[121] J. McCarthy, *The inversion of functions defined by Turing machines*, in C. E. Shannon and J. McCarthy eds., Automata Studies, p. 177–181, Princeton University Press, Princeton, New Jersey, 1956. (19, 124)

[122] J. McCarthy, *Programs with Common Sense*, in D. V. Blake and A. M. Uttley eds., Mechanisation of Thought Processes, Proceedings of the Symposium, p. 75–91, Teddington, England, 1959, Her Majesty's Stationery Office. (11, 124)

[123] J. McCarthy, *Situations, Actions and Casual Laws*, Memo 2, Stanford University, Artificial Intelligence Project, 1963. Reprinted in [138, p. 410-417]. (11, 130)

[124] J. McCarthy, *Epistemological Problems of Artificial Intelligence*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 1038–1044, 1977. (11, 20)

[125] J. McCarthy, *Circumscription: A form of Non-Monotonic Reasoning*, Artificial Intelligence, 13(1-2):27–39, 171–172, 1980. (11, 20, 89)

[126] J. McCarthy, *Applications of Circumscription to formalising Common Sense Knowledge*, in Non-monotonic Reasoning, Proceedings of the International Workshop, p. 295–324, Oct. 1984. (21, 37, 63)

[127] J. McCarthy, *Applications of Circumscription to formalising Common Sense Knowledge*, Artificial Intelligence, 28(1):89–116, 1986. (11, 37, 63, 89)

[128] J. McCarthy, *Generality in Artificial Intelligence*, Communications of the ACM, 30(12):1030–1035, 1987. Turing Award Lecture. (79)

[129] J. McCarthy, *Concepts of Logical AI*, in Logic-based Artificial Intelligence, p. 37–52, Kluwer Academic Publishers, 2001. (130, 131)

[130] J. McCarthy, *Actions and Other Events in Situation Calculus*, in Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, 2002. (130)

[131] J. McCarthy, *Problems and projections in Computer Science for the next forty-nine years*, Journal of the ACM, 50(1):73–79, 2003. (26, 79)

[132] J. McCarthy and T. Costello, *Combining Narratives*, in Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, 1998. (130)

[133] J. McCarthy and P. J. Hayes, *Some philosophical problems from the standpoint of Artificial Intelligence*, Machine Intelligence, 4:463–502, 1969. (10, 11, 48, 52, 64, 109, 130, 131)

[134] D. V. McDermott and J. Doyle, *Non-Monotonic Logic I*, Artificial Intelligence, 13:41–72, 1980. Reprinted in [84, pp. 111–126]. (11)

[135] R. Milner, *Processes: A mathematical model of computing agents*, in Logic Colloquium, p. 157–174, North-Holland, Amsterdam, 1973. (39)

[136] J. Minker, *Editorial Introduction*, Theory and Practice of Logic Programming, 1(1), 2001. (6)

[137] M. L. Minsky ed., *Semantic Information Processing*, MIT Press, 1968. (142)

[138] J. D. Monk, *Mathematical Logic*, Springer-Verlag, Berlin, 1976. (53)

[139] S. H. Muggleton, *Inductive Acquisition of Expert Knowledge*, Ph.D. Thesis, University of Edinburgh, 1986. (56, 77)

[140] S. H. Muggleton and L. D. Raedt, *Inductive Logic Programming: Theory and Methods*, Journal of Logic Programming, 19-20:629–679, 1994. (56, 77)

[141] A. Ness, *'Truth' As Conceived by Those Who Are Not Professional Philosophers*, Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, II. Hist. Filos. Klasse, IV, 1938. (18)

[142] I. Newton, *Philosophiae Naturalis Principia Mathematica*, Printed for the Royal Societry by Joseph Streater, London, first edition, 5 July 1686. (65)

[143] N. J. Nilsson, *SHAKEY the Robot*, Technical Report 323, SRI, AI Center, Menlo Park, California, Apr. 1984. (33)

[144] J. Pearl, *Probabilistic Semantics for Nonmonotonic Reasoning*, in R. Cummins and J. Pollock eds., Philosophy and AI: Essays at the Interface, p. 157–187, MIT Press, 1991. (77)

[145] J. Pearl, *Reasoning with Cause and Effect*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 1437–1449, 1999. (77)

[146] E. Pednault, *ADL: Exploring the middle ground between STRIPS and the Situation Calculus*, in Principles of Knowledge Representation and Reasoning, Proceedings of the International Conference, p. 324–332, 1989. (63)

[147] T. Persson and L. Staflin, *A Causation Theory for a Logic of Continuous Change*, in Artificial Intelligence, Proceedings of the European Conference, p. 497–502, 1990. (63)

[148] Plato, *Meno*, Oxford University Press, 1953. Translated by Benjamin Jowett. (30)

[149] Plato, *Parmenides*, Oxford University Press, 1953. Translated by Benjamin Jowett. (70)

[150] K. R. Popper, *The Logic of Scientific Discovery*, Harper and Row, New York, 1934. (143)

[151] K. R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, 1963. (19)

[152] K. R. Popper, *Logica della Scoperta Scientifica*, Einaudi, Torino, 1995. Italian translation of [151]. (77)

[153] V. R. Pratt, *Semantical considerations on Floyd-Hoare Logic*, in Foundations of Computer Science, Proceedings of the IEEE Symposium, p. 109–121, 1976. (63)

[154] A. N. Prior, *Time and Modality*, Clarendon Press, Oxford, 1957. (50)

[155] A. N. Prior, *Past, Present and Future*, Clarendon Press, Oxford, 1967. (50)

[156] R. Quirk, S. Greenbaum, G. Leech and J. Svartvik, *A Comprehensive Grammar of the English Language*, Longman, 1985. Index by David Crystal. (34, 52, 116, 117, 132, 133)

[157] H. Reichenbach, *The Direction of Time*, University of California Press, 1956. (51)

[158] R. Reiter, *On Closed World Data Bases*, in H. Gallaire and J. Minker eds., Logic and Data Bases, Proceedings of the Workshop, p. 119–140, Plenum Press, New York, 1978. (11, 20)

[159] R. Reiter, *On Reasoning by Default*, in default reasoning ed., Theoretical Issues in Natural Language Processing, Proceedings of the 2nd Symposium, Urbana, Illionis, 1978. (11)

[160] R. Reiter, *A Logic for Default Reasoning*, Artificial Intelligence, 13(1-2):81–132, 1980. (11)

[161] R. Reiter, *Circumscription implies Predicate Completion (sometimes)*, in Artificial Intelligence, Proceedings of the AAAI Conference, p. 418–420, 1982. (11, 20, 89)

[162] R. Reiter, *The Frame Problem in the Situation Calculus: A simple solution (sometimes) and completeness result for goal regression*, in V. Lifschitz ed., Artificial Intelligence and Mathematical Theory of Computation, p. 359–380, Academic Press, 1991. (63)

[163] R. Reiter, *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*, MIT Press, 2001. (11, 26, 109, 130, 131)

[164] E. Sandewall, *An approach to the Frame Problem and its Implementation*, Machine Intelligence, 7:195–204, 1972. (11, 19, 20)

[165] E. Sandewall, *Filter Preferential Entailment for the logic of action in almost continuous worlds*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 894–899, 1989. (11, 89, 90)

[166] E. Sandewall, *Causal Qualification and Structure-Based Ramification*, in Logical Formalizations of Commonsense Reasoning, Proceedings of the Symposium, 1993. (69, 71)

[167] E. Sandewall, *The Range of Applicability of Non-monotonic Logics for the Inertia Problem*, in Artificial Intelligence, Proceedings of the International Joint Conference, 1993. (11, 22, 36)

[168] E. Sandewall, *Features and Fluents: The Representation of Knowledge about Dynamical Systems*, vol. 30 of Oxford Logic Guides, Oxford University Press, 1994. (12, 22, 23, 32, 34, 36, 37, 38, 42, 44, 45, 46, 47, 48, 49, 50, 60, 63, 85, 91)

[169] E. Sandewall, *Cognitive Robotics Logic and its Meta-theory: Features and Fluents Revisited*, Electronic Transactions on Artificial Intelligence, 2(3-4):307–329, 1998. `www.ep.liu.se/ej/etai/1998/010/`. (12)

[170] E. Sandewall. *List of Individual Researchers in 'Nonmonotonic Reasoning about Actions and Change'*, 1999. `http://www.etai.org`. (11)

[171] E. Sandewall, *Book Review: M. Shanahan, 'Solving the Frame Problem'*, Artificial Intelligence, 123(1,2):271–273, Dec. 2000. (26, 104)

[172] E. Sandewall et al. *Ontologies for Actions and Change (Continuation of the debate held at the International Joint Conference on Artificial Intelligence, Workshop on Nonmonotonic Reasoning, Actions, and Change)*. News Journal of the ETAI, October-November 1997. `http://www.ida.liu.se/ext/etai/actions/nj/9710-2/`. (24)

[173] E. Sandewall and Y. Shoham, *Non-monotonic Temporal Reasoning*, in D. M. Gabbay, C. J. Hogger and J. A. Robinson eds., Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 4, ch. 7, p. 439–498, Oxford University Press, 1995. (12, 23, 34, 37, 38, 42, 61, 63, 88, 124, 130)

[174] D. Sangiorgi, *Bisimulation: From the origins to today*, in Logic in Computer Science, Proceedings of the IEEE Symposium, p. 298–302, 2004. (38)

[175] J. Sebelik and P. Stepanek, *Horn Clause Programs for Recursive Functions*, in K. L. Clark and S. Å. Tärnlund eds., Logic Programming, p. 324–340, Academic Press, 1982. (79)

[176] M. P. Shanahan, *Prediction is Deduction but Explanation is Abduction*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 1055–1060, 1989. (100)

[177] M. P. Shanahan, *Representing Continuous Change in the Event Calculus*, in Artificial Intelligence, Proceedings of the European Conference, p. 598–603, 1990. (90)

[178] M. P. Shanahan, *A Circumscriptive Calculus of Events*, Artificial Intelligence, 77(2):249–284, 1995. (20, 88, 89)

[179] M. P. Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press, 1997. This work was conducted while the author was a Senior Research Fellow at University of London, Department of Computer Science at Queen Mary College. (11, 12, 20, 21, 24, 64, 67, 79, 88, 89, 92, 94, 95, 100, 109, 124)

[180] M. P. Shanahan, *The Event Calculus Explained*, vol. 1600 of Lecture Notes in Artificial Intelligence, p. 409–430, 1999. (71, 88, 89, 90, 94, 95)

[181] M. P. Shanahan, *An Abductive Event Calculus Planner*, Journal of Logic Programming, 44(1,3):207–239, 2000. (88, 89, 100)

[182] M. P. Shanahan, *An attempt to formalise a non-trivial benchmark problem in commonsense reasoning*, Artificial Intelligence, 153:141–165, 2004. (12, 103, 104)

[183] C. E. Shannon and J. McCarthy eds., *Automata Studies*, Princeton University Press, 1956. (11)

[184] Y. Shoham, *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*, Ph.D. Thesis, Yale University, 1986. Published as [187]. (11, 20)

[185] Y. Shoham, *Nonmonotonic Logic: Meaning and Utility*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 388–393, 1987. (11, 20, 21, 36, 40)

[186] Y. Shoham, *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*, MIT Press, 1988. (11, 20, 90, 145)

[187] M. Simi, *Viewpoints subsume Belief, Truth and Situations*, in Artificial Intelligence, Proceedings of the Italian Conference, vol. 549 of Lecture Notes in Artificial Intelligence, p. 38–47, 1991. (76)

[188] J. Simpson and E. Weiner eds., *Oxford English Dictionary*, Oxford University Press, second edition, 1989. Twenty-volumes (1989), together with Addi-

tions Series volumes 1-2 (1993) and volume 3 (1997). Electronic edition, 2002. (30, 35)

[189] R. Smith, *Medical journals are an extension of the marketing arm of pharmaceutical companies*, Public Library of Science: Medicine, 2(5):364–366, 2005. (17)

[190] R. Spier, *The history of the peer-review process*, Trends in Biotechnology, 20(8):357–8, 2002. (31)

[191] S. Å. Tärnlund, *Horn Clause Computability*, BIT, 17(2):215–226, 1977. (79)

[192] Tarnow, D. Young and Cohen, *Coauthorship in pathology, a comparison with physics and a survery-generated and member-preferred authorship guideline*, MedGenMed, 6(3):1–2, 2004. (7)

[193] E. Tarnow, *Coauthorship in Physics*, Science and Engineering Ethics, 8(2):175–190, 2002. (7)

[194] A. Tarski, *Pojęcie prawdy w językach nauk dedukcyjnych (On the concept of truth in languages of deductive sciences)*, in Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III Sciences Mathématiques et Physiques, vol. 34, 1933. In Polish. (146)

[195] A. Tarski, *O pojciu wynikania logicznego (On the Concept of Logical Consequence)*, Przeglad Filozoficzny (Philosophical Review), 39:58–68, 1936. In Polish. (146)

[196] A. Tarski, *Über den Begriff der Logischen Folgerung (On the Concept of Logical Consequence)*, in Actes du Congrès International de Philosophie Scientifique, vol. 7, p. 1–11, Paris, 1936. In German. (146)

[197] A. Tarski, *The Semantic Conception of Truth and the Foundations of Semantics*, Philosophy and Phenomenological Research, 4(3):341–376, 1944. (10, 17, 19, 29, 36)

[198] A. Tarski, *Contributions to the theory of models I*, Indagationes Mathematicae, 16:572–581, 1954. (36)

[199] A. Tarski, *A Lattice-theoretical Fix-point Theorem and its Applications*, Pacific Journal of Mathematics, 5:285–309, 1955. (82)

[200] A. Tarski, *On the Concept of Logical Consequence (1936)*, in Logic, Semantics, Metamathematics: papers from 1923 to 1938, p. 409–420, Hackett, second edition, 1983. English translation of [197] by J. H. Woodger. (19)

[201] A. Tarski, *The Concept of Truth in Formalised Languages (1929–1933)*, in Logic, Semantics, Metamathematics: papers from 1923 to 1938, p. 152–278, Hackett, second edition, 1983. English translation of [195] by J. H. Woodger. (10)

[202] A. Tarski, *On the Concept of Following Logically (1936)*, History and Philosophy of Logic, 23:155–196, 2002. English translation of [196] by M. Stroinska and D. Hitchcock. (19)

[203] A. Tarski et al., *Undecidable Theories*, Studies in Logic and the Foundations of Mathematics, North-Holland, 1953. 2nd ed. 1968; 3rd ed. 1971. (29, 36)

[204] M. Thielscher, *Introduction to the Fluent Calculus*, Electronic Transactions on Artificial Intelligence, 2(3-4):179–192, 1998. (130)

[205] M. Thielscher, *From Situation Calculus to Fluent Calculus*, Artificial Intelligence, 111(1-2):277–299, 1999. (130)

[206] A. M. Thomson, *The Refereeing Process*, in P. Hills ed., Publish or Perish, p. 11–26, Peter Francis Publishers, Norfolk, UK, 1999. (16)

[207] D. S. Touretzky, J. F. Horty and R. H. Thomason, *A clash of intuitions: The current state of non-monotonic multiple inheritance systems*, in Artificial Intelligence, Proceedings of the International Joint Conference, p. 476–482, 1987. (22)

[208] A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proceedings of the London Mathematical Society, 2(42):230–65, 1937. (10, 17, 18, 124, 125, 126)

[209] A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem: A Correction*, Proceedings of the London Mathematical Society, 2(43):544–6, 1937. (124)

[210] A. M. Turing, *Intelligent Machinery*, Technical report, National Physical Laboratory, 1948. (124)

[211] A. M. Turing, *Computing Machinery and Intelligence*, Mind, 49(236):433–460, 1950. (19, 124)

[212] M. H. van Emden and R. A. Kowalski, *The Semantics of Predicate Logic as Programming Language*, Journal of the ACM, 23(4):733–742, 1976. (79, 80, 88)

[213] J. van Heijenoort ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Harvard University Press, 1967. Third printing, 1976. (140)

[214] J. T. Wilson, *Responsible Authorship and Peer Review*, Science and Engineering Ethics, 8(2):155–174, 2002. (17)

[215] L. A. Zadeh, *Fuzzy Sets*, Information and Control, 8(3):338–353, 1965. (53)

[216] L. A. Zadeh, *Fuzzy Logic*, Computer, 21(4):83–93, Apr. 1988. (53)

INDEX

LONDON
A.D. MMVIII