1 Introduction

The prejudicial difficulty encountered in the development of both General Proof Theory in the sense of Prawitz [73, 74] and the so-called Artificial Intelligence in the sense of Turing [99, 100] is that the procedure for establishing the validity of a sentence consists essentially of certain mental operations. However, that these operations are biological processes of some sort is of no help, as far as the true nature of these processes remains unknown. On the other hand, since the *Organon* of Aristotle, *reasoning* has revealed itself as an object of scientific enquiry. These "mental operations", in fact, can be investigated through their linguistic representation, that is through explicit reasoning formally expressed in a language. The quality of the results depends on the quality of the linguistic representation reflecting the structure of these operations. The objective difficulty [1] is that such a structure does not appear to be unique, nor does it appear to be static. An important branch of formal reasoning, namely *non-monotonic reasoning*, has focused the efforts of a specialised and fragmented community for about 30 years. The fragmentation is due to this non unique and evolving structure. On one hand, the community does not want to agree on the development of a single logic, although Tarskian first order logic still remains the starting point of it all. On the other hand, branches of it apply constraints in order to favour the research in situation calculus and event calculus against linear logic with modalities, for example, so that valuable insight may be neglected or discouraged because of one's favourite logic. The fact is that new logics do emerge, each of which is claimed to be elegant, intuitive and expressive. The fact is also that despite all claims, the different logics do impose quite a natural division among the different speakers, and there is no precise knowledge of what is their problem-solving power. The community calls for freedom, but also needs a method of measuring the *objective* progress, both in front of the specific community and the mainstream.

In 1994 a methodology of research has been proposed in order to classify and compare certain non-monotonic reasoning methods. The methodology is based on the following idea: to classify these methods against a taxonomy of explicit simplifying assumptions so to measure the objective *range of applicability* or *problem-solving power* of the reasoning methods being assessed. Although the methodology has been proven successful so far, the community has not yet taken active part in it. Understanding the advantages of its general practice, together with a higher than usual learning curve and the usual lack of time, are the most lamented difficulties.

In this paper, the interested reader will be firstly introduced to the context, the idea, the definition and the results of this methodology (section 2). The comparison of this methodology with the standard scientific methodology will then follow, taking classical Newtonian Mechanics as a case study. As Newton's theory implicitly involves a number of simplifying assumptions, these assumptions are enumerated and classified (section 3). The boundaries of the result are then outlined.

[1] Additional objective difficulties are that the reasoning systems being developed so far are disembodied [12, 14] and non-social [49]. In [102, 51, 50] Wooldridge and Jennings have compiled an authoritative survey of theoretical results on embodied (hybrid) reasoning systems, together with a list of commercial applications.

2 Features and Fluents

Mathematicians are like Frenchmen: whatever you say to them they translate into their own language, and forthwith it is something entirely different. (Johann Wolfgang von Goethe)

2.1 The context

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 [2] that *truth is defeasible.*

Monotonicity is a well known property of classical and modal logic, and is also an important property of scientific reasoning at large. If Γ is the explicit state of knowledge of a theory T_i, the relation $Cn_i(\Gamma) \supset Cn_i(\Gamma \cup \{\alpha\})$ means that all theorems in Γ that depend on α, either directly or indirectly, have failed together with their own consequences. The more numerous the dependencies on α, the more numerous the falsified theorems. If T_j is a certain other theory based on T_i, i.e. it uses T_i as it's founding theory, the state of knowledge of T_j will be affected too. The state of knowledge of an authoritative discipline can not afford such revolutions, which is why Mathematics got seriously involved in its own foundations and, in particular, in the proof of the non-contradictoriness of its funding theories: arithmetic and analysis. That was Hilbert's 1904 [45, 46] program for a Reductive Proof Theory. In 1931 [36] Gödel proves that in any arithmetic system there are propositions which can not be proved or disproved *within the system.* In 1938 [37] Gödel reduces the consistency of set theory to that of the Axiom of Choice and the Continuum Hypothesis (namely the first of Hilbert's 23 problems [44]), and in 1963 Cohen extends the result to the negation of AC and CH. By *force majeure*, the resulting idea with the foundations of Mathematics is that the coherence of a theory must be a relative notion, as the absolute notion of internal coherence can never be achieved.

Five years after Gödel's theorem, in 1936 [96, p. 417], Tarski introduced explicitly the *semantic notion of (absolute) truth*, embodied in the notion of logical consequence as semantical counterpart to the syntactical notion of derivability [7, p. 319]. This notion of truth was already developed by Gentzen to a considerable degree in 1932 [35] and is analogous to a far-reaching extent to that of Bolzano [8], presented over a hundred years earlier. Due to the application of logic to the study of the foundations of mathematics, this notion is such that a sentence (a thesis, typically) needs to be either true or false regardless of who reads it, as well as of when and where its truth is evaluated; i.e. a sentence must not change its truth or falsity by, for example, changing room or reader or passing time, nor is it allowed for an intermediate state of knowledge.

[2] *The thesis of 'Proofs and Refutations' [56] is that the development of mathematics does not consist (as conventional philosophy of mathematics tells us it does) in the steady accumulation of eternal truths. Mathematics develops, according to Lakatos, in a much more dramatic and exciting way - by a process of conjecture, followed by attempts to 'prove' the conjecture (i.e. to reduce it to other conjectures) followed by criticism via attempts to produce counter-examples both to the conjectured theorem and to the various steps in the proof.* (Worral [103])

The extension of symbolic logic to the notion of relative truth, of which non-monotonicity is a special case, is the ongoing Copernican revolution in modern symbolic logic. Despite the different implementations, the gathering idea is that *truth is relative* to a formalised context, where the context may be the explicit knowledge of a formal theory, a database (eg. the heap of a programming language), the mental state of an agent (eg. a set of beliefs), a social context for cooperating and negotiating agents, or the position in space at a certain point in time. As the context may change, *time* is a good way of ordering the contexts and *actions* are a good way of justifying their change. For truth to be relative, however, it also needs to be *consistent* with the known facts [47, p. 13]; in other words, it needs to be non-trivial. As knowledge is a formal context itself, a general theory of relative truth must allow for its change and must be able to verify its own consistency. Gödel's 1931 result is not expected to apply, as the theory is not required to infer absolute truths and its consistency is the result of a continuous validation against incoming knowledge and observations.

On the idea, it is important to observe that the challenge for a notion of relative truth is far from novel. It was Aristotle to observe that sentences vary in their truth and falsehood with the times at which they are made or held, just as concrete things have different qualities at different times. In more recent days, besides non-monotonicity, it is a duty to mention *pragmatics* [67, 66] [3], *modal temporal logic* [75, 76, 71, 57] [4] and *belief revision* (see, eg. [32, 33]). On the applications, it is important to observe that the rigour of classical logic and set theory is often referred to as the model to be used when specifying properties of mechanical procedures for von Neumann calculators, so that monotonicity is expected to be a basic property of all software products, including formal programming languages, which interpreters are such mechanical procedures themselves. It is not surprising that the design of such products follows the same pattern described by Lakatos for the development of Mathematics. In principle, a non-monotonic model would allow such mechanical procedures to face an inconsistency, possibly improving themselves autonomously.

[3] *The study of language (or semiotic) was partitioned in (Morris, 1938) into three branches –syntax, semantics, and pragmatics– that may be characterised roughly as follows. Syntax is concerned solely with relations between linguistic expressions; semantics with relations between expressions and the objects to which they refer; and pragmatics with relations among expressions, the objects to which they refer, and the users or contexts of use of the expressions.* (Montague, 1968)

[4] The adequacy of modal logic from the standpoint of AI is still an open debate. It is since 1969 [65], for example, that McCarthy manifests his doubts. More recently, the following address:

"*I propose that a robot be able to introduce new modalities as new predicates. In logician's terminology, this changes the language but not the logic. I have regarded introducing new modalities to modal logical systems as changing the logic.*

What concerns me [...] is the possible implication that there will always be a person around to decide what formalism to use. Our formalism for modality needs to be expressive enough, so that we can imagine the robot deciding for itself what formalism to use for a problem." (J. McCarthy [64])

The debate seems to originate from the thesis of extensionality, that owes its name to Russell (Principia Mathematica, I, pag. XIV, 659 sgg.) and Carnap [16, ch. 67]. The thesis states that for any non-extensional system S_n there is an extensional system S_e for which S_n is reducible to S_e. As the most important intensional sentences are modal, the thesis states the reducibility of modal sentences into non-modal sentences. This is nothing more than a thesis, and Carnap himself did not validate it [17, ch. 32].

Results and Methods

In 1969 [65, p. 487] the novel challenge of using logic for representing natural-language information about the effects of natural and artificial events in the common-sense world, lead McCarthy and Hayes to the discovery of the *Frame Problem.* The discovery is that classical mathematical logic can not implicitly express the non-effects of events, and that the explicit representation of such knowledge would result in an enormous amount of sentences. In 1972 [82, p. 200] Sandewall first addressed the frame problem using a non-monotonic logical formalism, as the very first result of a field known today as *Non-monotonic Reasoning about Actions and Change* (NRAC). Due to the relevance of the discovery, and the perceived urge to deliver in computer science, much of the later work after 1972 consisted in double checking the property of non-monotonicity, the first historical results being that of the Closed World Assumption rule of Reiter [78] and the Negation as Failure of Clark [19] in logic and databases. The novel field of research lead to a cornucopia of different formalisms and techniques, including the combination of logic with differential equations [83]. Two basic difficulties motivated the different approaches. The prime difficulty is that in classical mathematical logic the well formed formulas and the notion of logical consequence are context-free; representing the context is then the first technical problem, where Tarski's original definition of logical consequence is either preserved, so that the context is represented directly by means of sentences of the language, like in the situation calculus [65], or modified so to explicitly refer to a domain or world of interpretation, like in modal logic. The other difficulty is due to the fact that much of the studies were conducted away from the proper field of symbolic logic, its rigour and its journals.

In 1986 the following criticism appeared in the literature:

> *The non-standard nature of the various systems [non-monotonic formalisms] and their diversity has made it hard to gain a good understanding of them and to compare among them.* [92]

A proof-theoretic and a model-theoretic approach were proposed in order to classify and compare the different systems.

Gabbay proposed to consider the output of those systems. He raised the question "when does a provability relation represent a non-monotonic consequence relation?" and formulated an answer along the lines of Tarski [97, p. 31] and Scott [91, p. 414], namely the provability relation is a non-monotonic consequence relation iff it satisfies the rules of reflexivity, restricted monotonicity (i.e. $A \subseteq B \subseteq Cn(A)$ implies $Cn(A) \subseteq Cn(B)$) and transitivity (or cut) [29, p. 455]. The development of this approach lead to what Gabbay refers to as the axiomatic approach to non-monotonic reasoning [61, 54, 28, 43, 52, 90] in the context of sub-structural logics [5] [30].

Shoham approached the problem from a model-theoretic point of view. In 1986 he introduced explicitly the notion of non-monotonic logical consequence [92] [93, pages 73-77], defined in terms of a preference-relation

[5] Sub-structural logics are non-classical logics defined by rejecting or restricting some of Gentzen's structural rules. The most important example is Brouwer's intuitionist logic. Other examples are relevant logic, BCK logic and Girard's linear logic. Lambek's calculus of syntactic categories is another interesting example, that belongs primarily to mathematical linguistics and has deep connections with Category Theory. See [31] for the details. See also [69, 48] for the sub-structural logic of Bunched Implications and its extension to the "Frame Introduction" rule.

on models (either classical or modal). This notion resulted to be a generalisation of McCarthy's predicate circumscription [63, 60], which in turn is a generalisation of Clark's predicate completion within Logic Programming [78, 19, 20, 80]. The works of Bossu and Siegel [9], of Halpern and Moses [38], of Doyle [26, 15] and variants of Reiter's default logic [79] were also either subsumed by or reformulated in terms of Shoham's notion. This approach has survived almost all competing approaches, leading to the discovery of the filtering technique [84] and to a further generalisation of the main definition [61].

Gabbay's proof-theoretic approach and Shoham's model-theoretic approach are related by Makinson's theorem for which every preferential-entailment relation satisfies the conditions of inclusion, idempotence and transitivity [62, p. 73].

The picture may look nice at this point. There is an established and respectable model-theoretic approach, an established and respectable proof-theoric approach, and there is a formal relation between them. However, non-monotonic reasoning demands for more (see [62]), as not all default logics are preference logics, for example. But there is an higher level of difficulties. Despite the improved nature of the non-monotonic formalisms, it soon became evident that Shoham's criticism was still affecting the field:

> *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 counterexamples. According to the standard research methodology 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.* [86]

According to the example-based methodology, in fact, no general approach is followed, tailor-made reasoning methods are just designed and compared against single examples, and there is no precise knowledge of what is their problem-solving power. A novel methodology was then proposed by Sandewall in 1994 [87, 88]. Although the novel methodology was originally designed against the single example-based methodology, it applies to the proof-theoretic and the model-theoretic approaches too, which design was *not* inspired by single examples, the classification and comparison of the different logics is merely structural (inference rules, preferential entailment methods), and yet there is no precise knowledge of what is their problem-solving power. The general reading is that the connection between the motivating problems and the proposed method can not be tenuous or indirect, as it would pose dangers of over-abstraction and irrelevance.

In what follows we summarise the idea, the definition and the results of Sandewall's meta-theoretic approach to non-monotonic reasoning.

2.2 The idea

The analysis of the reality using models admits a gap among the reality and the models: the solution of a problem always is, as a matter of fact, the solution of the representation we built of the real problem. The model will always be a very limited description of the reality, however it must represent with reasonable accuracy those aspects of interest for the solution of the decisional problem with which we are dealing. Such analysis is then affected by one intrinsic methodological difficulty, namely the provability of the soundness and completeness relations for a given representation. The soundness refers to the accuracy of the representation, namely whether the real problem has been successfully represented within the chosen formal language. The completeness refers instead to the reproducibility of the real problem from the given formal problem, as well as to the recognisability whether the given formal problem is a real problem at all.

The meta-theoretic approach poses itself between the reality and the models, by providing an explicit and formal account of (1) what are the epistemological and ontological assumptions we implicitly involved when representing problems, and (2) what is the relation between the semantics of each logic and the underlying semantics that formally defines the chosen assumptions. The underlying semantics consists of an ego-world game semantics that is meant to capture our intuitions or *simplifying assumptions*. The meta-theory provides a taxonomy of those assumptions, and the different logics can be gathered in congruence classes modulus the particular assumptions they satisfy, in determining their *range of applicability*. The original aim is to provide a precise and uniform account of the various approaches in the field, so that their properties could be easily and concisely compared. Driven by this aim, the underlying semantics is meant to be considerably more transparent and intuitive than the logics themselves. Of course it is not possible to prove that the underlying semantics captures our intuitions correctly, or that the simplifying assumptions are appropriate, however the difficulty of proving the main soundness and completeness relations for a given representation is delayed by one non-trivial step, where the assumptions of the informer/researcher are explicit and formally defined, and the additional difficulty of admitting several formal languages and tailor-made techniques finds an adequate and elegant solution. Within hybrid systems, the underlying semantics is also meant to be related in a mathematically precise way to both the physical description of the world in which the system/robot is operating and to the physical characteristics of the system/robot itself, including its sensory and actuatory systems.

The three pages of section 2.3 summarise the complete definition and basic properties of the broadest class in [87], namely $\mathcal{K}$-**IA**. Section 2.4 summarises the assessment criterion and the assessment results obtained against the $\mathcal{K}$-**IA** class. References therein are meant to invite and facilitate the reading of the original work.

2.3 The definition

The *world* consists in a finite collection of *objects*. Each object may have individual *properties*. Objects may have *relationships* between them. Such properties and relations are referred to as *features*, and they are assumed to be observables. Simple examples of features are "the *colon* of the board is white" and "the block b_i is *on* the block b_j". Objects with a particular property or involved in a relation are arguments for features, like in $color(board)$ and $on(block\_i, block\_j)$. The domain for all time-points is denoted by $\mathcal{T}$ (0,1,2,3,...), the domain for all feature names is denoted by $\mathcal{F}$ ("*color*", "*on*", etc.), the domain for all objects is denoted by $\mathcal{O}$ ("board","*block_j*", etc.), the domain for all feature values is denoted by $\mathcal{V}$ ("white", "true", "false"). An *observation* in a four-tuple h in $\mathcal{H} = \mathcal{T} \times \mathcal{F} \times 2^{\mathcal{O}} \times \mathcal{V}$. A *state* of the world at time τ is a set σ_τ of elements in $\mathcal{S} = \mathcal{F} \times 2^{\mathcal{O}} \times \mathcal{V}$. A *fluent* is a feature traced over time.

Common sense reasoning about the world is understood as a board game between two players: "ego" and "image-level world".

Definition 2.1 (Board) *Let a number of* temporal constant symbols s_i *and* t_i *and a number of* object constant symbols o_i *be given* [6]*. The board consists of the five-tuple* $\langle \mathcal{B}, M, H, \mathcal{P}, \mathcal{C} \rangle$*, called finite development.*

- $\mathcal{B}$ *is a subset of the temporal domain* $\mathcal{T}$*, which in turn is the set of all natural numbers. The smallest and the largest elements of* $\mathcal{B}$ *are respectively* 0 *(the "origo") and* $n_{\mathcal{B}}$ *(the point in time expressing the current "now").*
- M *is a mapping which assigns values in* $\mathcal{B}$ *to some or all of the temporal constant symbols* [7]*, and values in* $\mathcal{O}$ *to all of the object constant symbols.*
- H *is a* history *of the image-level world from the origo of time-points up to* $n_{\mathcal{B}}$*, defined as the mapping* $H : \mathcal{B} \to \mathcal{S}$*.*
- $\mathcal{P}$ *is the* past-action set*; it is a set of tuples* $\langle s, t, E \rangle \in \mathcal{B} \times \mathcal{B} \times \mathcal{E}$*,* $s < t$*, where* s *and* t *represent respectively the start time and the end time of the action* E*.*
- $\mathcal{C}$ *is the* current-action set*; it is a set of tuples* $\langle s, E \rangle \in \mathcal{B} \times \mathcal{E}$ *where* $s \leq n_{\mathcal{B}}$ *is the start time of the action* E*.* ■ *[87, page 18]*

Definition 2.2 (Board revision) *If* J *and* J' *are two finite developments,* J' *is said to be a correct revision of* J *iff the following conditions hold.* $\mathcal{B} \subseteq \mathcal{B}'$ *and if* $b \in \mathcal{B}' - \mathcal{B}$ *then* $b > n_{\mathcal{B}}$*.* $M \subseteq M'$*. The restriction of* H' *to* $[0, n_{\mathcal{B}}]$ *is equal to* H*.* $\mathcal{P} \subseteq \mathcal{P}'$ *and if* $\langle s, t, E \rangle \in \mathcal{P}' - \mathcal{P}$ *then* $t = n_{\mathcal{B}'}$ *and* $\langle s, E \rangle \in \mathcal{C} - \mathcal{C}'$*. Finally, if* $\langle s, E \rangle \in \mathcal{C} - \mathcal{C}'$ *then either* $s = n_{\mathcal{B}}$ *or* $\langle s, n_{\mathcal{B}'}, E \rangle \in \mathcal{P}'$*, and if* $\langle s, E \rangle \in \mathcal{C}' - \mathcal{C}$ *then is* $s = n_{\mathcal{B}}$*.* ■ *[87, page 19]*

Each correct revision reports an update of the history, and the above condition on H implements the solution to the frame problem.

---

[6] The use of the two symbols s_i and t_i is only for convenience reasons, and carries no formal significance.

[7] The reason for allowing a number of undefined temporal constant symbols is that they are going to be assigned values greater than $n_{\mathcal{B}}$, but the history H has not yet reached that far, so that the exact value is still undetermined.

Definition 2.3 (Trajectories) *The image-level world has a set-theoretic description* $\langle Infl, Trajs\rangle$ *given in terms of influenced features and corresponding trajectories. If* E *is an action and* $\sigma_s = H(s)$ *is a state at time* s*,* $Infl(E, \sigma_s)$ *represents the set of those features which may be affected if the action* E *is initiated in* σ_s*, while* $Trajs(E, \sigma_s)$ *represents the set of possible trajectories of* E *if* E *is initiated in* σ_s*, and consists of a set of finite nonempty sequences of partial states each of which assigns values to those features appearing in* $Infl(E, \sigma_s)$ *but not to other features. Each single trajectory defines an action alternative. The tuple* $\langle Infl, Trajs\rangle$ *then consists of a* propositional *table where the description of each trajectory is given in a detailed case by case: if* σ_s *is the state of the world at time* s*, a trajectory in* $Trajs(E, \sigma_s)$ *is a finite nonempty sequence* $\langle \sigma_{s+1}, \ldots, \sigma_t\rangle$ *of partial states from* $s+1$ *up to a certain time-point* t*,* $t \neq s+1$*, where the trajectory ends.* ■ *[87, pages 75,170 and section 3.4]*

Definition 2.4 (Rules of the Game) *The* game *between an ego and the image-level world starts with the board in an initial configuration*

$$\langle \mathcal{B} = \{0\}, M, H(0), \mathcal{P}, \mathcal{C}\rangle$$

where $H(0)$ *is the initial state of the world. Then the players take turns.*

- *The ego can do one of the followings on its turn:*
  - $\mathcal{C}' = \mathcal{C} \cup \{\langle s, E\rangle\}$*,* $s = n_{\mathcal{B}}$
  - $\mathcal{C}' = \mathcal{C} \setminus \{\langle s, E\rangle\}$*,* $\mathcal{P}' = \mathcal{P} \cup \{\langle s, t, E\rangle\}$*,* $t = n_{\mathcal{B}}$
- *The image-level world can do one of the followings on its turn:*
  - *if* $\langle n_{\mathcal{B},E}\rangle \in \mathcal{C}$ *then*
    $\mathcal{B}' = \mathcal{B} \cup \{n_{\mathcal{B}} + k\}$
    $H' = H \triangleright h$ *where* $h = \langle \sigma'_1, \sigma'_2, \ldots, \sigma'_k\rangle \in Trajs(E, H(n_{\mathcal{B}}))$
    $\mathcal{P}' = \mathcal{P} \cup \{\langle n_{\mathcal{B}}, n_{\mathcal{B}} + k, E\rangle\}$
  - *if* $\mathcal{C} = \varnothing$ *then*
    $\mathcal{B}' = \mathcal{B} \cup \{n_{\mathcal{B}} + 1\}$
    $H' = H \triangleright \langle\varnothing\rangle$
    $\mathcal{P}' = \mathcal{P}$

The update operator $\triangleright$ *is defined as follows: if* H *is an history over* $\mathcal{B} = [0, t]$ *and* $h = \langle \sigma'_1, \sigma'_2, \ldots, \sigma'_k\rangle$ *is a trajectory of* E*, then* $H' = H \triangleright h$ *is the updated (extended) history over* $\mathcal{B}' = [0, t+k]$ *such that:*

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

The operator $\oplus : \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ *is such that* $\sigma \oplus \sigma'$ *is the state* $\sigma^\star$ *where* $\sigma^\star(f) = \sigma'(f)$ *if* $\sigma'(f)$ *is defined, and* $\sigma^\star(f) = \sigma(f)$ *otherwise. In particular* $H \triangleright \langle\varnothing\rangle$*, where* $\langle\varnothing\rangle$ *is the null trajectory, extends* H *from* t *to* $t+1$ *such that* $H(t+1) = H(t)$*.* ■ *[87, pages 19,20,74,76,78]*

The game is non-deterministic because of the initial configuration of the board and the selection of the trajectory h from $Trajs$.

Proposition 2.1 *The rules of the game correctly revise the board.*

Definition 2.5 (Reasoning Problem) *A* scenario *for W is a set of games where W was the world player. A* scenario description *in $\mathcal{K}$-$\mathbf{IA}$ is the description of W itself and of how its state has changed over time, and it consists of the five-tuple $\Upsilon \equiv \langle \mathcal{K}, \mathcal{O}, \langle \mathbf{IA}, LAW \rangle, SCD, OBS \rangle$.*

- *$\mathcal{K}$ is the assumed* epistemological assumption*, defined as follows: (1)* elementary actions *are allowed, namely the schedule specifies each action in terms of action symbol, arguments, starting and ending time; (2) there are no other actions besides those mentioned in the schedule; (3) action laws must explicitly specify all the features they may influence (an empty set of postconditions is not allowed), together with a possibly empty set of preconditions for their applicability; (4) for every combination of action designator and initial state, there is a definition for how to perform the action from that state.*
- *$\langle \mathbf{IA}, LAW \rangle$ is the* logical description *of the set-theoretic description $\langle Infl, Trajs \rangle$ of the world. The component $\mathbf{IA}$ is the assumed family of ontological assumptions, for which actions may have alternative results ($\mathbf{A}$) and the following property applies ($\mathbf{I}$):*

  Strict inertia: The state of the world persists unchanged until an action occurs that causes a change, therefore the transition to a new state. Features of objects in the world do not change their value unless under the direct and known effect of an action.

  The component LAW consists of the direct translation of the information in the table $\langle Infl, Trajs \rangle$, as a set of action laws *in* Full Trajectory Normal Form *(this is analogous to how Full Disjunctive Normal Form in propositional logic corresponds to a truth-table).*
- *SCD, the* schedule*, corresponds to the $\mathcal{P}$ component of the board at the end of all games. OBS is a finite set of observations.* ■ *[87, pages 23,26,42,44,160 and section 8.2]*

Definition 2.6 *The classical model set for Υ in $\mathcal{K}$-$\mathbf{IA}$ is defined as follows:*

$$[\![\Upsilon]\!] = [\![LAW[SCD] \cup OBS]\!]$$

where $LAW[SCD]$ is the expansion of all formulae in SCD using their corresponding definitions in LAW. ■ *[87, pag. 185]*

Definition 2.7 *The intended model set for Υ in $\mathcal{K}$-$\mathbf{IA}$ is defined as follows:*

$$\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) = \{\langle M, H \rangle \mid \langle \mathcal{B}, M, H, \mathcal{P}, \mathcal{C} \rangle \in Mod(\Upsilon)\}$$

where $Mod(\Upsilon)$, called complete development set, is obtained as follows: (1) construct the unique image-level world W that is precisely specified by the logical description $\langle \mathbf{IA}, LAW \rangle$, then select an arbitrary ego and an arbitrary initial state $H(0)$; (2) generate all possible developments which can be obtained in games between them; (3) add an arbitrary M component to each development; (4) restrict the set of developments to those where all formulas in $SCD \cup OBS$ are satisfied, and where there is a one-to-one correspondence between actions in the development and action statements in SCD. ■ *[87, pages 46,180,182-185]*

Proposition 2.2 $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}()$ *is a non-monotonic logical operation.*
PROOF. *For any distinct Υ and Υ' in $\mathcal{K}$-$\mathbf{IA}$, $SCD \subseteq SCD'$ does not imply $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) \subseteq \Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon')$.* ■ *[87, pag. 194]*

Proposition 2.3 $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) \subseteq [\![\Upsilon]\!]$. PROOF. *The classical model set $[\![\Upsilon]\!]$ allows models which do not satisfy inertia and models where there are additional actions besides those specified in SCD.* ■ *[89, p. 40] [87, pag. 186]*

2.4 The results

As the intended model set is a subset of the classical (Tarski) model set (proposition 2.3), preferential entailment methods are a suitable model-theoretic machinery to reconcile them.

Definition 2.8 (Preference Logic) *Let $\mathcal{L}$ be a standard logic, namely a logic with the usual compositional model-theoretic semantics.*

- *Let $\ll$ be a strict partial order on interpretations for $\mathcal{L}$. Intuitively $I_1 \ll I_2$ will mean that the interpretation I_2 is* preferred *over the interpretation I_1. $\mathcal{L}$ and $\ll$ define a new logic $\mathcal{L}_\ll$. We will call such logics* preference logics.
- *Let α, β be in $\mathcal{L}$. α* preferentially entails *β, written $\alpha \models_\ll \beta$, if for any M, if $M \models_\ll \alpha$ then $M \models \beta$, or, equivalently, if the models (preferred and otherwise) of β are a superset of the preferred models of α.*
- *$\mathcal{L}_\ll$ is* monotonic *if for all $\alpha, \beta, \gamma \in \mathcal{L}$, if $\alpha \models_\ll \gamma$ then also $\alpha \wedge \beta \models_\ll \gamma$.* ■

Definition 2.9 (Range of Applicability) *Let $\mathcal{L}_\ll$ be a preference logic, let Υ be a reasoning problem and let ζ be the mapping defined in terms of $\ll$ that selects those members of $[\![\Upsilon]\!]$ which are minimal according to $\ll$, so that the maximally preferred models are the selected ones:*

$$\zeta([\![\Upsilon]\!]) = Min(\ll, [\![\Upsilon]\!]) = \{I \in [\![\Upsilon]\!] \mid \neg\exists\, I' \,.\, I' \in [\![\Upsilon]\!] \wedge I' \ll I\}$$

$\mathcal{L}_\ll$ is correct *for Υ iff $\zeta([\![\Upsilon]\!]) = \Sigma\,(\Upsilon)$. The* range of applicability *of $\mathcal{L}_\ll$ is the class of all Υ for which $\mathcal{L}_\ll$ is provably correct.* ■

The assessment of the range of applicability is meant to guarantee that for a certain class of reasoning problems, the logic always gives the correct set of conclusions. The problem of establishing whether a given preference logic solves a given reasoning problem is then reduced to saying whether the problem itself belongs to the class for which the logic is provably correct. A classification of reasoning problems is available in [87, page 168].

In order to determine whether a certain range of applicability can be improved, one needs to investigate about its upper bound. A given range of applicability is, in principle, a lower-bound class, while the upper-bound class is a set of reasoning problems such that $\zeta([\![\Upsilon]\!]) \neq \Sigma(\Upsilon)$ for any Υ outside the set. When the upper-bound class equals the lower-bound class, that is the logic is sound and complete with respect to the underlying semantics that formally defines that precise class of reasoning problems, then the obtained range is exact and no additional improvement is possible.

In [87] Sandewall reports that his review of reasoning methods from the literature lead to the definition of five preferential entailment methods which upper-bound were proven to be the $\mathcal{K}$**r-IA** subclass [8] of $\mathcal{K}$-**IA** [87, page 288]. Attempts to design methods of broader range of applicability resulted in the following proposition and the discovery of the filtering technique.

Proposition 2.4 *There does not exist any selection function ζ which satisfies $\zeta([\![\Upsilon]\!]) = \Sigma_{\mathcal{K}\text{-}\mathbf{IA}}\,(\Upsilon)$ for all Υ in $\mathcal{K}$-**IA**.*
PROOF. *Consider a world where there is a single propositional feature p*

[8] The interested reader may refer to [87, pages 42,163,166] for a list of epistemological and ontological sub-characteristics of $\mathcal{K}$-**IA**. Their definition constrain the main definition 2.5, so to define subclasses of $\mathcal{K}$-**IA**.

and a single action A whose effect is like tossing a coin, i.e. the truth value of p after the action is non-deterministically assigned, regardless of its value when the action starts. Let Υ be the scenario description where both SCD and OBS are empty, and let Υ' be the scenario description where OBS is empty and SCD consists of the single formula $[4,5]A$. Then $[\![\Upsilon]\!] = [\![\Upsilon']\!]$ but $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) \neq \Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon')$. ■ *[87, p. 187]*

The conceptual basis of the above method is the application of the preference relation on the whole set of premises Υ. The negative result by proposition 2.4, and in particular its proof, suggests the *filtering technique*, namely the partitioning of the premises and the application of different selection functions to the classical model set of the resulting and distinct sets of premises. The set of selected models is then chosen by Filter Preferential Entailment (FPE) as $Min(\ll, [\![LAW[SCD]]\!]) \cap [\![OBS]\!]$, so that the minimisation is imposed *before* the observations. FPE was first described by Sandewall in 1989 [84]. Crawford and Etherington later suggested that such a separation of system description formulae and observation formulae is consistent with the working practice of Qualitative Reasoning [21].

Proposition 2.5 *[87, page 214]*

$$Min(\ll, [\![LAW[SCD]]\!]) \cap [\![OBS]\!] \subseteq Min(\ll, [\![LAW[SCD] \cup OBS]\!])$$

Corollary 2.1 *FPE is monotonic with respect to OBS, so that if SCD and OBS filter preferentially entail α, and $OBS \subseteq OBS'$, then SCD and OBS' also filter preferentially entail α. FPE is non-monotonic with respect to SCD. [87, page 214]*

Seven methods were designed according to the filtering technique [87, page 298]. Among them, the Pointwise Minimisation of Occlusion with Nochange Premises (PMON) is the simplest to be correctly applicable to $\mathcal{K}$-**IA**.

Is the meta-theoretic approach applicable to preference logics only? The following assessment shows that Logic Programming does benefit from it too.

Proposition 2.6 *Abductive Logic Programming with integrity constraints is sound and complete with respect to the $\mathcal{K}$-**IbsAd** class.*
PROOF. *In [23, 25] the Abductive Logic Programming with integrity constraints has been proven sound and complete with respect to the action description language A [34]. The result consists in a sound and complete transformation from A domain descriptions to* open logic programs *with integrity constraints, where the reasoning procedure adopted for the resulting programs is the SLDNFA Resolution Rule [24]. In [98] the language A has been proven sound and complete with respect to the $\mathcal{K}$-**IbsAd** class. The thesis holds per transitivity.* ■

Proposition 2.6 first appeared in [10]. The interested reader may refer to [59] for the computational complexity of the language A.

The introduction to the novel approach is now complete. If the reader feels persuaded and confident, he/she is now ready to take active part in it. The definition of an underlying semantics will follow the given pattern, unless evidence is given for a better approach than the ego-world game. The assessment of a logic, on the other hand, does not follow any specific pattern and the study of the available assessments is not necessarily useful in this respect, although reading such proofs is good practice and gives a great deal of insight on the assessed logic. What is left is general practice.

3 Case Study: Classical Physics

The underlying semantics that formally defines the broadest class in the Features and Fluents taxonomy, defines the knowledge representation method and the intended model set for each reasoning problem where accurate and complete information about actions ($\mathcal{K}$), together with *strict inertia* in *integer time* and alternative results of actions (**IA**) are the assumed simplifying assumptions. In order to make an explicit distinction between the reality and the models, two levels of description for a system were assumed: the material-level and the image-level [87, page 3]:

> *The material-level consists in the description of the physical system, or can be even identified with it, and uses continuous time and continuous-valued state variables that change according to the laws of Physics.*
>
> *The image-level consists in the abstract representation of the material-level and models the way the world is perceived by the agent; it is obtained from the material-level by means of a perception (abstraction) function, and uses discrete-valued state variables.*

In the first part of this section, the image-level is extended to the case of continuous-valued state variables. The assumed simplifying assumptions are as follows: accurate and complete information about actions ($\mathcal{K}$), strict inertia in continuous time, continuous change and alternative results of possibly concurrent and independent actions (**RACi**). In the genuine style of the underlying semantics for $\mathcal{K}$-**IA**, the set of all the intended answers for each reasoning problem in $\mathcal{K}$-**RACi** is built simulating infinite game developments between a trajectory-semantics ego and a trajectory-semantics world. The new underlying semantics subsumes the underlying semantics for $\mathcal{K}$-**IA** and admits a trajectory-semantics world with explicit continuous time and a combination of encapsulated and/or non-encapsulated actions with discrete and/or continuous change, depending on the need. The definition of $\mathcal{K}$-**RACi** first appeared in [10].

It was said: "of course it is not possible to prove that the underlying semantics captures our intuitions correctly, or that the simplifying assumptions are appropriate". As classical Newtonian Physics implicitly involves a number of epistemological and ontological assumptions, is it possible to enumerate all these assumptions and capture them by means of ego-world game developments in Sandewall's style? In the second part of this section, exploiting the theorem from Augustin Louis Cauchy on the solution of first-order linear differential equations, the subclass $\mathcal{K}$**sp-RAdCi** of $\mathcal{K}$-**RACi** is shown to include the epistemological and ontological assumptions that were implicitly involved by Newton in his three principles of Mechanics (proposition 3.5), the same Mechanics that the test of time has proven to be appropriate for explaining a good part of the physical world as we commonly perceive it. The boundaries of the result are outlined.

3.1 The definition

The western philosophy dates back to Aristotle the notion of time as measurable order of the movement. According to Aristotle (*Physics*, IV, 11, 219 b 1), time is not change in itself, but *the measure of change according to a before and an after*, where movement is assumed as primary notion.

Since time evolves because of change, the absence of change determines 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 [77] for which time has a growth direction.

The notion of time for the $\mathcal{K}$-**IA** world is Aristotelian. In fact: (1) when an action is initiated by the ego, that is the current-action set is not empty, then the world increases the current *now* by k time-points, where k is the duration of the invoked action; (2) when no action is initiated by the ego, that is the current-action set is empty, then the world initiates, executes and terminates a null action which trajectory is the null trajectory $\langle\varnothing\rangle$ of duration 1, so that the current *now* is increased by one time-point. Therefore, executed actions are the image-level justification of flowing time and determine, de facto, the way the image-level world beats time. Concerning the direction of time, actions may also have reversible effects on feature values; in fact, due to the correct revision, if $b \in \mathcal{B}' - \mathcal{B}$ then is $b > n_{\mathcal{B}}$.

In pursuing the aim of extending the given notion of time to the continuum case, we meet two problems.

The first problem is that no action can be physically initiated and terminated such that its duration is infinitesimal and its repeated execution is the image-level justification to continuous time. The world we perceive is fully discrete, and time and continuity are abstractions of our mind. At the image-level the problem with beating the continuous time *formally* consists in the well known non-existence of the successor function for real numbers, as the image-level world could just advance time by increasing the current *now* by an infinitesimal quantity ε, and lose a continuum of time-points between the *now* $n_{\mathcal{B}}$ and $n' = n_{\mathcal{B}} + \varepsilon$. We then approach the problem from St. Augustine's viewpoint [5, ch. 24,26]; movement is not assumed as primary notion, it is time that permits to describe it as velocity (in mathematical terms, as a continuous function of time).

The second problem, which is tightly connected with the first one, consists of the famous paradox of Zeno of Elea (Italy, 489 B.C.), which simply occurs as follows: the trajectory of a running action can never describe the continuous change during the scheduled time interval. The issue can be characterised as follows. The image-level world (definition 2.3) has a set theoretic description given in terms of influenced features and corresponding trajectories. For a given action E and state σ_s, $Infl(E, \sigma_s)$ represents the set of those features which may be affected if the action E is initiated in the state σ_s, while $Trajs(E, \sigma_s)$ represents the set of possible trajectories of the action E if the action E is initiated in σ_s and consists in a set of finite nonempty sequences of partial states each of which assigns values to those features appearing in $Infl(E, \sigma_s)$ but not to other features. Due to the assumption of *strict inertia* in *discrete (integer) time*, the set theoretic description consists in a propositional table where the description of each trajectory is given in a detailed case by case: if σ_s is the state of the world at time s, a trajectory in $Trajs(E, \sigma_s)$ is a finite nonempty sequence $\langle \sigma_{s+1}, \ldots, \sigma_{s+k} \rangle$, $k \sqsupset 1$. For example, the image-level world for the celebrated Yale Shooting Scenario [39, 40] is as follows:

Initial state σ	$Infl(Fire, \sigma)$	$Trajs(Fire, \sigma)$
$\{a:\mathtt{T}, l:\mathtt{T}, p:\mathtt{T}\}$	$\{a, l\}$	$\{\langle\{a:\mathtt{T}, l:\mathtt{F}\}, \{a:\mathtt{F}, l:\mathtt{F}\}\rangle\}$
$\{a:\mathtt{T}, l:\mathtt{T}, p:\mathtt{F}\}$	$\{a, l\}$	$\{\langle\{a:\mathtt{T}, l:\mathtt{F}\}, \{a:\mathtt{F}, l:\mathtt{F}\}\rangle\}$
$\{a:\mathtt{T}, l:\mathtt{F}, p:\mathtt{T}\}$	$\varnothing$	$\{\langle\varnothing\rangle\}$
$\{a:\mathtt{T}, l:\mathtt{F}, p:\mathtt{F}\}$	$\varnothing$	$\{\langle\varnothing\rangle\}$
$\{a:\mathtt{F}, l:\mathtt{T}, p:\mathtt{T}\}$	$\{l\}$	$\{\langle\{l:\mathtt{F}\}\rangle\}$
$\{a:\mathtt{F}, l:\mathtt{T}, p:\mathtt{F}\}$	$\{l\}$	$\{\langle\{l:\mathtt{F}\}\rangle\}$
$\{a:\mathtt{F}, l:\mathtt{F}, p:\mathtt{T}\}$	$\varnothing$	$\{\langle\varnothing\rangle\}$
$\{a:\mathtt{F}, l:\mathtt{F}, p:\mathtt{F}\}$	$\varnothing$	$\{\langle\varnothing\rangle\}$

According to the first statement, the firing action takes place in two time units; at the fist time-unit the gun becomes unloaded ($l:\mathtt{F}$) and the turkey is still alive ($a:\mathtt{T}$), while at the second time-unit the gun remains unloaded and the turkey dies ($a:\mathtt{F}$). If we assume the value domain for features is a continuous domain, then the set-theoretic description of the image-level world results in a continuum of possible initial states, together with a tuple $\langle Infl(E,\sigma), Trajs(E,\sigma)\rangle$ for each initial state σ and action E, and a continuum of ordered partial states for each trajectory in $Trajs(E,\sigma)$. Such set-theoretic description does not fit the above propositional table.

The given ontology falls beyond $\mathcal{K}$-**IA**.

3.1.1 Continuous Time

Let $clock^0$ (image-level master-clock) be a strictly monotonic rising function. *clock* is defined either as an *internal* or as an *external* function. An internal clock is, we assume, a computable function. In being internal, the agent is rationally aware of the flux of time, so to know exactly how the time is beaten. An external clock has no internal formal definition and may also be a non-computable function; it is simply invoked and the actual value, i.e. the "now", is obtained for magic. In being external, the agent is still aware of the flux of time but has no *given* understanding of how the time is beaten. An example of internal clock is the following: *clock* is initially zero, then

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

which is how the time is beaten in $\mathcal{K}$-**IA**. An example of external clock is the function which reads the clock pulse of a computer hardware; in this case the resulting underlying semantics is hybrid, since at least this extra-logical component is involved. The function beating the continuous time is purely theoretical; it is clearly not computable, and therefore external.

Definition 3.1 *The* basic time structure *is* $\langle\mathcal{T}; \{+^2, clock^0\}, \{=^2, \sqsubseteq^2\}\rangle$, *defined with the obvious first order language* $\mathcal{L}_\Sigma$ *for time-points, built on top of the signature* $\Sigma = \{\{+^2, clock^0\}, \{=^2, \sqsubseteq^2\}\}$. *We assume* $t_0 \in \mathcal{T}$.

- *Axioms for the sum function. For all pair of time-points* s *and* t *in* $\mathcal{T}$ *is defined a third time-point in* $\mathcal{T}$, *written* $s+t$, *such that:*

  S1 **:** $\forall\, r, s, t \in \mathcal{T}\ .\ (r+s)+t = r+(s+t)$

  S2 **:** $\forall\, s, t \in \mathcal{T}\ .\ s+t = t+s$

  S3 **:** $\forall\, t \in \mathcal{T}\, .\, t+t_0 = t$

- *Axiom for the clock function. For all pairs of subsequent calls of clock, ν and ν' in $\mathcal{T}$,*

  C1 : $\nu \sqsubset \nu'$

- *The equality is a symmetric relation, that is*

  E1 : $\forall\, s,t \in \mathcal{T} \,.\, (s = t) \rightarrow (t = s)$

- *For all pair of time-points s and t, it is possible to establish whether s precedes t ($s \sqsubseteq t$) or t precedes s ($t \sqsubseteq s$) according to the following axioms for the order predicate:*

  O1 : $\forall\, t \in \mathcal{T} \;.\; t \sqsubseteq t$

  O2 : $\forall\, t \in \mathcal{T} \;.\; t_0 \sqsubseteq t$

  O3 : $\forall\, s,t \in \mathcal{T} \;.\; s \sqsubseteq t \wedge t \sqsubseteq s \Rightarrow s = t$

  O4 : $\forall\, r,s,t \in \mathcal{T} \;.\; r \sqsubseteq s \wedge s \sqsubseteq t \Rightarrow r \sqsubseteq t$

  O5 : $\forall\, s,t \in \mathcal{T} \;.\; s \sqsubseteq t \vee t \sqsubseteq s$

  O6 : $\forall\, r,s,t \in \mathcal{T} \;.\; r \sqsubseteq s \rightarrow r + t \sqsubseteq s + t$

  O7 : *Let A and B be non empty subsets of $\mathcal{T}$ such that $a \sqsubseteq b$ for all $a \in A$ and $b \in B$. Then exists $\xi \in \mathcal{T}$ such that $a \sqsubseteq \xi \sqsubseteq b$ for all $a \in A$ and $b \in B$.*

S1 is the usual axiom for the associativity, S2 for the commutativity and S3 for the neuter element. O1 is the usual axiom for the reflexivity, O3 for the anti-symmetry, O4 for the transitivity, O5 (optional) for the totality and O7 (optional) is Cantor-Dedekind's axiom of completeness. The neuter element t_0 is the origo *of time-points. The relation $s \sqsubset t$ is an abbreviation for $s \sqsubseteq t \wedge s \neq t$. The relation $s \sqsubseteq t$ ($s \sqsubset t$) is equivalent to $t \sqsupseteq s$ ($t \sqsupset s$).* ∎

If we assume axiom *O5* and *clock* as the successor function of Peano's arithmetic (recursively defined as $clock = clock + 1$, where $t_0 = 0$ is the origo of time-points and initial value for *clock*, and 1 is the duration of the null trajectory), the resulting structure is the time structure for $\mathcal{K}$-**IA**. In what follows we assume *O7* and *O5*, i.e. $\mathcal{T}$ continuous and linearly ordered; the resulting time structure is the classical structure $\mathcal{R}^+$ for the positive real numbers augmented with the axiom $C1$ for the master clock function.

3.1.2 Continuous Change

The following simplifying assumptions are adopted: (1) for all $E \in \mathcal{E}$ and $\sigma \in \mathcal{S}$, $Infl(E, \sigma)$ is a finite set (each action must influence a finite number of features), (2) as $Infl$ is a set, if $f_i, f_j \in Infl(E, \sigma)$ then $f_i \neq f_j$, (3) for all $E \in \mathcal{E}$ there is a finite number of alternatives and a finite number of initial states σ such that $Infl(E, \sigma) \neq \varnothing$.

From the apriorism, a finite number of ordered partial states defines a discrete trajectory, that can be also explicitly written if reasonably small. Clearly a continuum of ordered partial states cannot be explicitly written, unless with a characteristic property that *implicitly* defines them by selection from a common domain. Because of the nonexistence of a "next partial state", the order relation between partial states must necessarily be the order relation between time-points. The order relation $\sqsubseteq$ applies as follows on members of $\mathcal{H}$: for all $h_1 = \langle t_1, f_1, \{\ldots\}, v_1 \rangle$ and $h_2 = \langle t_2, f_2, \{\ldots\}, v_2 \rangle$ in $\mathcal{H}$, $h_1 \sqsubseteq h_2$ iff $t_1 \sqsubseteq t_2$. As we assumed *O5*, the set $(\mathcal{H}; \sqsubseteq)$ is a totally ordered set. How do we describe trajectories?

Definition 3.2 *A single-feature trajectory descriptor, or partial fluent, is a function $\varphi\colon \mathcal{T}^3 \to \mathcal{V}$ which domain is $Dom(\varphi) = \{\langle s, \tau, t\rangle \in \mathcal{T}^3 : s \sqsubseteq \tau \sqsubseteq t\}$. The duration of $\varphi(s, \tau, t)$, where $\varphi(s, \tau, t)$ is the invocation of the partial fluent φ, is that unique member l of $\mathcal{T}$ such that $s + l = t$.* ■

Partial fluents can be internal and static, internal and dynamic, or external of two different species.

An internal static partial fluent is a computable function which definition is known and does not change passing time. A partial fluent of this type is, for example, a mathematical function $\varphi : \mathbb{R}^{3+} \to \mathbb{R}$, or a function $\varphi : \mathbb{N}^3 \to \{\mathtt{T}, \mathtt{F}\}$. For a given feature f_j such that several actions are concurrently running and influencing f_j, instead of separately consider all involved partial fluents we can just consider their combined function φ_j, like in Newtonian Mechanics. This type of partial fluent also allows to simulate lack of knowledge when complete knowledge about the change is not required, or the change is not exactly known; for example

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

where w is the expected value of a precondition involving the feature f_j, *unknown* represents $\varkappa f_j$ (read "occluded f_j"), and v is the terminal value of the postcondition. The tuple $\langle t, f, unknown\rangle$ is an abbreviation for $\bigvee_i \langle t, f, \{\ldots\}, v_i\rangle$, varying i over all possible tuples $\langle t, f, \{\ldots\}, v_i\rangle$ in $\mathcal{H}$ (assuming a continuous value domain for features, i is necessarily a continuous index), and express the *don't know* non-determinism.

An internal dynamic partial fluent is a computable function which definition is known and can only change according to internal criteria. The function may be updated, for instance, during a learning process.

An external partial fluent of first specie is a computable function which definition is known and is allowed to change passing time according to external criteria. An example is the following. Suppose the function $\varphi_i : \mathbb{R}^3 \to \mathbb{R}^3$ represents the i-th airway within a certain airspace. The invocation $\varphi_i(s, \tau, t)$, varying τ in $[s, t]$, represents the trajectory of an airplane, and is generated by an air-traffic scheduler, which is supposed to be "external" to the airplane co-pilot system, so that for all τ in $[s, t]$ no other airplane is using the i-th airway. The co-pilot system "knows" φ_i at time τ, and may drive the airplane accordingly.

An external partial fluent of second specie is a partial function which definition is not necessarily known, and may also be a non computable function. An example is the non-predictable trajectory of an hostile aircraft.

As mentioned in the introduction, the image-level consists in the abstract representation of the material-level, and is obtained from the material-level by means of a perception function, defined as follows:

> "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 assumptions: *(1)* the argument and value are histories over the same interval; *(2)* if H is a history over $[0, t]$, $s < t$, and $H_{[0,s]}$ is the restriction of H to the interval $[0, s]$, then $Perc(H)_{[0,s]} = Perc(H_{[0,s]})$." (Sandewall, [87])

According to the definition, both the image-level and the material-level are given in terms of histories. However, we need observational data in order to make the model realistic. As Brooks suggested in [13], the real world should play instead of its description and the agent must sense the environment often, so to have an up-to-date idea of what is happening in the world. If the material-level is to be identified with the physical system, then no formal histories are available and no mapping $Perc$ is possible. A different definition is then needed. Just like an human being is able to focus and release the attention on a particular feature of a particular object, there is no reason why an agent could not similarly initiate and terminate a "sense" action for the particular feature of interest. We will refer to this ability with the term "hi-level sensing", so to distinguish it from the low-level one consisting of the mere perception of raw data through the physical sensors. The key instrument for hi-level sensing is the external partial fluent of the second specie, here adopted as perception function. If we assume to deal with robots where only distances, bumps and a directional beacon are registered by the sensors, for example, then hi-level sensing and low-level sensing may be considered as being the same process, provided no noise is involved. The perception function will then use the raw-data provided by the sensor as value for its corresponding feature. If complex data are involved, like the image of the world through an infrared camera, or noise in the sensors must be taken into account, then hi-level sensing will consist in the interpretation of the raw-data provided by the low-level sensing process; the perception function will then serve as such an interpreter, whatever the definition. At start-up time, the typical robot will have some or all of its sensors switched on, and will then start receiving data through them. In the underlying semantics this corresponds to saying that at the beginning of the game the ego will initiate a "sense" action for some or all of the sensors of its vehicle, according to an agreed mapping between sensors and features, and each sensing action will run concurrently and independently with the others, reading its sensor and updating the value of the corresponding feature. If a sensor dies, the corresponding "sense" action(s) will be terminated.

Definition 3.3 (Trajectories) *Let E be an action and σ_s a state of the world at time s such that $Infl(E,\sigma_s) \neq \varnothing$, where $Infl(E,\sigma_s)$ represents the set of those features which may be affected if the action E is initiated in σ_s. Let n be the number of alternatives of the action E initiated in σ_s and let $\varphi_{1j},\ldots,\varphi_{ij},\ldots,\varphi_{nj}$ be the partial fluents that are associated to each feature f_j in $Infl(E,\sigma_s)$. Furthermore, for all φ_{ih} and φ_{ik} associated to f_h and f_k in $Infl(E,\sigma_s)$, where $i=1,\ldots,n$, is $s \notin Dom(\varphi_{ih})$, $s \notin Dom(\varphi_{ik})$ and $duration(\varphi_{ih}) = duration(\varphi_{ik})$. The i-th trajectory of E initiated in σ_s is then defined as the nonempty totally-ordered set $(T_i(E,\sigma_s);\sqsubseteq)$, where*

$$T_i(E,\sigma_s) = \{\langle \tau, f_j, \{\ldots\}, \varphi_{ij}(s,\tau,t)\rangle \in \mathcal{H} : \tau \in Dom(\varphi_{ij}), f_j \in Infl(E,\sigma_s)\}$$

Given a trajectory $T_i(E,\sigma_s)$, $h(\tau) = \{\langle t, f, \{\ldots\}, v\rangle \in T_i(E,\sigma_s) : t = \tau\}$ represents the partial state of the world at time τ according to $T_i(E,\sigma_s)$. If $s = Min\{\tau : \langle \tau, f, \{\ldots\}, v\rangle \in T_i(E,\sigma_s)\}$ and $t = Max\{\tau : \langle \tau, f, \{\ldots\}, v\rangle \in T_i(E,\sigma_s)\}$, the duration *of the trajectory $T_i(E,\sigma_s)$ is defined as the unique member l of $\mathcal{T}$ such that $s + l = t$. The set $Trajs(E,\sigma)$ is then defined as the set of all possible trajectories of E initiated in σ.* ■

Definition 3.4 (Action Laws) *A* <u>*trajectory formula*</u> *for a given feature f_j in $\mathcal{F}$ is the first-order formula $\forall \tau \in [s,t] \subset \mathcal{T} \,.\, [\tau] f_j \doteq \varphi_j(\tau)$ where the descriptor φ_j is a partial fluent defined over $D \subseteq [s,t] \subset \mathcal{T}$, and $s \neq t$. Let M*

be a mapping and $\theta = \{s/M(\mathbf{s}), t/M(\mathbf{t})\}$ *a valuation. The meaning of a trajectory formula* $\forall \tau \in [s,t] \subset \mathcal{T} . [\tau] f_j \doteq \varphi_j(s,\tau,t)$ *is the lattice* $(S_j(\theta); \sqsubseteq)$, *where* $S_j(\theta) = \{\langle \tau, f_j, \{\ldots\}, \varphi_j(s\theta, \tau, t\theta)\rangle \in \mathcal{H} : \tau \in Dom(\varphi_j)\}$. *In what follows,* S *will represent a trajectory formula and* $S(\theta)$, *for a given valuation* θ, *will represent its meaning. An action law in* Full Trajectory Normal Form *with no alternatives is a mapping of an action statement to a conjunction of trajectory formulas*

$$[s,t]A \Rrightarrow \bigwedge_{j=1}^{m} S_j$$

where $A \in \mathcal{E}$, *the conjunction* $\bigwedge_{j=1}^{m} S_j$ *represents a trajectory set, and each* S_j *is the corresponding trajectory formula for the feature* f_j. *The Full Trajectory Normal Form for actions with alternative results is the mapping*

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

where 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 formulas* S_{ij}, *each of which corresponds to the feature* f_j *in the alternative* i.

The following conventions are imposed. Within the same action law, alternatives involve the same set of features and partial fluents have the same duration, which defines the duration of the action itself. Each action type has influence over those features occurring in the action definition, while all other feature values remain unchanged. ■

Proposition 3.1 *The FTNF subsumes the FTNF for the discrete case.*

PROOF. Let Pre be a conjunction of statements like $[s]f_n = v_n$, and let $Post$ be a conjunction of statements like $[s,t]f_m := v_m$; the former represents the initial expected values for all the features involved in a precondition, while the latter describes the trajectory for all the features involved in a postcondition. The FTNF for the discrete case is as follows:

Basic case:	$[s,t]A \Rrightarrow [s,t]f := v$
Random effect:	$[s,t]A \Rrightarrow [s,t]f := v \vee [s,t]f := \neg v$
Simple conditional:	$[s,t]A \Rrightarrow Pre \Rightarrow Post$
Disjunctive preconditions:	$[s,t]A \Rrightarrow (Pre_1 \vee \ldots \vee Pre_k) \Rightarrow Post$
Disjunctive postconditions:	$[s,t]A \Rrightarrow Pre \Rightarrow (Post_1 \vee \ldots \vee Post_k)$
Generic alternative:	$[s,t]A \Rrightarrow C_1 \vee \ldots \vee C_k$

where the generic alternative is a disjunction of C_i statements, each of which is like the body $Pre \Rightarrow Post$ of a simple conditional, where Pre may also be empty, while an empty set of postconditions is not allowed, and each Pre_i in C_i and Pre_j in C_j involve the same set of features.

The basic case is trivial. The random effect is that particular instance of the generic alternative for which two or more distinct values are non-deterministically assigned to the same feature. In order to avoid compromising the scenario description's consistency, the disjunction is intended to be exclusive. The conditional with disjunctive preconditions is an alternative writing for a generic alternative where each C_i has the same postcondition $Post$. The conditional with disjunctive postconditions is an alternative writing for a generic alternative where each C_i has the same precondition Pre.

The simple conditional, by exploiting partial fluents expressiveness, may be rewritten as follows: (1) write a partial fluent φ_n for each element $[s]f_n = v_n$ in Pre; (2) write a partial fluent ψ_m for each element $[s,t]f_m := v_m$ in Post; (3) for each partial fluent ψ_m such that there exists a φ_n for the same feature, *complete* ψ_m with the definition of φ_n then eliminate φ_n; (4) conjunct all obtained partial fluents. ■

Example 3.1 *The action law*

$$[s,t]Fire \Rrightarrow [s]l \Rightarrow ([s,t]l := false \wedge [s,t]a := false)$$

is rewritten as

$$[s,t]Fire \Rrightarrow \forall\, \tau \in [s,t]\,.\,[\tau]a \;\check{=}\; to\_die(s,\tau,t) \wedge [\tau]l \;\check{=}\; to\_fire(s,\tau,t)$$

where the partial fluents to_die and to_fire are defined as follows:

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

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

Definition 3.5 (Rules of the Game) *The game between an ego and the image-level world starts with the board in an initial configuration:*

$$\langle [0, clock], M, H(0), \mathcal{P}, \mathcal{C} \rangle$$

where $H(0)$ is the initial state of the world. Then the clock function starts beating time.

- *At time $\nu = clock$ the ego communicates its decision to initiate a certain action or activity E, by adding the tuple $\langle \nu, E \rangle$ to $\mathcal{C}$ (we assume synchronous communication between the ego and the world). Then the move of the world consists in modifying the board as follows:*

  1. *At time ν, the world non-deterministically selects a trajectory T_i from the trajectory set $Trajs(E, H(\nu))$.*
  2. *From ν to $\nu' = \nu + l$, $l = duration(T_i(E, H(\nu)))$, the world's state changes according to the chosen T_i, that is $H(\rho, f_j) = \varphi_{ij}(\nu, \rho, l)$ for all ρ in $Dom(\varphi_{ij}(\nu, \rho, l))$.*
  3. *At time ν' the world adds the tuple $\langle \nu, \nu', E \rangle$ to $\mathcal{P}$ and resets $\mathcal{C}$ to the empty set (this communicates a "terminated" message to the ego).*

  The ego can also terminate E earlier than ν', let say at the present time $\nu'' = clock \sqsubset \nu'$, by removing $\langle \nu, E \rangle$ from $\mathcal{C}$ and adding $\langle \nu, \nu'', E \rangle$ to $\mathcal{A}$. In that case the world persists in its values at ν'' and at subsequent time-points until another initialisation message is received from the ego.

- *The ego has not decided to initiate an action at time ν, so that the current-action set $\mathcal{C}$ is empty. In this case the world's state simply persists in its values at ν and at subsequent new time-points until an initialisation message is received from the ego.*

- *No other rules apply.* ■

Proposition 3.2 *The rules of the game correctly revise the board.*

3.1.3 Language and Models

Let $\mathcal{T}$ be the timepoint domain (section 3.1.1), $\mathcal{F}$ the set of all feature symbols, $\mathcal{V}$ the domain of all feature values, and $\mathcal{E}$ the set of all action symbols. Let $\langle \mathcal{H}, \sqsubseteq \rangle$ be the lattice which elements, called *observations*, are members of $\mathcal{H}$ and the order relation $\sqsubseteq$ applies as follows: $\langle t_1, f_1, \{\ldots\}, v_1 \rangle \sqsubseteq \langle t_2, f_2, \{\ldots\}, v_2 \rangle$ iff $t_1 \sqsubseteq t_2$. The tuple $\langle t, f, \{\ldots\}, \mathit{unknown} \rangle$ is an abbreviation for $\bigvee_i \langle t, f, \{\ldots\}, v_i \rangle$, varying i over all possible tuples $\langle t, f, \{\ldots\}, v_i \rangle$ in $\mathcal{H}$. Let $\langle \mathcal{D}, \sqsubseteq \rangle$ be the lattice which 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: $\langle s_1, t_1, A_1 \rangle \sqsubseteq \langle s_2, t_2, A_2 \rangle$ iff $s_1 \sqsubseteq s_2$. The order relation $\sqsubset$ is an abbreviation for $\sqsubseteq \wedge \neq$. The relation $\langle s_1, t_1, A_1 \rangle = \langle s_2, t_2, A_2 \rangle$ simply means that A_1 and A_2 start at the same timepoint, while $\langle s_1, t_1, A_1 \rangle \sqsubset \langle s_2, t_2, A_2 \rangle$ means that A_1 starts earlier than A_2. Let Υ be a scenario description.

- The OBS part of Υ is a sub-lattice of $\langle \mathcal{H}_\Upsilon, \sqsubseteq \rangle$, which elements are members of $\mathcal{H}_\Upsilon = \mathcal{T} \times \mathcal{F}_\Upsilon \times 2^{\mathcal{O}} \times \mathcal{V} \subseteq \mathcal{H}$, where F_Υ is the set of all features explicitly occurring in Υ.
- The SCD part of Υ is a sub-lattice of $\langle \mathcal{D}, \sqsubseteq \rangle$. Each tuple in SCD specifies the starting time, the ending time and the action symbol of an action scheduled for execution.
- The function $\Rrightarrow : \mathcal{D} \to 2^{\mathcal{H}_\Upsilon}$ maps each schedule's occurrence in a set of non-empty lattices of observations. The function $\Rrightarrow$ is parametric on the action type, and the LAW part of Υ consists in the definition of $\Rrightarrow$ as a set of action laws in Full Trajectory Normal Form, one law for each action type.

As for $\mathcal{K}$-**IA**, a *scenario* for W is a set of games where W was the world player, and a *scenario description* is the description of W itself and of how its state has changed over time. The definition of scenario description for $\mathcal{K}$-**RA** is almost identical to definition 2.5 of scenario description for $\mathcal{K}$-**IA**. The difference between the definitions consists in the new assumption of "strict inertia in continuous time", represented with the ontological designator **RA**, the new set-theoretic description of the world (definition 3.3) and the corresponding logical description (definition 3.4). A scenario description is a recursive definition for problem domain specification, where observations in OBS and occurrences of actions in SCD are regarded as being *true propositions*, and action laws in LAW are regarded as being *rules* which from a number of true propositions they permit to infer other true propositions. In the specific case of action laws, inferred propositions can only be observations, i.e. tuples in $\mathcal{H}$.

The definition of classical model set for Υ in $\mathcal{K}$-**RA** is identical to definition 2.6, where the expansion of all formulae in SCD using their corresponding definitions in LAW is now possible by means of definition 3.3.

The definition of intended model set for Υ in $\mathcal{K}$-**RA** is identical to definition 2.7, where **RA** occurs in lieu of **IA** and the rules of the game are those of definition 3.5.

Proposition 3.3 $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) \subseteq \Sigma_{\mathcal{K}\text{-}\mathbf{RA}}(\Upsilon) \subseteq [\![\Upsilon]\!]$.

PROOF. *The model set $\Sigma_{\mathcal{K}\text{-}\mathbf{RA}}(\Upsilon)$ equals $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon)$ in case of discrete worlds, while $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon)$ is empty in case of continuous or almost-continuous worlds, so that $\Sigma_{\mathcal{K}\text{-}\mathbf{IA}}(\Upsilon) \subseteq \Sigma_{\mathcal{K}\text{-}\mathbf{RA}}(\Upsilon)$. The set $[\![\Upsilon]\!]$ allows models which do not satisfy inertia and models where there are additional actions besides those specified in SCD, so that $\Sigma_{\mathcal{K}\text{-}\mathbf{RA}}(\Upsilon) \subseteq [\![\Upsilon]\!]$.* ■

3.1.4 Concurrency

So far we allowed only one action to be executed at a time. This is a fairly strong simplification, for which the dynamical system is inhabited by a single agent, the agent can execute a single action at the time, and the action takes complete control of the influenced features. The more complex the dynamical system, the more it is expected to happen. A typical dynamical system is inhabited by two or more autonomous agents that are either *cooperating and negotiating*, *combating* or *independent and not interfering*. Each single agent can execute more than a single action at a time, and each action may have to share the control of the influenced features. From the standpoint of the game, as the vehicle of each agent is assumed to be part of the world, it is the world that performs all actions, so to change the feature values. Sharing the control of the influenced features is then a duty of the world, as player. The most interesting case is certainly the most general one, at least from the standpoint of Distributed Systems and Distributed AI. In this section, driven by the single aim of capturing classical Newtonian Mechanics, we are concerned with the simpler case only.

The case of "concurrency of actions" is designated, in Sandewall's taxonomy, with the **C** ontological assumption. The simpler case of "concurrency of independent actions" is obtained by specialising **C** with the **i** ontological sub-characteristics. No structural extensions are required for the underlying semantics of a world with concurrent independent actions. When playing the game, if the ego initiates several independent actions at the same time τ, then the world simply chooses one trajectory function non-deterministically. If $\varphi_j^1, \ldots, \varphi_j^n$ partial fluents influence the same feature f_j during $[\tau, \mu]$, each one independently on the other, then only one partial fluent will be considered for that feature as the composed partial fluent $\varphi_j^1 \circ \ldots \circ \varphi_j^n$ of the given partial fluents, and its definitional domain will be the closed and limited interval in time $[\tau, \mu]$. The symbol "$\circ$" represents the composition operator. When such composed function is a constant function c, then persistence arise for the influenced feature if and only if 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). The rules just given allow the case of n ego players, each player acting concurrently and independently from the others.

Scenario descriptions in $\mathcal{K}$-**RACi** are substantially identical to scenario descriptions in $\mathcal{K}$-**RA**, except for action occurrences which are now allowed to overlap in at least one time-point. The intended and classical model set are defined as usual, and the relation between $\mathcal{K}$-**RACi** models and $\mathcal{K}$-**RA** models is described by the following proposition:

Proposition 3.4 *Let $\Upsilon_1, \ldots, \Upsilon_n$ be scenario descriptions, each of which is of the form $\Upsilon_i = \langle \mathcal{K}, \mathcal{O}, \langle \mathbf{RA}, LAW_i \rangle, SCD_i, OBS \rangle$. Assume also that $\cup_{i=1}^n SCD_i$ is a schedule appropriate for the* **RACi** *ontological family. The following relation holds:*

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

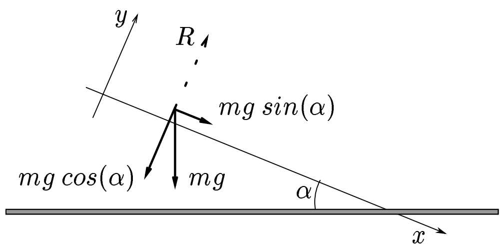
PROOF. *Independent concurrent actions are compositional, as their joint effect is the sum of their individual effects.* ■

Corollary 3.1 $\Sigma_{\mathcal{K}\text{-}\mathbf{RA}}(\Upsilon) \subseteq \Sigma_{\mathcal{K}\text{-}\mathbf{RACi}}(\Upsilon) \subseteq [\![\Upsilon]\!]$

3.2 Examples

The following is an example of reasoning problem in $\mathcal{K}$-**RACi**.

Scenario 3.1 (Galileo's inclined plane) *Given is an inclined plane with no friction, and let α be the angle between the inclined plane and the horizontal plane. We place an object of mass m on the inclined plane, then we release the object. Will the object move along the inclined plane? What are the involved "actions" at a given point in time τ and what is their "effect"?*



According to the given description, the mass of the object is m and the angle is α; then $[0]mass \hat{=} m$ and $[0]angle \hat{=} \alpha$ are in the OBS part of the scenario description. Furthermore, two forces are always exerted on the object: the weight force $W = mg$ and the reaction R of the plane (no other forces apply), then the SCD part of the scenario description consists of $[\Box]W(o)$ and $[\Box]R(o)$, where o is an object constant. Finally, the LAW part of the scenario description consists of the action laws in Full Trajectory Normal Form for W and R:

$$[s,t]W(obj) \Rrightarrow \forall\tau \in [s,t].[\tau]force(obj) \dot{=} \varphi_1(s,\tau,t)$$
$$[s,t]R(obj) \Rrightarrow \forall\tau \in [s,t].[\tau]force(obj) \dot{=} \varphi_2(s,\tau,t)$$

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 $[0]mass * g * cos([0]angle)$, while φ_2 is defined as $-[0]mass * g * cos([0]angle)$. According to the ego-world game, the force exerted on the object at time τ is

$$H(\tau, force(obj)) = \varphi_1(s,\tau,t) \circ \varphi_2(s,\tau,t) = mg\ sin(\alpha)$$

As the scenario was not explicit on α, a number of possible situations are associated to it:

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

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, like in figure 1, and we obtain the case of inertia with influencing actions. If $\alpha = \frac{\pi}{2}$, then the inclined plane is a vertical plane, so that $\varphi_2(s,\tau,t) = 0$. Passing time α may change as the effect of a lifting

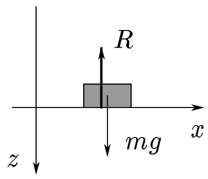


Figure 1: Inertia

action; as the action laws are parametric in α, the underlying semantics is tolerant in that respect.

Galileo's inclined plane then involves the frame [9] problem, as well as the prediction [10] and causal-qualification [11] problems. The frame problem finds its solution with Newton's principle of inertia. The property of *strict inertia* (page 9) is a generalization of Newton's principle to all features of objects in the image-level world, where the position in space is one of such features. The temporal projection problem here consists in predicting the motion of the object given the initial conditions of position and velocity has been assigned to it, and is addressed in its general terms by proposition 3.5. The causal-qualification problem occurs as follows. In Physics, the gravity force normally causes the objects to fall; the phenomenon may be fully inhibited because of the resistance exerted by an horizontal plane, which causes the object to rest in its position according to the principle of inertia. The underlying semantics implements the same solution via the operator $\circ$; the free fall corresponds to the case $\alpha = 0$, and the situation where the phenomenon is inhibited corresponds to the case $\alpha = \frac{\pi}{2}$, also referred to as "inertia with influencing actions", where the operator $\circ$ is used as shown.

This above scenario plays a practical role when it comes to the embodiment of an intelligent agent. As robots are resource-bounded embodied agents, it is not difficult to picture the situation where a robot with wheels, for example, is faced with the task of climbing an inclined platform with the available energy, taking into consideration its own weight and the minimum expense of energy that is required in order to achieve this goal with a certain desired speed. Although no planning techniques are given in this paper, the above design shows how the problem could be represented and reasoned about within the $\mathcal{K}$**sp**-**RAdCi** family of simplifying assumptions.

The following variant leads to the dividing-instant [12] problem.

Scenario 3.2 (first variant) *In addition to the inclined plane scenario, we assume $\alpha = 0$ and a force F which is exerted on the object of mass m, which direction is perpendicular to the plane and the verse moves along the reaction of the plane. After reaching a certain hight h, the object will return to the plane. Is the object raising or falling at the apex of its revolution?*

According to the classical mathematical account of the continuum, if a function $f(x)$ admits two different values for the same x, then the function

[9] Although the terms has been used already throughout the paper, it is worth here reminding its meaning. *The frame problem is the problem of how to express the facts about the effects of actions and other events in such a way that it is not necessary to explicitly state for every event, the fluents it does not affect.* (McCarthy, "Concepts of Logical AI", 1998)

[10] The Temporal Projection (Prediction) Problem [40] is the problem of how to determine facts about future situations. Given an initial description of the world (some facts that are true), 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?

[11] The Causal-Qualification Problem [85] 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.

[12] The Dividing-Instant Problem is commonly attributed to Plato (*Parmenides*). According to the AI literature, the problem consists in establishing what is the value of a certain feature at that instant in time dividing two time periods where the feature has neat different values.

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 first-order 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 formally enforces the intuition that things change smoothly. An immediate consequence of continuity, respected by all systems of qualitative reasoning, is that a changing quantity must pass through all intermediate values (Bolzano's theorem [13]) so that, in particular it passes through the dividing-instant point. The dividing-instant problem then seems to arise *only if* actions with instantaneous effects are allowed. In that case, the consequences that are associated in violating either laws must be accepted. As trajectories are nonempty sets, i.e. instantaneous effects are *not* allowed, the dividing-instant problem is *not* expected to occur in $\mathcal{K}$-**IA** and $\mathcal{K}$-**RACi** scenarios.

The following variant leads to the ramification [14] problem:

Scenario 3.3 (second variant) *In addition to the inclined plane scenario, the block o_2 of mass m_2 is placed on the block o_1 of mass m_1, and $\alpha = 0$. The scenario is identical to Baker's ice-cream scenario [6], where a person (the block o_1) is holding the ice-cream (the block o_2). A tiny force F is then exerted on block o_1, i.e. the person slowly moves to a different location. We should conclude that block o_2 moves together with block o_1, namely the ice-cream moves together with the person.*

The problem is solved via the operator $\circ$. Under the condition of a tiny F, the friction between the blocks is a null resistance, the two blocks are not expected to change its respective position and they form an isolated system of its own. The block o_2 rests on the block o_1, according to the known case of inertia with influencing actions. Block o_1, in the meantime, moves or rectilinear and uniform motion, until the tiny force F applies. The Block o_2, that forms a single system together with o_1, moves too.

The following scenario involves both the causal qualification and the structure-based ramification problems:

Scenario 3.4 *Five domino tokens are initially "up" and aligned in such a way that if the token $i-1$ falls then the token i is pushed. Successively, a* push *on the first token and a* wait *action occur. One expects the last token to fall after a short while. An action is then performed that inserts and holds a card between the fourth and the fifth token. After the waiting period, one expects the last token to persist in its original state.*

The structure-based ramification problem is solved via the $\circ$ operator, as no distinction is made between the force that pushes on the first token and the successive ones; a token falls only if it is pushed somehow. The causal-qualification problem is solved in exactly the same way; as no force applies on the last token, there is no reason why it should fall.

[13] If f is continuous on a closed interval $[a, b]$ and $f(a)$ and $f(b)$ have opposite signs, then there exits a number c in the open interval (a, b) such that $f(c) = 0$.

[14] The Structure-Based Ramification Problem [85] is the phenomenon where the world consists of a number of objects with relationships between them, and changes in one object (eg. the effects of actions on that object) are dependent on properties of adjacent objects in the structure.

We conclude with a folk scenario:

Scenario 3.5 (Achilles and the tortoise) *The tortoise defied the swift-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 scenario is properly an example of rectilinear and uniform motion in Newtonian Mechanics. The scenario has the following description:

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

where $\mathbf{s}_1$ and $\mathbf{s}_2$ are temporal constants, while s, t, t_1 and 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 description clearly falls within $\mathcal{K}$**sp-RAdCi**, and the trajectory of the running action describes the continuous change during the scheduled time interval. 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 can define the succession (τ_n) as follows:

$$\begin{array}{l}
S_A(\tau_1) = d \\
S_A(\tau_{n+1}) = S_T(\tau_n)
\end{array}$$

so that $S_A(\tau_{n+1}) = V_A * \tau_{n+1} = d + V_T * \tau_n = S_T(\tau_n)$. As $V_A * \tau_1 = S_A(\tau_1) = d$, then the following relations hold:

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

Zeno asserts Achilles will never reach the tortoise, that is the sum of all intervals $(\tau_{n+1} - \tau_n)$ is infinite. 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 as 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 duration, distance and sum of an infinite serie. According to the resulting account on series, we have:

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

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

in fact $(\star)$ is $\frac{V_T}{V_A} < 1$ per hypothesis. The swift-footed Achilles will then reach the tortoise at τ^*, in fact is $S_A(\tau^*) = S_T(\tau^*)$, that is

$$[\tau^*]position(Achilles) = [\tau^*]position(Tortoise)$$

The reader will surely *not* blame us for the mathematical solution. Over more than twothousand years of philosophical disquisitions, in fact, the paradox of Zeno still remains an unsolved problem within pure reasoning.

3.3 Assessment and Boundaries

The search for a small number of principles that could accurately describe the whole range of observed phenomena, and formally explain those which escaped common-sense, has been the major task of Physics for centuries, where Physics played the role of experimental science and Mathematics served as language and computational instrument. The outcome of such an effort was the paper *Philosophiae Naturalis Principia Mathematica* by Newton [68], which theory is based on three fundamental principles together with (nowadays) a number of well refined numerical techniques, and is taught at all degrees of education. As classical Newtonian Physics implicitly involves a number of epistemological and ontological assumptions, is it possible to enumerate all these assumptions and capture them by means of ego-world game developments in Sandewall's style? The following proposition 3.5 gives a partial result in this sense. Its boundaries are outlined.

Theorem 3.1 (Cauchy, 1820 [18]) *Let a and b be continuous functions on an interval $I \subset \mathbb{R}$. Given $t_0 \in I$ and $x_0 \in \mathbb{R}$, exists and is unique the solution of 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:

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

■

Proposition 3.5 *The epistemological and ontological assumptions that are involved with Newton's three principles of Mechanics are described by the* $\mathcal{K}$**sp-RAdCi** *family of assumptions.*

PROOF.

> First principle: *In an inertial reference frame, a body persists in its state of rest or of rectilinear uniform motion unless compelled to change that state by some force acting on it.*
>
> Second principle: *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 $F = ma$.*

> Principle of independence of simultaneous actions: *The net force exerted on a body is the vectorial sum of the individual forces exerted on it by the material systems interacting with it.*
>
> Third Principle: *If a body exerts a force on another body, the latter exerts on the former a force of equal magnitude and opposite sign, that is* $F_{12} = -F_{21}$.

By expressing with $x(t)$ the position in space of the object at time t, the equation $F = ma$ — where F is the component of the external force along the direction of the movement, m is the mass (constant) of the object and a is the acceleration of the object — can be written as the second order differential equation $F(t, x(t)) = mx''(t)$, 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 of such differential equation, that satisfies the initial conditions of position and velocity has been assigned to the object. The equation $x''(t) = \frac{1}{m}F(t, x(t))$ then reduces to the following problem [15]:

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

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

The first principle, or *principle of inertia*, finds an immediate correspondence with the *property of strict inertia.* The first and second principle then invoke the $\mathcal{K}$**sp-RAdCi** class of assumptions. The epistemological sub-characteristic **s** restricts $\mathcal{K}$ to the case of *complete knowledge about the initial state*, which is the necessary condition in order to apply Cauchy's theorem. The epistemological sub-characteristics **p** restricts $\mathcal{K}$ to the case of *no information at any later state than the initial one.* The ontological characteristics **R**, already needed for the first law, finds here its motivation on the applicability of Cauchy's theorem and on the whole underlying mathematical theory of the continuum. 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 world at the time when the action started, which is properly the result of applying Cauchy's theorem. The ontological sub-characteristics **i** restricts **C** to the case of *concurrency of independent actions*, which definition (per 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.

Finally, the information privided by the *principle of action and reaction* represents no technical challenges, and can be explicitly represented in the scenario description. ■

We shall now justify the use of the term "partially" in our thesis.

---

[15] Although Kuipers does not acknowledge Cauchy in his book, he does refer to a Cauchy problem as the definition of a "dynamical system" [55, page 316]. The terminology is the one adopted within Qualitative Reasoning.

The boundaries of the result are as follows.

On the first principle. The following classical experiment (ghost forces) triggers the issue.

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; 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.

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* of the experiments the disk moved along a straight line. According to classical Physics, an *observer* is needed. It can be located in an inertial or non-inertial reference frame, with consequences in his comprehension of phenomena. The first camera was located in the inertial reference frame, 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 given example is a reasoning problem for a theory of relative truth, where the contexts are the inertial and the non-inertial system. The related extension of the underlying semantics, so to include this sort of problems, would require the ability to see the same problem from different viewpoints, which we see it as a special case of the more general ability to switch context, possibly changing the underlying reasoning method. The work by Attardi and Simi on the formalization of viewpoints [1, 94, 2, 3, 4] may be a good starting point in this respect.

On the third principle. Also known as the *principle of action and reaction*, it deduces one of the fundamental and most popular principles of the whole Physics, namely 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 constructions are classical and given for granted. There are no known exceptions to the principle of preservation of the quantity of motion. In an 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 simply 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 the underlying semantics 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 1981, like the program BACON [58]. The issue is certainly novel within the metatheoretic approach. The term "partial" is here well justified; although the information provided by the third principle can be explicitly represented in the scenario description, and presents no evident simplifying assumptions, 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.

Another boundary arises from a letter of Einstein to Popper [72]:

> *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!* (A. Einstein. Old Lyme, 11 September 1935)

Although the full $\mathcal{K}$-**RACi** class allows for worlds with incomplete knowledge about the initial state, Einstein's example clearly does not fall into the class. The work by Pearl [70] may be a starting point in this respect.

We conclude by comparing the two methodologies of research. In Physics, the methodology of research has been summarised by Rosati [81] as follows: (we translate from the original Italian text)

> *Physics is addressed as a quantitative science based on experiments that can be performed on the world we are inhabiting; the results of the experiments are systematised and correlated according to the scientific method, which is the founding basis not only of Physics, but also of the other quantitative sciences like Chemistry, Biology or Astronomy. The observation of a phenomenon consists in the accurate analysis of its characteristics, of the circumstances that lead to it together with all the influencing factors; in order to facilitate the observation, the phenomenon is let to occur in conditions that can be accurately reproduced so that the process can be repeated and its causes varied. The scientific method consists in searching theories that are valid and capable of explaining the broadest possible range of experimental facts by means of a small number of principles which are supported by the observations. Those principles are accepted as postulates within Physics, which validity is founded on their internal coherence and the ability to explain the most diverse phenomena by means of quantitative mathematical information processing. Since no certainty can be achieved on the non-contradictoriness of the principles, and it is always possible that a later experiment may invalidate them, we have that: in Physics, a theory is never completely proved.*

In Nonmonotonic Reasoning about Actions and Change, the methodology of research has been summarised by Davis [22] as follows:

> *The basic approach used here, as in much of the research on automating common-sense reasoning, is to take a number of examples of common-sense inference in a common-sense 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.*

A quick comparison reveals that each "example" of the latter does not even play the role that a single "experiment" may play on the former. The problem is well known, as it was first raised by Hayes in 1978 [41, 42] when complaining against AI's emphasis on toy worlds. That observation from Hayes led to what is now referred to as Qualitative Reasoning (see, eg. [55]). A general aspect of the same problem was also summarised by Hanks and McDermott [40] as follows:

> *Non-monotonic 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.*

The same problem was then emphasised by Brooks [12] a few years later, after the failure of the SHAKEY [27] project. Brooks emphasis led to what still is the approach in MIT AI-Lab, namely research in Artificial Intelligence without the *knowledge representation and reasoning* part. There are, however, additional issues of interest. The "*search for theories that are valid and capable of explaining the broadest possible range of experimental facts by means of a small number of principles which are supported by the observations*" as well as the "*systematisation and correlation of the results*" of Physics research do not find their counterpart in NRAC research, where the design of "*a formal language*" for addressing "*a number of examples*" is preferred, in determining lack of generality. This aspect of the scientific method is now recovered by the meta-theoretic approach we summarised in section 2, where explicit simplifying assumptions are used when representing and reasoning about problems, the systematisation and correlation of the results is an explicit issue in research, and theories are classified according to their range of applicability. Both methods share the same lack of certainty on the non-contradictoriness of the principles, although classical Physics has stood the test of time and can be considered as being fairly robust at this point, as far as the study of the utterly small is not of interest. As NRAC research is just about thirty years old, and the meta-theoretic approach has been around for about a third of that time, it is now easy to foresee that one of its logics will be able to switch between different reasoning methods depending on the different reasoning problem, where each reasoning method is assessed against the taxonomy of simplifying assumptions, although the day will come for an additional problem falling outside each of these classes, so that the work of one of us would still be required to understand the properties of that and similar problems, define the corresponding underlying semantics, and then extend and assess the given logic.

3.4 More results

All methods cited in section 2.4 are assessed aginst subclasses of $\mathcal{K}$-**RACi**. In [11], within Logic Programming, an algebraic semantics has been designed and proven sound and complete with respect to the full $\mathcal{K}$-**RACi** class.

4 Conclusion

We summarised the complete definition, the basic properties and the results of the meta-theoretic approach to non-monotonic reasoning, including the definition of the $\mathcal{K}$-**IA** and $\mathcal{K}$-**RACi** classes of reasoning problems.

The underlying semantics for $\mathcal{K}$-**IA** (section 2.3) defines the intended model set for each reasoning problem where "accurate and complete information about actions" is the epistemological assumption, represented by the designator $\mathcal{K}$, and "strict inertia in discrete (integer) time, discrete change and alternative results of sequential actions" are the ontological assumptions, represented by the designator **IA**.

The underlying semantics for $\mathcal{K}$-**RACi** (section 3.1) defines the intended model set for each reasoning problem where "accurate and complete information about actions" is the epistemological assumption, represented by the designator $\mathcal{K}$, and "strict inertia in continuous time, continuous change and alternative results of possibly concurrent and independent actions" are the ontological assumptions, represented by the designator **RACi**.

The extension of $\mathcal{K}$-**IA** to $\mathcal{K}$-**RACi** (section 3.1) and the assessment results (sections 2.4, 3.4) imply that the already successful meta-theoretic approach applies also to logics for the continuum, as well as to Logic Programming in general. The open problem remains as to how the proof-theoretic and the model-theoretic approaches apply to the $\mathcal{K}$-**RACi** family of assumptions. Within the proof-theoretic approach the problem appears to be novel, although there are early investigations on the relation between intuitionist logic and the continuum (see, eg. [53]). Within the model-theoretic approach, it was Prior who first considered propositions which could change their truth-values with the passage of time, according to a certain modal criterion of validity. According to Prior's account, it results that there are formulae which are not valid when time is taken to be continuous but which are valid when time is taken to be discrete. The additional problem with Prior's approach is that propositions are truth-valued; the case of interest demands for discrete and/or continuous change, hence requiring a many-sorted and many-valued extension of Prior's modal criterion, if this extension is possible at all.

Of main concern in this paper was the *partial* classification of the simplifying assumptions for a classical theory of causation: Newtonian Mechanics. Exploiting the theorem from Augustin Louis Cauchy on the solution of first-order linear differential equations, the $\mathcal{K}$**sp**-**RAdCi** subclass of $\mathcal{K}$-**RACi** is shown to include the epistemological and ontological assumptions that were implicitly involved by Newton in the first and the second of his three principles of Mechanics. The boundaries of the result are outlined, suggesting the need for a further extension of the underlying semantics in two main directions: viewpoints, and learning from inconsistencies.

Acknowledgements

The École Normale Supérieure in Pisa is acknowledged for allowing the search of the mentioned work from Augustin Louis Cauchy in its private collection of antiques. The original work, known to be dated 1820, resulted to be a collection of unpublished lecture notes, supposed lost until recently.

References

[1] G. Attardi and M. Simi. Metalanguage and reasoning across viewpoints. In *European Conference on Artificial Intelligence (ECAI)*, 1986.

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

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

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

[5] St. Augustine. Confessions, 401 A.D.

[6] A. B. Baker. Non-monotonic Reasoning in the Framework of the Situation Calculus. *Artificial Intelligence*, 49:5–23, 1991.

[7] E. W. Beth. *The foundations of mathematics.* North-Holland, 1965.

[8] B. Bolzano. *Wissenschaftslehre.* Sulzbach, 1837.

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

[10] S. Brandano. $\mathcal{K}$-**RACi**. In *Non-monotonic Reasoning, Action and Change, Proceedings of the International (IJCAI) Workshop*, pages 9–16, Stockholm, Sweden, August 1999.

[11] S. Brandano. Features and Fluents for Logic Programming: Non-simulative Algebraic Semantics. In J. Cunningham and D. Gabbay, editors, *Formal and Applied Practical Reasoning (FAPR), Proceedings of the 3rd International Conference*, pages 131–143, Imperial College, 2000. ISSN: 1469-4166.

[12] R. A. Brooks. Intelligence without Reason. In *International Joint Conference on Artificial Intelligence (IJCAI), Proceedings*, pages 569–595, 1991. Computers and Thought Award.

[13] R. A. Brooks. Intelligence without Representation. *Artificial Intelligence*, 47:139–159, 1991.

[14] R. A. Brooks. *Cambrian Intelligence: The Early History of the New AI.* MIT Press, 1999.

[15] A. L. Brown and Y. Shoham. New results on semantical non-monotonic reasoning. In *Non-Monotonic Reasoning, Proceedings of the International Workshop*, volume 346 of *Lecture Notes in Artificial Intelligence*, pages 19–26. Springer-Verlag, 1989.

[16] R. Carnap. *Logische Syntax der Sorache.* 1937.

[17] R. Carnap. *Meaning and Necessity.* 1957.

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

[19] K. L. Clark. Negation as Failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases, Proceedings of the Workshop*, pages 293–322. Plenum Press, New York, 1978.

[20] K. L. Clark. *Predicate Logic as a Computational Formalism.* PhD thesis, University of London, Queen Mary College, Department of Computer Science, 1980.

[21] J. M. Crawford and D. Etherington. Formalizing reasoning about change: A Qualitative Reasoning approach (preliminary report). In *National (USA) Conference on Artificial Intelligence (AAAI), Proceedings*, 1992.

[22] E. Davis. *Representation of Commonsense Knowledge.* Morgan Kauffman, 1990.

[23] M. Denecker. *Knowledge Representation and Reasoning in Incomplete Logic Programming.* PhD thesis, University of Leuven, 1993.

[24] M. Denecker and D. De Schreye. SLDNFA: an abductive procedure for normal abductive programs. In *International Joint Conference and Symposium on Logic Programming*, pages 686–700. MIT Press, 1992.

[25] M. Denecker and D. De Schreye. Representing incomplete knowledge in Abductive Logic Programming. *Journal of Logic and Computation*, 5(5):553–578, 1995.

[26] J. Doyle. A Truth Maintenance System. *Artificial Intelligence*, 12:231–272, 1979.

[27] N. J. Nilsson ed. SHAKEY the robot. Technical Report 323, SRI, Menlo Park, California, 1984.

[28] M. Freund. Supra-compact Inference Operations. In *Non-monotonic and Inductive Logic, Proceedings of the 1st International Workshop*, volume 543 of *Lecture Notes in Artificial Intelligence*, pages 59–73. Springer-Verlag, 1990.

[29] D. M. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In K. R. Apt, editor, *Logics and Models of Concurrent Systems*, pages 439–457. Springer-Verlag, 1985.

[30] D. M. Gabbay. A General Theory of Structured Consequence Relations. In *Substructural Logics*, pages 109–151. Oxford University Press, 1993.

[31] D. M. Gabbay, editor. *Substructural Logics.* Oxford University Press, 1993.

[32] P. Gärdenfors. *Knowledge in Flux: Modelling the dynamics of Epistemic States.* MIT Press, 1990.

[33] P. Gärdenfors. Belief Revision: A Vademecum. In A. Pettirossi, editor, *Meta-Programming in Logic (META), Proceedings of the Third International Workshop*, volume 649 of *Lecture Notes in Computer Science*, pages 1–10, Uppsala, Sweden, June 10-12, 1992. Springer-Verlag.

[34] M. Gelfond and V. Lifschitz. Representing Actions and Change by Logic Programs. *Journal of Logic Programming*, 17, 1993.

[35] G. Gentzen. Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen. *Mathematische Annalen*, 107:329–350, 1932. Reprinted as "On the existence of independent axiom systems for infinite sentence systems" in [95] chapter 1.

[36] K. Gödel. On Formally Undecidable Propositions of Principia Mathematica and Related Systems. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. English translation in [101] pages 596–616.

[37] K. Gödel. The consistency of the axiom of choice and the generalized continuum hypothesis. *Annals of Mathematics Studies*, 3, 1940.

[38] J. Halpern and Y. Moses. Towards a Theory of Knowledge and Ignorance: Preliminary Report. In K. R. Apt, editor, *Logics and Models of Concurrent Systems*, pages 459–476. Springer-Verlag, 1985.

[39] S. Hanks and D. McDermott. Default Reasoning, Non-monotonic Logics, and the Frame Problem. In *National (USA) Conference on Artificial Intelligence (AAAI), Proceedings*, pages 328–333, Los Altos, CA, 1986. Morgan Kauffman.

[40] S. Hanks and D. McDermott. Non-monotonic Logic and Temporal Projection. *Artificial Intelligence*, 33:379–412, 1987. Best Paper Award at AAAI-1986.

[41] P. J. Hayes. The Naive Physics Manifesto. In D. Michie, editor, *Expert systems in the microelectronic age*. Edinburgh University Press, Edinburgh, Scotland, 1978.

[42] P. J. Hayes. The Second Naive Physics Manifesto. In J. R. Hobbs and R. C. Moore, editors, *Formal Theories of the Commonsense World*. Ablex Publishing Corporation, Norwood, New Jersey, 1985.

[43] H. Herre. Compactness properties of non-monotonic inference operations. In *Logics in Artificial Intelligence, Proceedings of the European Workshop (JELIA)*, volume 838 of *Lecture Notes in Artificial Intelligence*, pages 19–33. Springer-Verlag, 1994.

[44] D. Hilbert. Mathematische Probleme. In *Internationalen Mathematikerkongress*, pages 253–297, Paris, 1900.

[45] D. Hilbert. Über die Grundlagen der Logik und der Arithmetik. In *Internationalen Mathematikerkongress*, pages 174–185, Heidelberg, 1904. In English: *On the foundations of Logic and Arithmetic, in [101]*.

[46] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–190, 1926.

[47] W. Hodges. *Logic*. Pelican Books, 1977.

[48] S. Ishtiaq and P. W. O'Hearn. BI as an Assertion Language for Mutable Data Structures. In *Principles of Programming Languages (POPL), Proceedings of the Symposium*, 2001.

[49] N. R. Jennings. Agent-Based Computing: Promise and Perlis. In *International Joint Conference on Artificial Intelligence (IJCAI), Proceedings*, pages 1429–1436, 1999. Computers and Thought Award.

[50] N. R. Jennings, K. Sycara, and M. Wooldridge. A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, 1:7–38, 1998.

[51] N. R. Jennings and M. Wooldridge. Applications of Intelligent Agents. In *Agent Technology: Foundations, Applications and Markets*. SV, 1998.

[52] Y. Kaluzhny and D. Lehmann. Deductive Non-monotonic Inference Operations. *Journal of Logic and Computation*, 5(1):111–122, 1995.

[53] S. C. Kleene and R. Vesley. *The Foundations of Intuitionistic Mathematics, Especially in relation to Recursive Functions*. North-Holland, 1965.

[54] S. Kraus, D. Lehmann, and M. Magidor. Non-monotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence*, 44(1):167–207, 1990.

[55] B. J. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, 1994.

[56] I. Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, 1999 edition, 1976.

[57] L. Lamport. What Good is Temporal Logic? *Information Processing*, pages 657–668, 1983.

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

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

[60] V. Lifschitz. Circumscription. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 297–352. Oxford University Press, 1994.

[61] D. Makinson. General Theory of Cumulative Inference. In *Non-Monotonic Reasoning, Proceedings of the 2nd International Workshop*, volume 346 of *Lecture Notes in Artificial Intelligence*, pages 1–18. Springer-Verlag, 1989.

[62] D. Makinson. General Patterns in Non-monotonic Reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 35–110. Oxford University Press, 1994.

[63] J. McCarthy. Circumscription – A form of Non-Monotonic Reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[64] J. McCarthy. Modality for Robots. In *On the adequacy of Modal Logic from the standpoint of AI*. Debate at http://www.etaij.org/rac, 1999.

[65] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of Artificial Intelligence. *Machine Intelligence*, 4:463–502, 1969.

[66] R. Montague. Pragmatics (1968). In R. H. Thomason, editor, *Selected Papers of Richard Montague*. Yale University Press, 1974.

[67] C. Morris. *Foundations of the Theory of Signs*. Chicago, 1938.

[68] Isaac Newton. *Philosophiae Naturalis Principia Mathematica*. July 5 1686.

[69] P. W. O'Hearn and D. Pym. The Logic of Bunched Implications. *Bulletin of Symbolic Logic*, 5(2):215–244, 1999.

[70] J. Pearl. Reasoning with Cause and Effect. In *International Joint Conference on Artificial Intelligence (IJCAI), Proceedings*, pages 1437–1449, Stockholm, Sweden, 1999.

[71] A. Pnueli. The Temporal Logic of Programs. In *Foundations of Computer Science, Proceedings of the IEEE Symposium*, pages 46–67, 1977.

[72] K. R. Popper. *The Logic of Scientific Discovery.* Harper and Row, 1959.

[73] D. Prawitz. Towards a Foundations of a General Proof Theory. In P. Suppes, L. Henkin, A. Joja, and GR. C. Moisil, editors, *Logic, Methodology and Philosophy of Science IV, Proceedings of the 1971 International Congress*, pages 225–250. North-Holland, 1973.

[74] D. Prawitz. On the idea of a General Proof Theory. *Synthese*, 27:63–77, 1974.

[75] A. N. Prior. *Time and Modality.* Oxford University Press, 1957.

[76] A. N. Prior. *Past, Present and Future.* Oxford University Press, 1967.

[77] H. Reichenbach. *The Direction of Time.* University of California Press, 1956.

[78] R. Reiter. On Closed World Data Bases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, 1978.

[79] R. Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13(1–2):81–132, 1980.

[80] R. Reiter. Circumscription implies Predicate Completion (sometimes). In *National (USA) Conference on Artificial Intelligence (AAAI), Proceedings*, pages 418–420. Morgan Kauffman, 1982.

[81] S. Rosati. *Fisica Generale*, volume 1. Casa Editrice Ambrosiana, Milano, 1986. Pisa Lectures, in Italian. Reprinted in 1988.

[82] E. Sandewall. An approach to the frame problem and its implementation. *Machine Intelligence*, 7:195–204, 1972.

[83] E. Sandewall. Combining logic and differential equations for describing real-world systems. In *Principles of Knowledge Representation and Reasoning (KR), Proceedings*, pages 412–420, 1989.

[84] E. Sandewall. Filter Preferential Entailment for the logic of action in almost continuous worlds. In *International Joint Conference on Artificial Intelligence (IJCAI), Proceedings*, pages 894–899, Detroit, USA, 1989.

[85] E. Sandewall. Causal Qualification and Structure-Based Ramification. Technical Report LiTH-IDA-R-92-42, Department of Computer and Information Science, Linköping University, Sweden, December 1992. Accepted to Common Sense'93; Austin, Texas.

[86] E. Sandewall. The Range of Applicability of Non-monotonic Logics for the Inertia Problem. In *International Joint Conference on Artificial Intelligence (IJCAI), Proceedings*, Chambery, France, 1993.

[87] E. Sandewall. *Features and Fluents: The Representation of Knowledge about Dynamical Systems*, volume 1. Oxford University Press, 1994. The errata corrige is available at the author's Internet Address.

[88] E. Sandewall and Y. Shoham. Non-monotonic Temporal Reasoning. In *Handbook of Artificial Intelligence and Logic Programming*. Oxford University Press, 1994.

[89] E. Sandewall and Y. Shoham. Non-monotonic Temporal Reasoning. Technical Report IDA-RKL-94-TR11, Department of Computer and Information Science, Linköping University, Sweden, 1994. Published as [88].

[90] K. Schlechta. Some Results on Classical Preferential Models. *Journal of Logic and Computation*, 2(6):675–686, 1992.

[91] D. Scott. Completeness and Axiomatizability in Many-Valued Logic. In L. Henkin, editor, *Proceedings of the Tarski Symposium*, pages 411–435. American Mathematical Society, 1974.

[92] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. PhD thesis, Yale University, 1986. Published as [93].

[93] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, second edition, 1988.

[94] M. Simi. Viewpoints subsume Belief, Truth and Situations. In *Trends in Artificial Intelligence, Proceedings of the 2nd Congress of the Italian Association for Artificial Intelligence*, volume 549 of *Lecture Notes in Artificial Intelligence*, pages 38–47, 1991.

[95] M. E. Szabo, editor. *The Collected Papers of Gerhard Gentzen*. Studies in Logic and the Foundations of Mathematics. North-Holland, 1969.

[96] A. Tarski. On the Concept of Logical Consequence. In *Actes du Congrès International de Philosophie Scientifique*, Paris, 1936. Reprinted in [97].

[97] A. Tarski. *Logic, Semantics, Meta-Mathematics: Papers from 1923 to 1938*. Oxford University Press, second edition, 1956.

[98] M. Thielscher. An Analysis of Systematic Approaches to Reasoning about Actions and Change. In *Artificial Intelligence : Methodology, Systems, Applications, Proceedings of the International Conference (AIMSA)*, 1994.

[99] A. M. Turing. Computing Machinery and Intelligence. *Mind*, 59:433–460, 1950.

[100] A. M. Turing. Computing Machinery and Thought. In E. A. Feigenbaum and J. Feldman, editors, *Computers and Thought*. McGraw-Hill, New York, 1963.

[101] J. van Heijenoort, editor. *From Frege to Gödel: A source Book in Mathematical Logic*. Harvard University Press, 1967.

[102] M. Wooldridge and N. R. Jennings. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2), 1995.

[103] J. Worrall. Imre lakatos (1922-1974) : philosopher of mathematics and philosopher of science. *Z. Allgemeine Wissenschaftstheorie*, 5(2):211–217, 1974.